Cronfa
Swansea University's Research Repository

An investigation on the Mutagenicity of Compounds using Benchmark Dose and Dose-Response Data for Research in Genetic Toxicology

Yusuf Hussien BSc (Hons)

Submitted to Swansea University for the degree of Doctor of Philosophy

2025

Copyright: The Author, Yusuf Hussien, 2026.

DECLERATION

This work has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree.

Signed.....

30/06/2025

Date.....

STATEMENT 1

This thesis is the result of my own investigations, except where otherwise stated. Other sources are acknowledged by footnotes giving explicit references. A bibliography is appended.

Signed.

30/06/2025

Date.....

STATEMENT 2

I hereby give consent for my thesis, if accepted, to be available for photocopying and for interlibrary loan, and for the title and summary to be made available to outside organisations.

Signed..

30/06/2025

Date.....

STATEMENT 3

The University's ethical procedures have been followed and, where appropriate, that ethical approval has been granted.

Signed.....

30/06/2025

Date.....

Acknowledgements

This PhD has been a transformative experience, and it would not have been possible without the guidance, support, patience and encouragement of many individuals and institutions. Firstly, I would like to express my deepest gratitude towards Dr George Johnson for allowing me to carry out this research under his supervision, supporting me through my development as a researcher, providing opportunities for me to travel to present my research at global conferences and introducing me to a network of world-class scientists and institutions in pharmacology and toxicology. This gratitude also extends to Dr Katherine Chapman for providing me with valuable feedback, constructive criticism and encouragement throughout this process. Furthermore, this PhD project has greatly benefited from the cooperation of the scientists at Litron Laboratories, Drs Erica Briggs, Steven Bryce, Jeffrey Bemis, Svetlana Avlasevich, Adam Conrad, who generously provided most of the data used in this project and shown me incredible kindness during my visit to their facilities. On top of that, I would like to give a special thanks to Dr Stephen Dertinger for being my "third supervisor" throughout this PhD and providing me with so many thought-provoking ideas. I wanted to thank Drs Carole Yauk, Ashley Allemang, Stefan Pfuhler, Marc Beal, Wen Sun, Maik Schuler and Andrew Williams for provisioning me with intellectual exchanges, stimulating my ability to research, assisting me with any programming/coding concerns, and encouraging me to pursue my goals and ambitions as a scientist. I would also like to express my gratitude to Giel Hendricks and the all the members at Toxys B.V. for allowing me to engage in close collaborations with them and continuing to work with them in the future for the development of my career. I'm also truly thankful for the funding provided to me by the Health and Environmental Science Institute Genetic Toxicology Technical Committee as their financial support has been a major factor in allowing me to transform this project into a full PhD and I am forever indebted to them for their generosity. Finally, I want to thank the friends and family who have stuck by me through this endeavour and a special thanks to my father, who worked incredibly hard through his entire life to subsidize my wellbeing and supporting me in making my academic dreams come true. Thank you, Swansea University, for allowing me to compile this thesis as it is a testament to the collective efforts of all these remarkable individuals who have been part of this journey.

Summary (Abstract)

Despite the current availability of analytical techniques for genotoxicity studies, limitations in quantitative methods restrict researchers from numerically analysing and explaining how various environmental, chemical, and physical sources cause genetic damage to cells. Benchmark dose serves as an important metric when gathering dose-response data from genotoxic compounds and multiple studies assess information on both to appropriately determine doses that result in genetic abnormalities for visualisation, interpretation, and review. The ability to map different compounds according to their genotoxicity allows for the comparative analysis of these compounds to be normalised and formatted in a coherent manner whilst retaining the depth and complexity that they initially had during the processing phase of each dataset. This investigation utilises data generated across multiple tests and assays (Litron Multiflow Genotoxicity Assay, Toxys ToxTracker Assay, TGx-DDI transcriptomic biomarker test, iScreen image analysis and *in vitro*-to-*in vivo* extrapolation) to quantitatively measure the genotoxic effects of various mutagenic compounds on cellular biomarkers (e.g. p53, γH2aX. phospho-histone H3, TGx-DDi, Bscl2, CENPA, MPM2, etc) using raw response (imaging induction or fold change/percentage) data generated from multiplexed flow cytometry imaging techniques to expand our understanding of mutagenicity whilst revealing the potential use of data analysis techniques to aid in genetics/cancer research. PROAST was used to calculate benchmark dose confidence intervals (BMD CI) that provide information on potency, and ToxPi was used to analyse the BMD CI across each endpoint and chemical, using approaches such as hierarchical clustering, principal component analysis (PCA) and providing a potency score. In conjunction with a weight-of-evidence approach, this combined analysis will inform us of the chemical grouping and potency across the multiple genetic toxicity endpoints and allow a quicker and more precise assessment of complex datasets. There will be benefits for hazard and risk assessment, while also providing increased weight to *in vitro* data sets.

Table of Contents

Declaration 2
Acknowledgements 3
Summary (Abstract) 4
List of Tables 10
List of Figures 12
Abbreviations 27

Chapter 1: General Introduction 31
1.1 Significance of Genetic Toxicology in Research 31
1.2 DNA Damage 32
1.3 Benchmark Dose-response Model & Dose-response Relationship 38
1.4 Genotoxicity Testing Models for Hazard Assessment 48
1.5 Modern Genotoxicity Testing Systems 53
1.5.1 Multiflow *In vitro* DNA Damage Assay for Modern Genotoxicity Testing 53
1.5.2 Microflow Assay Datasets 55
1.5.3 Toxys ToxTracker Assay 58
1.5.4 HESI GTTC TGx-DDI Transcriptomic Biomarker Test 63
1.5.5 The iScreen High-Content Imaging Platform 64
1.5.6 *In vitro In vivo* Extrapolatory modelling 66
1.6 Investigation Objectives, Genotoxicity Hazard Assessment Paradigms and Precautionary Considerations 69
Chapter 2: Benchmark Dose Analysis on the *in vitro* MultiFlow and MicroFlow Assays for Quantitative DNA Damage Biomarker Assessment for Aneugens or Clastogens Across Three Genotoxic Studies 72
2.1 Introduction 72
2.1.1 MultiFlow Assay Datasets Handling 73
2.1.2 MicroFlow Assay Datasets Handling 75
2.1.3 DNA damage biomarkers analysis to generate compound potency assessment and providing adverse outcome pathway information 75
2.1.4 The use of Benchmark Dose model to mark the point-of-departure and mark the margin of safety to determine genotoxic potency 76
2.1.5 Flow Cytometry as a scoring basis for *in vitro* Microflow and MultiFlow assay 77
2.1.6 The Three Components 77
2.1.6.1 Topoisomerase II inhibitors covering an array of mechanisms 77

2.1.6.2 Clastogens that form DNA-reactive metabolites through metabolic (S9) activation 84

2.1.6.3 Extended assessment of and expanded dataset on compound potency assessment 87

2.2 Materials and Methods 88

2.2.1 Dataset Acquisition 88

2.2.2 Cell treatments and MultiFlow processing for the assessment of identical targets with varying mechanisms of 8 topoisomerase II inhibitors 88

2.2.3 Cell treatments and MultiFlow processing to analyse metabolite and parent compound across 10 clastogens 93

2.2.4 Cell treatments and MultiFlow/MicroFlow processing to investigate the effects of data expansion on compound potency assessment and mode of action grouping using 30 aneugens/clastogens 95

2.2.5 PROAST 99

2.2.5.1 PROAST for Benchmark Dose Analysis. 99

2.2.5.2 PROAST Parameters Settings 103

2.2.6 Significance of model averaging to generate Confidence Intervals 106

2.2.7 ToxPi for visualisation and data dimensionality reduction and filtering 107

2.2.8 Unsupervised Machine Learning Algorithms: Hierarchical Clustering 114

2.2.9 Unsupervised Machine Learning Algorithms: K-Means Clustering 116

2.3 Results 117

2.4 Discussion 151

2.4.1 BMD CIs High-Low Charts 151

2.4.2 Topoisomerase II inhibitors Outputs 152

2.4.3 10 Clastogens Outputs 153

2.4.3.1 *hclust* functionality and metabolically separated ToxPi Outputs 155

2.4.4 Comprehensive 30 Aneugen/Clastogens Outputs 157

2.4.5 Improvements 158

2.5 Chapter Conclusions 159

Chapter 3: The combinational assessment of *in vitro* Toxtracker assay data with the TGx-DDi biomarker for genotoxicity hazard assessment on eighteen chemicals with positive/negative *in vitro* and *in vivo* cytotoxic calls 161

3.1 Introduction 161

3.1.1 The demand for combined approaches to toxicological for the innovation of new modern hazard assessment pipelines for mechanistic/potency information extraction 161

3.1.2 The Toxtracker Assay 162

3.1.3 TGx-DDI Transcriptomic Biomarker Test 165
3.1.4 Gene expression technologies in the TGx-DDI assay 168
3.1.5 Use of different cell lines for *in vitro* genotoxicity studies 169
3.1.6 The combinational ToxTracker and TGx-DDI testing assay 170
3.2 Materials and methods 173
3.2.1 ToxTracker Dataset curation 173
3.2.2 TGx-DDI Dataset curation 176
3.2.3 Working in PROAST 177
3.2.3.1 ToxTracker 177
3.2.3.2 TGx-DDI 178
3.2.4 ToxPi settings and calibrations 179
3.3 Results 180
3.4 Discussion 194
3.4.1 Pipeline Design Limitations 194
3.4.2 PROAST BMD individual and collective outputs 196
3.4.3 ToxPi Scoring Ranks and Profiles 197
3.4.4 ToxPi Hierarchical/Kmeans Clustering 199
3.4.5 Improvements 201
3.5 Concluding Remarks 202
Chapter 4: Developing Benchmark Dose-Based Analytical Pipelines for the iScreen High-Content Imaging Platform Mechanistic Classification of Aneugenic and Clastogenic Compounds in Genotoxicity Assessment 204
4.1 Introduction 204
4.2 Materials and Methods 208
4.2.1 Cell lines, culturing and treatment 208
4.2.2 Imagining Dataset Curation 209
4.2.3 Working in PROAST 213
4.2.4 ToxPi Data Visualisation 213
4.3 Results 214
4.4 Discussion 233
4.4.1 BMD CI Charts 234
4.4.2 ToxPi Outputs 235
4.4.3 Improvements 237

4.5 Chapter Conclusions 239

Chapter 5: Benchmark Dose and high-throughput-toxicokinetic modelling for *in vitro* to *in vivo* extrapolation (IVIVE) to generate administered equivalent doses (AEDs); a potential alternative to animal testing. 242

5.1 Introduction 242

5.2 Materials and Methods 252

5.2.1 Workflow for Data Handling and Assay Processing 252

5.2.2 PROAST procedure to extract BMDs 258

5.2.3 3compartmentss modelling 258

5.2.4 Css determination using HTTK to generate the AED 263

5.2.5 ToxPi Analysis 264

5.3 Results 265

5.3.1 *In vitro* PROAST Outputs 265

5.3.2 *In vivo* PROAST Outputs 273

5.3.3 Toxicokinetic Parameters 274

5.3.4 *In vitro* datasets extrapolated AED figures 275

5.3.5 *In vivo* datasets extrapolated AED figures 281

5.3.6 *In vitro* ToxPi Outputs 287

5.3.7 *In vivo* ToxPi Outputs 292

5.4 Discussion 297

5.4.1 PROAST BMD Outputs 297

5.4.2 Toxicokinetic Parameters and *in vitro/in vivo* AED Figures 299

5.4.3 ToxPi Figures 300

5.4.4 Improvements 303

5.5 Chapter Conclusions 304

Chapter 6: Thesis Conclusions 307

6.1 General Discussion 307

6.2 Thesis Main Findings 308

6.3 Genetic Toxicology Landscape 317

6.3.1 The BMD approach 317

6.3.2 ToxPi Visualisation 317

6.3.3 Analysis Framework Development 318

6.4 Final Conclusion 319

6.5 Future Research and Development 320
Appendices 321
Glossary 379
References 381

List of Tables

Table 2.1.6.1.2 – A table depicting each topoisomerase II inhibitor assessed and their respective structural classifications.

Table 2.1.6.1.3 – The Derek Nexus – an *in-silico* platform for QSAR – derived structural information derived from Wheeldon et al., (2020) for the 8 topoisomerase II poisons covered in this investigation, showing each compound, a description regarding each of their structural information and their respective skeletal formulae.

Table 2.2.2.1 – A table highlighting the chemicals used in this component of the Wheeldon et al., (2020) study.

Table 2.2.2.2 – This table highlights each flow cytometer device used and their respective fluidics settings (Bryce et al., 2016).

Table 2.2.3.1 – A summary of all 10 clastogens assessed, their respective CAS reference numbers, each MoA and metabolic properties, adapted from the Tian et al., (2020) paper.

Table 2.2.4.1 – The list of 30 compounds tested for the extended assessment of MultiFlow and MicroFlow assay datasets observed in Avlasevich et al., (2021), including their CASRN and MoA.

Table 3.2.1.1 – Table outlining the compounds assessed in both the ToxTracker and TGx-DDI assay.

Table 4.2.1.1 – Chemicals used in this investigation alongside their classifications and tested concentration range adapted from Sun et al., (2022).

Table 4.3.1 – iScreen Endpoint Nomenclature and Biological Interpretation.

Table 5.1 – Definitions for IVIVE and HTTK Abbreviations Table (Pearce et al., 2017).

Table 5.2.1.1 – A list of compounds investigated during this IVIVE chapter including their respective sources and individual MoA.

Table 5.2.2.1 – The xlsx file format adopted for the importation of chemical toxicokinetic parameter within the HTTK database using Bisphenol M and Bisphenol TMC as example chemicals – adapted from the presentation by Beal., (2022).

Table 5.3.2.1 – All the BMDL and BMDU figures produced for the Ki67+ biomarker across all compounds covering all tested days and weeks, where cisplatin, diethylnitrosamine and vinblastine sulphate produced successful BMD CI results, and all columns highlighted “NA” are instances where a BMD was no determined as dose response curve was able to be modelled.

Table 5.3.2.2 – All the BMDL and BMDU figures produced for the MN-RET biomarker across all compounds covering all tested days and weeks, where cisplatin, cyclophosphamide monohydrate, hydroxyurea, melphalan and vinblastine sulphate produced successful BMD CI results, and all columns highlighted “NA” are instances where a BMD was no determined as dose response curve was able to be modelled.

Table 5.3.2.3 – All the BMDL and BMDU figures produced for the MNHEP biomarker across all compounds covering all tested days and weeks, where cisplatin, diethylnitrosamine and vinblastine sulphate produced successful BMD CI results, and all columns highlighted “NA” are instances where

a BMD was no determined as dose response curve was able to be modelled.

Table 5.3.2.4 – All the BMDL and BMDU figures produced for the MUT RET biomarker across all compounds covering all tested days and weeks, where cisplatin, cyclophosphamide monohydrate, diethylnitrosamine, and melphalan produced successful BMD CI results, and all columns highlighted "NA" are instances where a BMD was no determined as dose response curve was able to be modelled.

Table 5.3.3.1 – Table depicting the Toxicokinetic parameters for each compound including their human Fub, Clint, logP, and molecular weight using the canonical SMILES and either the ADMET predictor, HTTK or Health Canada's *in vitro* toxicokinetic data produced in-house. Each compound's Css value was extracted using this TK information and any asterisked (*) parameters are considered unstable as their predictions were out of the domain of applicability for the model.

Table 6.2.1.1 – Summary table showing the assays used, endpoints assessed, chemicals tested and key findings across all chapters of the thesis.

List of Figures

Figure 1.2.1 – An overview of different DNA damaging agents and repair pathways, the type of DNA/genetic material damaged and different agent types that induce each type of DNA damage repair as depicted in the paper by Chatterjee & Walker (2017). The top panel illustrates the DNA damaging agents, the middle panel representing the damaged DNA caused by that agent and subsequently, the lower panel showing the associated DNA repair pathway.

Figure 1.2.2 – An illustration of indirect and direct DNA effect and the types of DNA damage that they can produce, such as ssDNA/dsDNA damage and base/bulky lesions within the DNA molecule (Huang & Zhou, 2021).

Figure 1.3.1 – Four possible dose-response extrapolations associated with toxicological studies producing linear, sublinear, superlinear or sigmoidal shapes to each curve once the threshold dose has been reached (Hassenzahl & Finkel, 2008).

Figure 1.3.2 - The Hill equation formula in terms of tissue response generated from the function, where "E" is the magnitude of response, "Emax" is the maximum response, "EC_{50}"is the Concentration which achieves 50% response, and "A" is the Drug concentration, which is also (Neubig, et al., 2003).

Figure 1.3.3 - The sigmoidal curve generated from the Hill equation where the percentage of maximum response (%) is plotted against an agonist's log drug concentration (nM) to describe dose response behaviour. The Hill coefficient reflects the overall curve shape rather than being a direct measure of the underlying binding mechanism. As a result, similarly shaped curves can arise from systems with multiple independent binding sites. The model also assumes saturable, monotonic relationship between concentration and response with the curves reflecting a steady-state response (Prinz, 2009).

Figure 1.3.4 – A depiction of the intended therapeutic range in relation to the toxicology dose-response range, marking the Margin of Safety (MOS), No Observed Adverse Effect Level (NOAEL), and the effective dose of compounds (Dorato & Engelhardt, 2005).

Figure 1.3.5 – A dose-response curve, which illustrates the benchmark dose (BMD) approach, which is positioned between the lower and upper bounds of the benchmark dose (BMDL and BMDU respectively). The BMDL and BMDU represents the uncertainty surrounding the BMD, which is currently at 10% uncertainty, meaning a 90% CI between the BMDL and BMDU. The benchmark response (BMR) represents a set percentage increase in response above the control (i.e. 50%) which is adjustable depending on the analysis required, which is adjusted to 50-100% (1/2-fold to 1-fold change respectively) above the control for this investigation (Wills et al., 2015). In Wills et al. (2015), this approach was applied to *in vivo* MN data using a series of fitted models alongside the BMR (relative to the background) to estimate the BMD to derive potency rankings and MoA information where the BMDL and BMDU reflect biological variability/experimental uncertainty.

Figure 1.4.1 – A flow chart/other chart showing the first *in vitro* stage and the second in-vivo stage of the hazard of the hazard identification process including the genotoxicity tests involved. Adapted from Committee on Mutagenicity of Chemicals in Food (2011) and accepted genotoxicity testing frameworks, including ICH S2(R1), FDA, EFSA and OECD genotoxicity test guidelines.

Figure 1.5.2.1 – The Histograms (graphs B and E) generated from SYTOX fluorescence are used to quantitatively evaluate cell cycle disruption. The bivariate plots (graphs A & D and C & F), which that are analysed under flow cytometry to quantify percentage EMA-positive events and to quantify the

incidence of nuclei, hypodiploid nuclei and MN using light scatter and fluorescence characteristics respectively (Bryce et al., 2007).

Figure 1.5.3.1 – This diagram reveals the system implemented to generate the Toxtracker assay through extensive whole-genome transcription profiling of mES cells once exposed to 40 different known chemical carcinogens. The genes activated post-exposure were identified to represent DNA damage, oxidative stress and unfolded protein responses being the main pathways identified across six genes that are represented as green fluorescence mES cell reporter lines (Toxys, 2024).

Figure 1.5.6.1 – This diagram illustrates how the integration of toxicokinetic (Plasma Protein Binding with Metabolic Stability) and toxicodynamic parameters enables the generation *in vitro in vivo* extrapolated (IVIVE) figures and estimated target tissue bioactivity concentration values, which contribute to the prediction of the administered equivalent dose (AED) in humans adapted from Rotroff et al. (2010).

Figure 2.2.1.1 – A schematic overview of the analytical workflow adapted from Avlasevich et al., (2021), illustrating the sequential processing steps from the MultiFlow assay data acquisition, the formatting and processing of BMD modelling, multivariate integration of ToxPi for the ranking profiling, and clustering of compounds.

Figure 2.1.6.1.1 – Chakravarti, Radha & Alla (2024) depicting how relevant analogues have increased positive QSAR prediction results and higher relevance values to the original query compound as opposed to the traditional Similar Analogues.

Figure 2.1.6.2.1 – A diagram demonstration the metabolic processing and potential metabolic activation, detoxification and toxicity produced by xenobiotics (Attia., 2010).

Figure 2.2.2.1 – Demonstration of gating process performed during flow cytometric analysis of p53, γH2aX and p-H3 where cells treated with solvent controls are on the left and cells exposed to each test chemical is on the right. Top panel showing γH2aX propidium iodine associated fluorescence of TK6 nuclei when exposed to solvent (left) or Mitomycin C (DNA cross-linking (dsDNA) breaking clastogen) (right). The middle panel showing p53 propidium iodine associated fluorescence of TK6 nuclei treated with solvent (left) or etoposide (Topoisomerase II inhibiting clastogen) (right). Finally, the bottom panel showing p-H3 propidium iodine associated fluorescence when TK6 nuclei were exposed to solvent (left) and griseofulvin (a mitotic spindle tubulin binding aneugen) (right) (Bryce et al., 2016).

Figure 2.2.5.2.1 – An individual dose response curve generated by PROAST v70.3 for Cisplatin S9-Blvrb endpoint in the ToxTracker assay. The raw concentration and GFP induction figures had the exponential models 3 and 5 applied to them, and a user selected CES of 0.5 with 200 model averaging runs. This generated 90% CI BMD values as the BMDL (0.034) and BMDU (0.038).

Figure 2.2.7.1 – The formula capturing the steps involved in calculating the ToxPi score adapted from (Auerbach et al., 2016).

Figure 2.2.7.2 – ToxPi Profile key illustrating the features and properties that produce the visual details for individual chemicals in dose-response datasets.

Figure 2.3.1.1 – A 90% BMD (μM) CI plot for 4-hour γH2AX response in TK6 cells exposed to eight topoisomerase II inhibitors. Ciprofloxacin did not generate a dose-response.

Figure 2.3.1.2 – A 90% BMD (μM) CI plot for 24-hour γH2AX response. Despite a generated BMD for ciprofloxacin at the 24-hour time point, the upper confidence interval is infinite, represented by an asterisk.

Figure 2.3.1.3 – A 90% BMD (μM) CI plot for 4-hour p53 response, measuring dose-response in the MultiFlow assay.

Figure 2.3.1.4 - A 90% BMD (μM) CI plot for 24-hour p53 response, measuring dose-response in the MultiFlow assay.

Figure 2.3.1.5 - A 90% BMD (μM) CI plot for 4-hour p-H3 response, indicates infinite upper bound values, marked by asterisks, for both flumequine and ciprofloxacin at the 4hr p53 response mark.

Figure 2.3.1.6 – A 90% BMD (μM) CI plot for 24-hour p-H3 response, measuring dose-response in the MultiFlow assay, where flumequine's upper bound was considered infinite, highlighted by an asterisk.

Figure 2.3.1.7 – The ToxPi potency ranking results for component/objective 1 illustrating the response of TK6 cells to topoisomerase II inhibitor exposure across 8 compounds. As described in section 2.2.7, the ToxPi scores on the y-axis ranges from 0-1 and represent a normalised set of dataset-dependent values from integrated genotoxicity endpoints. Higher ToxPi scores corresponds to greater relative genotoxic potency within a compound rather than an absolute measure of potency, where the error bars represent the CI of each compound. Therefore, in this dataset, Flumequine exhibited the lowest relative potency and Mitoxantrone has the highest.

Figure 2.3.1.8 – ToxPi profiles for component/objective 1 generated from BMD data for each chemical, showing the biomarkers and their ToxPi scoring for each end point. Each slice represents a biomarker at a given time. Shading in the centre of the circle represents uncertainty for each slice, where darker centres show more uncertainty. The magnitude of each slice (to either the edge or white line at the centre of the CI) represents the relative potency of each biomarker, and widths represent weightings. The brighter regions presented in each is the bootstrap CI. Overall profile size varies according to the normalised ToxPi score; therefore, smaller profiles reflect lower relative compound potency within the dataset rather than a reduced signal at any given endpoint. Additionally, a ToxPi profile key illustrating the colour coordinated arrangement of MultiFlow DNA damage biomarkers at 4hr and 24hr timepoints as slices and the magnitude for each slice represents the fold change values of the median fluorescence intensity generated by propidium iodine flow cytometry, which are modelled into BMD figures that have their CIs -log10(x) transformed to construct the following ToxPi slice profiles accordingly.

Figure 2.3.1.9 – Component 1 hierarchical AGNES clustering of each ToxPi profile generated through the complete clustering algorithm to show grouping of each topoisomerase II inhibitors based on similarity across four distinctive groupings based on their QSAR classifications described in Wheeldon et al., (2020) and Figure 2.1.6.1.3. Anthraquinones representing group 1 (yellow branch), podophyllotoxins for group 2 (purple branch), fluoroquinolones for group 3 (red branch), and isoflavones for group 4 (green branch) with emodin being the only compound that incorrectly clustered with isoflavone (genistein) instead of with the anthraquinones (mitoxantrone and doxorubicin).

Figure 2.3.1.10 – The K -means clustering generated with a user-defined number of clusters set at 4, using divisive algorithms and PCA plot to graphically plot compounds based on their integrated genotoxic biomarker response patterns. These compounds clusters on a PCA basis where more potent

compounds sit on the left (more negative PC1 axis) side, and compounds sat around lower values for the PC2 axis except for Teniposide (higher PC2). Each cluster correlates with the previous hierarchical clusters generated in figure 2.3.1.4, supporting the stability of the observed groupings across analytical methods. However, Teniposide appears to exhibit a larger Euclidian distance to Etoposide, likely due to reduced 4/24hr p53 response projections compared to other compounds and yH2AX and p-H3 reporters.

Figure 2.3.2.1 – BMDs (μM) with 90% CIs were generated using a CES of 0.3 in PROAST v65.5 using exponential models 3 and 5 to fit the dose-response curve for TK6 cells exposed to 10 clastogens. Here, γH2AX response at 4 hours in the absence of S9 metabolic activation was analysed.

Figure 2.3.2.2 – BMDs (μM) with 90% CIs were generated using PROAST for the p53 response at 4 hours in TK6 cells without S9 activation. Higher log10(BMD) values indicate lower potency.

Figure 2.3.2.3 – BMDs (μM) with 90% CIs were generated using PROAST for the γH2AX response at 24 hours in TK6 cells without S9 metabolic activation.

Figure 2.3.2.4 – BMDs (μM) with 90% CIs were generated using PROAST for the p53 response at 24 hours in TK6 cells without S9 activation, where cellular mitochondrial function remained intact whereas the S9 fraction itself contains phase I/II enzymes but lacks intact mitochondria.

Figure 2.3.2.5 – BMDs (μM) with 90% CIs were generated using PROAST for the γH2AX response at 4 hours in TK6 cells with S9 metabolic activation.

Figure 2.3.2.6 – BMDs (μM) with 90% CIs were generated using PROAST for the p53 response at 4 hours in TK6 cells with S9 metabolic activation. Higher log10(BMD) values indicate lower potency.

Figure 2.3.2.7 – BMDs (μM) with 90% CIs were generated using PROAST for the γH2AX response at 24 hours in TK6 cells with S9 metabolic activation.

Figure 2.3.2.8 – BMDs (μM) with 90% CIs were generated using PROAST for the p53 response at 24 hours in TK6 cells with S9 metabolic activation.

Figure 2.3.2.9 – ToxPi ranking for BMD data on 10 clastogens in the absence of S9 overall (1 = lowest potency and 10 = highest potency). ToxPi scores ranging between 0-1 where the higher score translates to higher genotoxic potency, with the calculation being described in section 2.2.7, and the error bars represent the CI of each compound

Figure 2.3.2.10 – oxPi ranking for BMD data on 10 clastogens in the presence of S9 overall (1 = lowest potency and 10 = highest potency). ToxPi scores ranging between 0-1 where the higher score translates to higher genotoxic potency, with the calculation being described in section 2.2.7, and the error bars represent the CI of each compound.

Figure 2.3.2.11 - ToxPi ranks for the 10 clastogens were separated based on the presence/absence of S9, with scores ranging from 0 to 1 produced through the methods given in section 2.2.7, where a higher score indicates greater genotoxic potencies. The error bars represent the confidence interval for each compound.

Figure 2.3.2.12 – ToxPi profile key illustrating each slice with their respective endpoints, including colour coordinated endpoints for with γH2AX and p53 endpoints, at 4/24-hour timepoints with S9 present (S9+) and absent (S9-).

Figure 2.3.2.13 – ToxPi profiles (with bootstrap CI) generated from BMD data for each chemical in component 2. The visualization displays biomarkers and their ToxPi scores for each endpoint. Each

slice represents a biomarker at a specific time point, with and without S9. The shading in the centre of each circle indicates uncertainty, with darker shades signifying greater uncertainty. The radius of each slice, extending from the centre of the circle to either the edge or a white line within the slice, represents the potency of each biomarker, while the width reflects user-defined weightings. The brighter regions indicate the bootstrap confidence interval (CI), with the outer edge showing the upper CI and the inner edge showing the lower CI for each slice.

Figure 2.3.2.14 – Euclidean hierarchical clustering was performed in R using the hclust function with the ward.D algorithm to reveal distinct groupings of each compound based on their S9 potency ratio CI range values, as described in (Wheeldon et al., 2021; Tian et al., 2020). The clustering was based on the AGNES approach in the original study but differences between the S9- and S9+ conditions were performed using a subtractive approach for each endpoint tested: 4-hour γH2AX, 4-hour p53, 24-hour γH2AX, and 24-hour p53. The branch height on the y-axis represents the degree of compound relatedness and is not attributed to any direct measurement unit, other than to describe the degree to similarity between compounds, where smaller heights indicating increased similarity among compounds. In this case, branch height reflects compound similarity based on metabolic-dependent potency differences. The hclust analysis was included to enable direct comparison between clustering derived from raw BMD values and ToxPi clustering based on transformed BMD inputs, thereby assessing the consistency of compound grouping across alternative analytical frameworks

Figure 2.3.2.15 – Complete hierarchical clustering of ToxPi profiles was performed using complete AGNES clustering to identify chemical groupings, each color-coded based on their genotoxic potency in the presence of S9. Green represents compounds with a significant increase in genotoxic potency when S9 is present, red indicates compounds with a moderate increase, yellow denotes no change in potency with the introduction of S9, and purple signifies a decrease in genotoxic capacity due to S9. These groupings correspond to the S9 potency ratio confidence intervals reported by (Wheeldon et al., 2021).

Figure 2.3.2.16 – ToxPi K-means divisive/PCA clustering where each compound is grouped based on variation across the ToxPi scores/genotoxicity and biomarker sensitivity. Each point represents a compound, and each colour/shape represents a cluster they have been placed into. The pattern located is that PC1 compounds with higher ToxPi scores are towards the left, and for PC2, compounds with increased potency with S9+ are at the bottom. Furthermore, the compounds cluster similarly in K-means colours/shapes according to their metabolic properties as described in the hierarchical cluster chart, with exception to mitomycin C and resorcinol: both compounds that have reduced potency with S9 presence. Cluster membership reflects similarity across the full multidimensional ToxPi space; therefore, compounds assigned to the same cluster may appear separated in the two-dimensional PCA projection due to variance captured in higher principal components and the forcing of the clusters based on the user selected K cluster number. While hierarchical provided clearer cluster groupings, the K-means clusters support the presence of broadly related compounds based on Euclidian proximity when evaluated across a 2D space.

Figure 2.3.3.1 – ToxPi ranking for BMD data on 30 aneugens/clastogens (1 = lowest potency and 30 = highest potency). ToxPi scores, which were calculated using the normalisation formulae described in section 2.2.7 ranging between 0-1 where the higher score translates to higher genotoxic potency and the error bars represent the CI of each compound.

Figure 2.3.3.2 – ToxPi profile key illustrating each slice, and their colour coordinated across the MultiFlow/MicroFlow endpoints for both S9± conditions for each biomarker at 4-hour and 24-hour timepoints. Slice colour ordering were selected to reflect mechanistic groupings of genotoxicity

biomarkers: green slices denoting downstream chromosomal damage (24hr MN and 24hr p53), red slices representing DNA lesion/DNA damage response signalling (4-hr p53 and 4-hr/24hr yH2AX) and blue slices indicating mitotic dysregulation associated with aneugenicity (4-hr/24hr p-H3 and 4-hr/24hr poly).

Figure 2.3.3.3 – ToxPi profiles (with bootstrap CI) generated from BMD data for all 30 aneugenic/clastogenic compounds assessed in this component, analysing potency across MultiFlow/MicroFlow endpoints at 4/24-hour timepoints in the presence/absence of S9. Each slice represents a biomarker at a specific time point, both with and without S9. The shading in the centre of each circle denotes uncertainty, with darker shades indicating higher levels of uncertainty. The radius of each slice, from the centre to either the edge or a white line within the slice, illustrates the potency of each biomarker, while the width corresponds to user-defined weightings. The brighter regions represent the bootstrap confidence interval (CI), with the outer edge marking the upper CI and the inner edge indicating the lower CI for each slice.

Figure 2.3.3.4 – Complete hierarchical clustering of ToxPi profiles was performed using complete AGNES clustering to identify chemical groupings, each color-coded based on their biomarker activation and MoA. ToxPi separated the compounds into their known subgroupings: clastogens (Clast), weak clastogens (W Clast), aneugenic tubulin binders (An: TB), aneugenic aurora kinase inhibitors (An: KI), irrelevant positives (IP), and non-genotoxicants (Non-G), with some compounds being identified as irrelevant positives for having unexpected projections as they were known non-genotoxic compounds, which included Brefeldin A and Thapsigargin..

Figure 2.3.3.5 – ToxPi K-means divisive/PCA clustering where each compound is grouped based on variation across the ToxPi scores/genotoxicity and biomarker sensitivity. Unlike previous chapters, as a fixed cluster number was not set, the compounds were instead manually observed, and compounds were grouped based on PC1 and PC2 analyses. PC1 compounds with higher ToxPi scores are towards the right of the graph, however, no such pattern was found for PC2 as more varying endpoints were assessed. Instead, compounds with similar modes of action grouped within closer Euclidian distances from each other, while distant compounds with greatly different biomarker activation ToxPi profiles had greater Euclidian distances between them. Importantly, as DNA damage signalling rather is primarily captured instead of direct mutagenicity, the clustering is better interpreted to reflect similar/differing degrees of genotoxic stress rather than discrete mechanistic classes.

Figure 3.1.2.1 – A representation of the genotoxic/non-genotoxic endpoints covered in the Toxys ToxTracker Assay indicating each of the six GFP reporters with the associated MoA (Hendriks et al., 2024).

Figure 3.1.3.1 – Heatmap of the TGx-DDI biomarker across 64 genes after TK6 cells were exposed to 12 μM etoposide after 4 hours and analysed using Affymetrix whole genome analysis (Allemang et al., 2020). Blue indicates a fold change decrease relative to the control and yellow indicates an increase in fold change. The Etop.0_4 band (representing etoposide) on the right of the heatmap was compared to the fold change profiles 13 known DDI compounds and 16 non-DDI reference compounds to predict whether etoposide at a concentration of 12uM is classified as DDI or non-DDI, where etoposide was classified as DDI.

Figure 3.1.6.1 – NAM-based testing strategy framework flowchart depicting the combination of MicroFlow, TGx-DDI and MultiFlow tests to group compounds into their respective MoA classifications –aneugens, clastogens, mutagens and non-genotoxicants (adapted from (Buick et al., 2021)).

Figure 3.3.1.1 – 90% BMD CIs produced and plotted by PROAST for the Blvrb endpoint, a biomarker for oxidative stress (ROS), across all 18 compounds assessed in the ToxTracker assay, where exponential models 3 and 5 were selected, the concentration was set as the independent variable and the GFP induction was set as the response with the covariate being the S9 conditions (S9+ / S9-). The CES was set to ±0.5 (CED-0.5) with 200 bootstrap runs to produce the BMDs, and graphical plotting of the CIs was formatted to Log10(BMD) on the x-axis with compound names on the y-axis. Compounds that did not return a BMD CI have a "NA" by the compound name, are highlighted in red, and have the '' symbols in the BMD CI response column. This included 5 fluorouracil (S9±), cyclohexanone (S9-), isobutyraldehyde (S9-), methyl methanesulfonate (MMS) (S9+), o-toluidine (S9-), quercetin (S9-), sodium diclofenac (S9-), and sodium saccharin (S9±).

Figure 3.3.1.2 – 90% BMD CIs produced and plotted by PROAST for the Bscl2 endpoint, a reporter of replication stress and ATR mediated DNA damage signalling, across all 18 compounds assessed in the ToxTracker assay, where exponential models 3 and 5 were selected, the concentration was set as the independent variable and the GFP induction was set as the response with the covariate being the S9 conditions (S9+ / S9-). The CES was set to ±0.5 (CED-0.5) with 200 bootstrap runs to produce the BMDs, and graphical plotting of the CIs was formatted to Log10(BMD) on the x-axis with compound names on the y-axis. Compounds that did not return a BMD CI have a "NA" by the compound name, are highlighted in red, and have the '' symbols in the BMD CI response column. This includes amitrol (S9+), cyclohexanone (S9-), ethionamide (S9+), eugenol (S9-), mms (S9+), o-toluidine (S9+), and sodium saccharin (S9-).

Figure 3.3.1.3 – 90% BMD CIs produced and plotted by PROAST for the Btg2 endpoint, a reporter for p53-mediated cell stress response and cell-cycle perturbation, across all 18 compounds assessed in the ToxTracker assay, where exponential models 3 and 5 were selected, the concentration was set as the independent variable and the GFP induction was set as the response with the covariate being the S9 conditions (S9+ / S9-). The CES was set to ±0.5 (CED-0.5) with 200 bootstrap runs to produce the BMDs, and graphical plotting of the CIs was formatted to Log10(BMD) on the x-axis with compound names on the y-axis. Compounds that did not return a BMD CI have a "NA" by the compound name, are highlighted in red, and have the '' symbols in the BMD CI response column. This includes amitrol (S9+), cyclohexanone (S9-), ethionamide (S9+), eugenol (S9-), mms (S9+), o-toluidine (S9±), sodium diclofenac (S9-), and sodium saccharin (S9+).

Figure 3.3.1.4 – 90% BMD CIs produced and plotted by PROAST for the Ddit3 endpoint, a biomarker for protein stress response, across all 18 compounds assessed in the ToxTracker assay, where exponential models 3 and 5 were selected, the concentration was set as the independent variable and the GFP induction was set as the response with the covariate being the S9 conditions (S9+ / S9-). The CES was set to ±0.5 (CED-0.5) with 200 bootstrap runs to produce the BMDs, and graphical plotting of the CIs was formatted to Log10(BMD) on the x-axis with compound names on the y-axis. Compounds that did not return a BMD CI have a "NA" by the compound name, are highlighted in red, and have the '' symbols in the BMD CI response column. This includes 5 fluorouracil (S9±), camptothecin (S9-), eugenol (S9+), isobutyraldehyde (S9-), o-toluidine (S9-), quercetin (S9+), and sodium saccharin (S9-).

Figure 3.3.1.5 – 90% BMD CIs produced and plotted by PROAST for the Rtkn, a reporter of ATM-mediated DNA damage signalling associated with double-strand break responses, endpoint across all 18 compounds assessed in the ToxTracker assay, where exponential models 3 and 5 were selected, the concentration was set as the independent variable and the GFP induction was set as the response with the covariate being the S9 conditions (S9+ / S9-). The CES was set to ±0.5 (CED-0.5) with 200

bootstrap runs to produce the BMDs, and graphical plotting of the CIs was formatted to Log10(BMD) on the x-axis with compound names on the y-axis. Compounds that did not return a BMD CI have a "NA" by the compound name, are highlighted in red, and have the '' symbols in the BMD CI response column. This includes amitrol (S9-), cyclohexanone (S9+), mannitol (S9-), mms (S9+), sodium diclofenac (S9-), and sodium saccharin (S9-).

Figure 3.3.1.6 – 90% BMD CIs produced and plotted by PROAST for the Srxn1 endpoint, a biomarker for oxidative stress (ROS), across all 18 compounds assessed in the ToxTracker assay, where exponential models 3 and 5 were selected, the concentration was set as the independent variable and the GFP induction, was set as the response with the covariate being the S9 conditions (S9+ / S9-) and compounds reflecting "NA" outcomes when no dose-response is produced. The CES was set to ±0.5 (CED-0.5) with 200 bootstrap runs to produce the BMDs, and graphical plotting of the CIs was formatted to Log10(BMD) on the x-axis with compound names on the y-axis Compounds that did not return a BMD CI have a "NA" by the compound name, are highlighted in red, and have the '' symbols in the BMD CI response column. This includes cyclohexanone (S9±), mms (S9+), o-toluidine (S9+), and sodium saccharin (S9+).

Figure 3.3.1.7 – 90% BMD CIs produced and plotted by PROAST for the TGx-DDI assay across all 18 compounds assessed, where exponential models 3 and 5 were selected, the concentration was set as the independent variable and the fold change was set as the response with the covariate being the S9 conditions (S9+ / S9-). The CES was set to ±0.5 (CED-0.5) with 200 bootstrap runs to produce the BMDs, and graphical plotting of the CIs was formatted to Log10(BMD) on the x-axis with compound names on the y-axis. 8/18 compounds had positive calls for and identified as DDI and, vinblastine/colchicine were classified as aneugens, making them undetectable and irrelevant for the TGx-DDI biomarker.

Figure 3.3.2.1 – Graph illustrating the ToxPi ranking derived from BMD data generated in Figures Figure 3.3.1.1 – 3.3.1.7, enabling the comparison of the 18 test compounds (1 = lowest potency and 18 = highest potency). ToxPi scores ranging between 0-1 where the higher score translates to higher genotoxic potency and the error bars represent the CI of each compound.

Figure 3.3.2.2 – ToxPi profile key illustrating each slice with their respective endpoints for each endpoint in the Toxtracker and TGx-DDI Assays including ATM/ATR mediated DNA damage (Bscl2 and Rtkn respectively), p53 mediated cell stress response (Btg2), oxidative stress (Blvrb and Srxn1), protein damage (Ddit3), and DDI (TGx-DDI) in the when metabolically activated (S9+) and inactivated (S9-).

Figure 3.3.2.3 – ToxPi profiles (with bootstrap CI) generated from BMD data for each chemical, showing the biomarkers and their ToxPi scoring for each end point. Each slice represents a biomarker at a given time period in the presence/absence of S9. Shading in the centre of the circle represents uncertainty for each slice (darker = more uncertain). The radius of each slice (from the centre of each circle to the either the edge or white line towards the edge of the slice) represents the potency of each biomarker and the widths the user defined weightings. The brighter regions presented in each is the bootstrap CI where outer edge shows the upper CI, and the inner edge is the lower CI for each slice.

Figure 3.3.2.4 – ToxPi profiles hierarchical clustering generated using complete AGNES clustering on modelled BMDL and BMDU data, each color-coded based on their genotoxic potency in the presence/absence of S9 to show how each of the 18 chemicals group together based on similarity in responses across the ToxTracker and TGx-DDI endpoints. The cluster highlights irrelevant positives

(IP), tubulin binding aneugens (An: TB), incorrectly identified compounds in the toxtracker assay (iiTox) and non-genotoxicants (Non-G).

Figure 3.3.2.5 – ToxPi K-means clusters generated using PCA and divisive clustering algorithms to group compounds on a 2D plane defined by the first (PC1) and second (PC2) principal components, representing greatest sources of variability in the data set using a K-value (K=1 for exploratory purposes to avoid forced partitioning). PC1 reflects overall ToxPi potency, with more potent compounds positioned to the left, while PC2 captures biomarker variation, separating compounds with greater S9− potency (below the axis) from those with greater S9+ potency (above the axis). Compounds with similar potency and biomarker response profiles cluster in closer proximity to each other.

Figure 4.1.1 – Flowchart demonstrating the identification of each compound's classification based on their MoA using the endpoints across the iScreen assay (adapted from (Sun et al., 2022)).

Figure 4.2.2.1 – Induction of MN and cytotoxicity from various compounds in A375 cells with red representing the rising/falling percentage of cytotoxicity with the increasing concentrations, where the y-axis showing the cytotoxicity percentage is on the right side of each graph. Additionally, the blue line represents the fold change increase/decrease in MN induction increase above the control (y-axis on the right) where (a) clastogens (cisplatin, ENU, MMC) (b) tubulin binders (estradiol, noscapine, taxol) (c) aurora kinase inhibitors (AMG900, barasertib, hesperadin) (d) non-genotoxicants (anisomycin, CCCP, tunicamycin). The grey region towards the higher concentrations represented the excessive cytotoxicity region where any concentration exceeding in ≥60% had their results removed from further analysis. Data shown are adapted from Sun et al., (2022).

Figure 4.3.1.1 – Results showing PROAST produced and plotted 90% BMD CIs for AMG900 across 10/13 endpoints that successfully produced BMD CIs with a CES set to ±0.5 and 200 bootstrap runs plotted on a Log10(BMD) x-axis with 4/24-hour endpoints on the y-axis. These endpoints cover MN identification, CENPA presence/absence in cells and γH2AX, MPM2 & H3 biomarker responses. The specific endpoints that did not produce any BMDs include: 24hr H2AXposMN, 4hr H2AXapo, and 4hr MPM2pos. These imply a lack of response in endpoints covering dsDNA damage (clastogens) and mitotic cell activity (tubulin binding aneugens), correlating with the knowledge that AMG900 is an aurora kinase inhibiting aneugen, despite some response produced by a series of γH2aX biomarkers.

Figure 4.3.1.2 – Results showing PROAST produced and plotted 90% BMD CIs for baraserntib across 8/13 endpoints that successfully produced BMD CIs with a CES set to ±0.5 and 200 bootstrap runs plotted on a Log10(BMD) x-axis with 4/24-hour endpoints on the y-axis. These endpoints cover MN identification, CENPA presence/absence in cells and γH2AX, MPM2 & H3 biomarker responses. The specific endpoints that did not produce any BMDs include: 24hr H2AXposMN, 4hr H2AXapo, 4/24h_nuclearH2AX, and 4hr MPM2pos implying a lack of response in dsDNA breaks (dsDNA) and mitotic cell activity (tubulin binding aneugens) biomarkers, indicative of the aneugens aurora kinase inhibiting MoA that baraserntib possesses.

Figure 4.3.1.3 – Results showing PROAST produced and plotted 90% BMD CIs for cisplatin across 12/13 endpoints that successfully produced BMD CIs with a CES set to ±0.5 and 200 bootstrap runs plotted on a Log10(BMD) x-axis with 4/24-hour endpoints on the y-axis. These endpoints cover MN identification, CENPA presence/absence in cells and γH2AX, MPM2 & H3 biomarker responses. The specific 24hr MPM2pos endpoint did not produce a BMD implying a lack of response in mitotic cell activity (tubulin binding aneugens) biomarkers, which is to be expected as cisplatin has a clastogenic

MoA. However, despite the unexpected activated for endpoints p-H3/MPM2/CENPA+, they demonstrated larger BMD CIs and higher BMD values implying lower potencies across those endpoints, meanwhile γH2AX related endpoints see narrow CIs with lower BMD figures suggesting higher activation/potency at those endpoints.

Figure 4.3.1.4 – Results showing PROAST produced and plotted 90% BMD CIs for colchicine across 12/13 endpoints that successfully produced BMD CIs with a CES set to ±0.5 and 200 bootstrap runs plotted on a Log10(BMD) x-axis with 4/24-hour endpoints on the y-axis. These endpoints cover MN identification, CENPA presence/absence in cells and γH2AX, MPM2 & H3 biomarker responses. The 24h_nuclearH2AX endpoint did not produce a BMDs, suggesting that a lower response from γH2AX biomarker within the nuclei, correlating to its primary MoA as a tubulin binding aneugen, however, a response across other γH2AX endpoints remains unexpected. The lower BMD values across most endpoints further imply a higher potency for colchicine.

Figure 4.3.1.5 – Results showing PROAST produced and plotted 90% BMD CIs for dasatinib across all 13 endpoints that successfully produced BMD CIs with a CES set to ±0.5 and 200 bootstrap runs plotted on a Log10(BMD) x-axis with 4/24-hour endpoints on the y-axis. These endpoints cover MN identification, CENPA presence/absence in cells and γH2AX, MPM2 & H3 biomarker responses. As dasatinib clastogenic MoA activates the γH2AX endpoint as expected, it further unexpectedly triggers responses from CENPA, MPM2 & H3 endpoints, likely due to external factors outside of dasatanib's primary MoA activating endpoints such as high potency inducing generalised cell stress producing unrelated endpoint responses. Furthermore, the nuclearH2AX endpoint has a wider CI suggesting more uncertainty surrounding the γH2AX endpoint; a key endpoint for clastogenic MoA.

Figure 4.3.1.6 – Results showing PROAST produced and plotted 90% BMD CIs for ETO across all 13 endpoints that successfully produced BMD CIs with a CES set to ±0.5 and 200 bootstrap runs plotted on a Log10(BMD) x-axis with 4/24-hour endpoints on the y-axis. These endpoints cover MN identification, CENPA presence/absence in cells and γH2AX, MPM2 & H3 biomarker responses. ETO has a known clastogenic MoA activates the γH2AX endpoint as expected, despite the activation of CENPA, MPM2 & H3 endpoints to lesser extents.

Figure 4.3.1.7 – Results showing PROAST produced and plotted 90% BMD CIs for MMC across 12/13 endpoints that successfully produced BMD CIs with a CES set to ±0.5 and 200 bootstrap runs plotted on a Log10(BMD) x-axis with 4/24-hour endpoints on the y-axis. These endpoints cover MN identification, CENPA presence/absence in cells and γH2AX, MPM2 & H3 biomarker responses. The missing endpoint is unexpectedly 24h_H2AXposMN, despite the primary clastogenic MoA of MMC, which may be down to several biological/experimental issues such as compound stability or cell line insensitivity to that endpoint.

Figure 4.3.1.8 – Results showing PROAST produced and plotted 90% BMD CIs for nocodazole across all 13 endpoints that successfully produced BMD CIs with a CES set to ±0.5 and 200 bootstrap runs plotted on a Log10(BMD) x-axis with 4/24-hour endpoints on the y-axis. These endpoints cover MN identification, CENPA presence/absence in cells and γH2AX, MPM2 & H3 biomarker responses. It activates key endpoints for its primary aneugenic tubulin binding MoA, such as CENPA, MPM2 and H3, however, it appears to also activate other unrelated endpoints, such as γH2AX.

Figure 4.3.1.9 – Results showing PROAST produced and plotted 90% BMD CIs for taxol across 11/13 endpoints that successfully produced BMD CIs with a CES set to ±0.5 and 200 bootstrap runs plotted on a Log10(BMD) x-axis with 4/24-hour endpoints on the y-axis. These endpoints cover MN identification, CENPA presence/absence in cells and γH2AX, MPM2 & H3 biomarker responses.

Taxol has a known tubulin binding aneugen and appears to mostly activates endpoints related to its MoA (CENPA, H3 and MPM2), however, other endpoints such as γH2AX are also somewhat activated. Furthermore, 4h_H2AXapo endpoint produced no BMD, however, unexpectedly, the 24h_MPM2pos produced no BMD, which is inconsistent with its MoA.

Figure 4.3.1.10 – Results showing PROAST produced and plotted 90% BMD CIs for tunicamycin across 4/13 endpoints that successfully produced BMD CIs with a CES set to ±0.5 and 200 bootstrap runs plotted on a Log10(BMD) x-axis with 4/24-hour endpoints on the y-axis. These endpoints cover MN identification, CENPA presence/absence in cells and γH2AX, MPM2 & H3 biomarker responses. Tunicamycin is a non-genotoxic compound and appears to activate 24h_MNnegCENPA, 24h_MPM2pos, 4h_MPM2posnuclei and 4h_H3pos endpoints, however, as no BMD is produced for the 24_hr MN endpoint, it is assumed that no micronuclei has formed, suggesting that no chromosomal damage/mis-segregation events had occurred within the cell samples and any BMD readings produced are likely due to certain types of DNA mutations/damage that does not result in major alterations to the chromosomes. Hence, tunicamycin may induce some degree of DNA damage but not any that induced numerical/structural chromosomal damage.

Figure 4.3.2.1 – ToxPi ranking of compounds based on relative genotoxic potency in the iScreen assay (1 = lowest potency, 10 = highest potency). Scores range from 0-1, with higher scores indicating greater potency in this assay. Error bars represent the confidence interval. The ranking reflects performance across the specific endpoint panel used here; for instance, cisplatin scored lowest despite its known human toxicity, highlighting that this ranking is assay-specific and not a measure of overall hazard.

Figure 4.3.2.2 – ToxPi profile key illustrating each biomarker examined in the iScreen assay as an individual slice with corresponding colours. Weightings reflect biological significance as described in section 4.2.4: the MN endpoint received the highest weighting at 3-fold (15.8%) as the primary genotoxicity marker representing the physical manifestation of chromosomal damage or whole chromosome loss. CENPA endpoints received a 2-fold weight (10.5%) , reflecting their critical role in distinguishing aneugenic (centromere-positive) versus clastogenic (centromere-negative) mechanisms. A baseline weight of 1-fold (5.3%) was assigned to the remaining endpoints (γH2AX, MPM2, and H3) to enable specific mechanism of action (MoA) identification while acknowledging their supporting role relative to the core genotoxicity endpoints.

Figure 4.3.2.3 – ToxPi profiles (with bootstrap CI) generated from BMD data for each chemical, showing the biomarkers and their ToxPi scoring for each end point. Each slice represents a biomarker at a given time. Shading in the centre of the circle represents uncertainty for each slice (darker = more uncertain). The radius of each slice (from the centre of each circle to the either the edge or white line towards the edge of the slice) represents the potency of each biomarker and the widths the user defined weightings. The brighter regions presented in each is the bootstrap CI where outer edge shows the upper CI, and the inner edge is the lower CI for each slice.

Figure 4.3.2.4 – Mcquitty hierarchical clustering of ToxPi profiles was performed using complete AGNES clustering to identify chemical groupings, each color-coded based on their biomarker activation and MoA. ToxPi separated the compounds into their known subgroupings: clastogens, weak clastogens, aneugenic tubulin binders, aneugenic aurora kinase inhibitors and non-genotoxicants.

Figure 4.3.2.5 – ToxPi K-means divisive/PCA clustering where each compound is grouped based on variation across the ToxPi scores/genotoxicity and biomarker sensitivity. Unlike previous chapters, as a fixed cluster number was not set, the compounds were instead manually observed, and compounds

were grouped based on PC1 and PC2 analyses. For this analysis, compounds with higher potencies had more negative PC1 values and compounds clustered within proximity to each other similarly to the hierarchical clustering performed in figure 4.3.2.4 explaining why MMC is significantly distanced from other compounds.

Figure 5.1.1 – A depiction of the reverse dosimetry and cross-species extrapolation applied from *in vitro* rodent models to obtain *in vivo* human toxicity figures produced by (Beal 2022 - Appendix).

Figure 5.1.2 – An evaluation of both chemical and genetic pharmacokinetic based toxicokinetic models, demonstrating the increased uncertainty surrounding the generic models as opposed to the chemical-specific model. For situations where chemical specific toxicokinetic data remain unavailable, the genetic model still demonstrates the ability to accurately model *in vivo* data despite the increased uncertainty (Produced by Beal, 2022 and adapted form Cohen Hubal et al., 2018).

Figure 5.1.3 – The depiction of both toxicodynamic and toxicokinetic parameters combining to produce the AED, adapted from Rotroff et al., (2010). Additionally, it further displays the data provided by ToxCast, *in vitro* generated data, registration documentation and computational modelling (IVIVE application).

Figure 5.2.1.2 – Diagram depicting the MutaFlow® Mouse or Rat Blood Thaw and Analysis high throughput protocol within the *pig-a* test performed using the *in vivo* Rat MutaFlow® kit. Blood samples are incubated with anti-CD59-PE to label wild-type erythrocytes (PE-positive) and mutant phenotypes (PE-negative). After adding anti-PE magnetic particles and Counting Beads, a portion is analysed by flow cytometry for pre-column cell-to-bead ratios and reticulocytes. The remaining sample passes through a magnetic column, where mutant cells and beads are collected in the eluate. Post-column eluates are stained and analysed to calculate mutant cell frequencies (Litron Laboratories, 2022).

Figure 5.2.3.1 – The equation for the 3compartmentss steady state model to generate the HTTK predicted Css where the definitions of each component are placed in the table below (Pearce et al., 2017).

Figure 5.2.3.2 – The (A) 1compartment, (B) 3compartment and (C) PBTK modelling diagrams showing flow from the gut lumen to the blood vessels, gut, lungs, the liver, kidney and rest of the body to predict the Css where definitions of each component are placed in the table below (Pearce et al., 2017).

Figure 5.2.3.3 – This graph depicting how the Css (μM) and daily dose (mg/kg/day) can be plotted against each other to generate a linear line, and the gradient (Css ÷ daily dose) is equivalent to the predicted "Css for 1mg/kg/day" (which can also be multiplied to get concentrations for other doses) (Wetmore et al., 2012).

Figure 5.2.3.4 – This graph shows the implementation of reverse dosimetry by swapping the axes, and the equivalent daily dose and Css (*in vitro* AC50) are plotted against each other to generate a slope where the gradient is mg/kg/day per Css for 1mg/kg/day. The final AED is finally found when AC50 ÷ Css (when simulated dose is 1mg/kg/day) (Wetmore et al., 2012). In this chapter, the calculations used are identical with the exception that the *in vitro* AC50 is replaced by the BMD.

Figure 5.3.1.1 – 90% BMD CIs produced using PROAST for the *in vitro* endpoints that successfully produced BMD CIs for cisplatin, which includes 4hr H2aX, 24hr MN, 24hr H2aX, 24hr poly, 24hr p53, and 24hr p-H3, once a CES of ±0.5 is selected with model averaging of 200 bootstrap runs being applied with the nested exponential models 3/5 approach. The log10(BMD) concentration (in μM) sits

on the x-axes, and each compound name on the y-axis where each bar depicts the BMD CI for each compound.

Figure 5.3.1.2 – 90% BMD CIs produced using PROAST for the *in vitro* endpoints that successfully produced BMD CIs for cyclophosphamide monohydrate, which includes 4hr p53, 4hr p-H3, 24hr H2aX, 24hr poly, 24hr MN and 24hr p53, once a CES of ±0.5 is selected with model averaging of 200 bootstrap runs being applied with the nested exponential models 3/5 approach. The log10(BMD) concentration (in µM) sits on the x-axes, and each compound name on the y-axis where each bar depicts the BMD CI for each compound.

Figure 5.3.1.3 – 90% BMD CIs produced using PROAST for the *in vitro* endpoints that successfully produced BMD CIs in diethylnitrosamine, which includes the 4hr H2aX biomarker, once a CES of ±0.5 is selected with model averaging of 200 bootstrap runs being applied with the nested exponential models 3/5 approach. Each individual model has the concentration on the x-axes, and each fold change value for each endpoint on the y-axis, where each BMDL and BMDU displayed under the BMD CI section. As the 4hr H2aX endpoint is the only endpoint produced, no comparative BMD chart is available as only one endpoint successfully modelled a BMD CI.

Figure 5.3.1.4 – 90% BMD CIs produced using PROAST for the *in vitro* endpoints that successfully produced BMD CIs for hydroxyurea, which includes 4hr H2aX, 4hr p-H3, 24hr H2aX, 24hr MN, 24hr p53, 24hr p-H3 and 24hr poly, once a CES of ±0.5 is selected with model averaging of 200 bootstrap runs being applied with the nested exponential models 3/5 approach. The log10(BMD) concentration (in µM) sits on the x-axes, and each compound name on the y-axis where each bar depicts the BMD CI for each compound.

Figure 5.3.1.5 – 90% BMD CIs produced using PROAST for the *in vitro* endpoints that successfully produced BMD CIs for in melphalan, which includes the 4hr H2aX, 4hr p53, 4hr p-H3, 24hr H2aX, 24hr MN, 24hr p53, 24hr p-H3, 24hr poly biomarker, once a CES of ±0.5 is selected with model averaging of 200 bootstrap runs being applied with the nested exponential models 3/5 approach. The log10(BMD) concentration (in µM) sits on the x-axes, and each compound name on the y-axis where each bar depicts the BMD CI for each compound.

Figure 5.3.1.6 – 90% BMD CIs produced using PROAST for the *in vitro* endpoints that successfully produced BMD CIs for in vinblastine sulphate, which includes the 4hr H2aX, 4hr p53, 4hr p-H3, 24hr H2aX, 24hr MN, 24hr p53, 24hr p-H3, 24hr poly biomarker, once a CES of ±0.5 is selected with model averaging of 200 bootstrap runs being applied with the nested exponential models 3/5 approach. The log10(BMD) concentration (in µM) sits on the x-axes, and each compound name on the y-axis where each bar depicts the BMD CI for each compound.

Figure 5.3.4.1 – A graph depicting the *in vitro* AED CIs in mg/kg-bw/day (Log_{10}AED) for cisplatin once 90% BMD CIs are generated using a CES of 0.5 using PROAST (as outlined in section 5.2 and 5.3.1) and each bound (BMDL and BMDU) was divided by the Css figures generated in HTTK using TK parameters to produce AEDL and AEDU values for endpoints 4hr H2aX, 24hr H2aX, 24hr p53, 24hr p-H3, 24hr poly, and 24hr MN.

Figure 5.3.4.2 – A graph depicting the *in vitro* AED CIs in mg/kg-bw/day (Log_{10}AED) for cyclophosphamide monohydrate once 90% BMD CIs are generated using a CES of 0.5 using PROAST (as outlined in section 5.2 and 5.3.1) and each bound (BMDL and BMDU) was divided by the Css figures generated in HTTK using TK parameters to produce AEDL and AEDU values for endpoints 24hr H2aX, 4hr p53, 24hr p53, 4hr p-H3, 24hr poly, and 24hr MN.

Figure 5.3.4.3 – A graph depicting the *in vitro* AED CIs in mg/kg-bw/day (Log_{10}AED) for diethylnitrosamine once 90% BMD CIs are generated using a CES of 0.5 using PROAST (as outlined in section 5.2 and 5.3.1) and each bound (BMDL and BMDU) was divided by the Css figures generated in HTTK using TK parameters to produce AEDL and AEDU values for the 4hr H2aX endpoint.

Figure 5.3.4.4 – A graph depicting the *in vitro* AED CIs in mg/kg-bw/day (Log_{10}AED) for hydroxyurea once 90% BMD CIs are generated using a CES of 0.5 using PROAST (as outlined in section 5.2 and 5.3.1) and each bound (BMDL and BMDU) was divided by the Css figures generated in HTTK using TK parameters to produce AEDL and AEDU values for the endpoints 4hr H2aX, 24hr H2aX, 24hr p53, 4hr p-H3, 24hr poly, and 24hr MN.

Figure 5.3.4.5 – A graph depicting the *in vitro* AED CIs in mg/kg-bw/day (Log_{10}AED) for melphalan once 90% BMD CIs are generated using a CES of 0.5 using PROAST (as outlined in section 5.2 and 5.3.1) and each bound (BMDL and BMDU) was divided by the Css figures generated in HTTK using TK parameters to produce AEDL and AEDU values for the endpoints 4hr H2aX, 24hr H2aX, 4hr p53, 24hr p53, 4hr p-H3, 24hr p-H3 24hr poly, and 24hr MN.

Figure 5.3.4.6 – A graph depicting the *in vitro* AED CIs in mg/kg-bw/day (Log_{10}AED) for vinblastine sulphate once 90% BMD CIs are generated using a CES of 0.5 using PROAST (as outlined in section 5.2 and 5.3.1) and each bound (BMDL and BMDU) was divided by the Css figures generated in HTTK using TK parameters to produce AEDL and AEDU values for the endpoints 4hr H2aX, 24hr H2aX, 4hr p53, 24hr p53, 4hr p-H3, 24hr p-H3 24hr poly, and 24hr MN.

Figure 5.3.5.1 – This graph depicts the *in vivo* AED CIs in mg/kg-bw/day (Log_{10}AED) for cisplatin once 90% BMD CIs are generated using a CES of 0.5 using PROAST (as outlined in section 5.2 and 5.3.1) and each bound (BMDL and BMDU) was divided by the Css figures generated in HTTK using TK parameters to produce AEDL and AEDU values for the endpoints Ki-67+ (3DperWx3W), MN-RET (3DperWx1W, 3DperWx3W and 7DperWx4W), MNHEP (3DperWx1W and 3DperWx3W) and MUT-RET (3DperWx3W and 7DperWx4W).

Figure 5.3.5.2 – This graph depicts the *in vivo* AED CIs in mg/kg-bw/day (Log_{10}AED) for cyclophosphamide monohydrate once 90% BMD CIs are generated using a CES of 0.5 using PROAST (as outlined in section 5.2 and 5.3.1) and each bound (BMDL and BMDU) was divided by the Css figures generated in HTTK using TK parameters to produce AEDL and AEDU values for the endpoints MN-RET (3DperWx1W, 3DperWx3W and 7DperWx4W), and MUT-RET (3DperWx3W and 7DperWx4W).

Figure 5.3.5.3 – This graph depicts the *in vivo* AED CIs in mg/kg-bw/day (Log_{10}AED) for diethylnitrosamine once 90% BMD CIs are generated using a CES of 0.5 using PROAST (as outlined in section 5.2 and 5.3.1) and each bound (BMDL and BMDU) was divided by the Css figures generated in HTTK using TK parameters to produce AEDL and AEDU values for the endpoints Ki-67+ (3DperWx1W, 3DperWx3W, 7DperWx4W), MNHEP (3DperWx1W and 3DperWx3W, 7DperWx4W) and MUT-RET (3DperWx1W, 3DperWx3W).

Figure 5.3.5.4 – This graph depicts the *in vivo* AED CIs in mg/kg-bw/day (Log_{10}AED) for hydroxyurea once 90% BMD CIs are generated using a CES of 0.5 using PROAST (as outlined in section 5.2 and 5.3.1) and each bound (BMDL and BMDU) was divided by the Css figures generated in HTTK using TK parameters to produce AEDL and AEDU values for the endpoints MN-RET (3DperWx1W and 3DperWx3W, 7DperWx4W), and MUT-RET (3DperWx3W, 7DperWx4W).

Figure 5.3.5.5 – This graph depicts the *in vivo* AED CIs in mg/kg-bw/day (Log_{10}AED) for melphalan once 90% BMD CIs are generated using a CES of 0.5 using PROAST (as outlined in section 5.2 and 5.3.1) and each bound (BMDL and BMDU) was divided by the Css figures generated in HTTK using TK parameters to produce AEDL and AEDU values for the endpoint MN-RET (3DperWx1W and 3DperWx3W, 7DperWx4W).

Figure 5.3.5.6 – This graph depicts the *in vivo* AED CIs in mg/kg-bw/day (Log_{10}AED) for vinblastine sulphate once 90% BMD CIs are generated using a CES of 0.5 using PROAST (as outlined in section 5.2 and 5.3.1) and each bound (BMDL and BMDU) was divided by the Css figures generated in HTTK using TK parameters to produce AEDL and AEDU values for the endpoints Ki-67+ (3DperWx1W), MN-RET (3D, 3DperWx3W), and MNHEP (3DperWx1W and 3DperWx3W).

Figure 5.3.6.1 – ToxPi ranking across all 6 compounds tested in this chapter, where the *in vitro* AEDL and AEDU were used by ToxPi across all endpoints (4hr H2aX, 24hr H2aX, 4hr p53, 24hr p53, 4hr p-H3, 24hr p-H3 24hr poly, and 24hr MN) to produce a single ToxPi score utilising the algorithm described in chapter 2.2.7 (all input values were added together, subtracted by the minimum, divided by the rage, multiplied by the weighting value and the component scores were added together and ranged between 0-1), where compounds with a ToxPi score value closer to 1 are considered higher in genotoxic potency and compounds closer to 0 are considered relatively less potent.

Figure 5.3.6.2 – ToxPi profile key with all slices associated with the specific *in vitro* timepoints and biomarkers, including colour coordination for endpoints H2aX, p53, p-H3, polyploidy and MN at both 4hr and 24hr timepoints.

Figure 5.3.6.3 – ToxPi profiles across all tested compounds displaying each overall ToxPi and slice score each using the AEDL and AEDU to produce each endpoint for each all endpoints tested (4hr H2aX, 24hr H2aX, 4hr p53, 24hr p53, 4hr p-H3, 24hr p-H3 24hr poly, and 24hr MN) and -log10(x) transformations are applied to all slices. Additionally, the weighting remained at 1-fold for all endpoints and BMD CIs were calculated using the default 1000 bootstrap runs in ToxPi with the properties of the profile were covered in chapter 2.

Figure 5.3.6.4 – AGNES hierarchical clustering for the *in vitro* IVIVE endpoints using the complete clustering algorithm based on ToxPi profiles produced from the *in vitro* extrapolated AED figures, which produced and identified distinct chemical groupings based on their respective similarities in MoA. Here, four distinct clusters were produced, where: cluster 1 comprises of vinblastine and melphalan, cluster 2 of hydroxyurea, cluster 3 of cisplatin and cluster 4 of cyclophosphamide monohydrate and diethylnitrosamine.

Figure 5.3.6.5 – K-means divisive/PCA clustering produced ToxPi scores, grouping compounds by genotoxicity and biomarker sensitivity. AGNES clustering revealed potency-based groupings: cisplatin, cyclophosphamide monohydrate, and diethylnitrosamine clustered closely, while melphalan grouped with vinblastine sulphate, and hydroxyurea appeared isolated with a lower PC2 value. More potent compounds showed higher PC1 values, while less potent ones had lower PC1 values.

Figure 5.3.7.1 – ToxPi ranking across all 6 compounds tested in this chapter, where the *in vivo* AEDL and AEDU were used by ToxPi across all endpoints Ki67+, MN-RET, MNHEP, and MUT-RET (in conditions 3DperWx1W, 3DperWx3W and 7DperWx4W) to produce a single ToxPi score utilising the algorithm described in chapter 2.2.7 (all input values were added together, subtracted by the minimum, divided by the rage, multiplied by the weighting value and the component scores were added together and ranged between 0-1), where compounds with a ToxPi score value closer to 1 are

considered higher in genotoxic potency and compounds closer to 0 are considered relatively less potent.

Figure 5.3.7.2 – ToxPi profile key with all slices associated with the specific *in vivo* biomarkers, including colour coordination for endpoints Ki67+, MN-RET, MNHEP, and MUT-RET at 3DperWx1W, 3DperWx3W and 7DperWx4W for each endpoint.

Figure 5.3.7.3 – ToxPi profiles across all tested compounds displaying each overall ToxPi and slice score each using the AEDL and AEDU to produce each endpoint for each all endpoints tested (Ki67+, MN-RET, MNHEP, and MUT-RET in either 3DperWx1W, 3DperWx3W and 7DperWx4W) and -log10(x) transformations are applied to all slices. Additionally, the weighting remained at 1-fold for all endpoints and BMD CIs were calculated using the default 1000 bootstrap runs in ToxPi with the properties of the profile were covered in chapter 2. Furthermore, the degree of greyness in the centre of each profile reflects the degree of missing data, where the darker shades represent more missing data for that ToxPi slice/biomarker.

Figure 5.3.7.4 – AGNES hierarchical clustering for the *in vivo* IVIVE endpoints using the complete clustering algorithm based on ToxPi profiles produced from the *in vivo* extrapolated AED figures, which produced and identified distinct chemical groupings based on their respective similarities in MoA. Here, four distinct clusters were produced, where: cluster 1 comprises of cisplatin and melphalan, cluster 2 of hydroxyurea and cyclophosphamide monohydrate, cluster 3 of vinblastine sulphate and cluster 4 of diethylnitrosamine.

Figure 5.3.7.5 – K-means divisive/PCA clustering produced ToxPi scores, grouping compounds by genotoxicity and biomarker sensitivity. Divisive clustering revealed potency-based groupings: hydroxyurea, cyclophosphamide monohydrate, and vinblastine sulphate clustered closely, while the Euclidian distance of all other remaining compounds appear to be larger, separating out the remaining compounds from each other.

Abbreviations

2-Amino-1-Methyl-6-Phenylimidazo [4,5-b] Pyridine (PhIP)

Adenosine Triphosphate (ATP)

Administered Equivalent Doses (AED)

Adverse Outcome Pathways (AOP)

Area Under the Curve (AUC)

Ataxia Telangiectasia and Rad-3 Related (ATR)

Ataxia Telangiectasia Mutated (ATM)

Base Excision Repair (BER)

B-cell Translocation Gene 2 (Btg2)

Benchmark Dose (BMD)

Benchmark Dose Lower Bound Confidence Limit (BMDL)

Benchmark Dose Upper Bound Confidence Limit (BMDU)

Benchmark Response (BMR)

Berardinelli-Seip Congenital Lipodystrophy (Bscl2)

Blood Flow to the Gut (Qgut)

Blood Flow to the Liver (Ql)

Blood Flow to the Liver (Qliver)

Body Weight (BW)

Cardiac Output/Blood Flow Through The Heart and Lungs (Qcardiac)

Centromere Protein A (CENPA)

Chemical Abstract Service Registry Number (CASRN)

Checkpoint Kinase 1 (CHK1)

Checkpoint Kinase 2 (CHK2)

Concentration at 50% Maximum Activity (AC50)

Confidence Interval(s) (CI(s))

Constant Dose Rate (Kdose)

Control Effect Size (CES)

Cyclin-Dependent Kinases (CDKs)

Deoxyribonucleic Acid (DNA)

Dimethyl Sulfoxide (DMSO)

DNA Damage Inducing (DDI)

Double-Stranded DNA (dsDNA)

Emerging Systems Toxicology for the Assessment of Risk (eSTAR)

Fraction Unbound, *in vitro* Ratio of Unbound to Total Concentration in Plasma (Fub)

Glomerular Filtration Rate (GFR)

Glomerular Filtration Rate (Qgfr)

Green Fluorescence Protein (GFP)

Gut Absorption Rate, Default of 1 h-1 (Kgutabs)

Health and Environmental Science Institute Genetic Toxicology Technical Committee (HESI GTTC)

Hepatic Clearance Modelled with Well-Stirred Approximation using Clmetabolism (Clwell-stirred)

High Throughput Toxicokinetic (HTTK)

Intrinsic Hepatic Clearance (Clint)

In-vitro-to-In-vivo Extrapolation (IVIVE)

Micronuclei/Micronucleus (MN)

Mismatch Repair (MMR)

Mode(s) of Action (MoA(s))

Mouse Embryonic Stem Cells (mES)

N-Nitrosodimethylamine (NDMA)

No Observed Adverse Effect Level (NOAEL)

Nuclear Factor Kappa Light Chain Enhancer of Activated B Cells (NF-kB)

Nucleotide Excision Repair (NER)

Organisation for Economic Co-operation and Development (OECD)

One Compartment Model (1compartment)

Phosphate Buffered Saline (PBS)

Phosphatidyl Inositol 3-Kinase-Like Family of Serine/Threonine Protein Kinase (PIKKs)

Phospho-histone H3 (p-H3)

Point of Departure (PoD)

Principal Component 1 (PC1)

Principal Component 2 (PC2)

Principal Component Analysis (PCA)

Quantitative Structural Activity Relationship (QSAR)

Reactive Oxygen Species (ROS)

Relative Increased Nuclei Count (RINC)

Relative Nuclei Count (RNC)

Single-Stranded DNA (ssDNA)

Steady-State Concentration in the Plasma (Css)

Three Compartment Model (3compartment)

Three Compartment Steady State Model (3compartmentss)

Transgenic Rodent Mutation (TGR) assay

Ultraviolet (UV)

Whole Liver Hepatic Clearance Scaled from Clint (Clmetabolism)

Chapter 1 – General Introduction

1.1 – Significance of Genetic Toxicology in Research

Genetic toxicology research studies the effects that chemical and physical agents have on genetic information, including Deoxyribonucleic Acid (DNA), chromosomes, chromatin and associated proteins. Genetic damage is induced by different mechanisms that lead to mutations, which induce permanent changes to DNA, causing genetic alterations that are associated with diseases (e.g., cancer), and impaired cellular function (Ren, Atyah, Chen & Zhou, 2017). However, beneficial genetic changes may also be retained across generations through natural selection, contributing to evolutionary adaptation (Ren, Atyah, Chen & Zhou, 2017). Genotoxic testing models play a crucial role in toxicology and pharmaceutical research, as each technique serves a specific purpose in identifying and analysing potential toxic threats to the human genome, facilitating the detection and characterisation of carcinogens to enable a deeper understanding of their properties (Kim et al., 2013). The scientific and technological advances made through this testing format has enabled pharmaceutical organisations to develop new drug risk assessment methods that evaluate genotoxicity across multiple biomarkers, improving our understanding of DNA and chromosomal damage, identification of genomic hazards and visualisation of the benefit of automating genotoxicity assays (Mahadevan et al., 2011).

In vivo and *in vitro* genotoxicity testing employs techniques to identify compounds that induce genetic damage through DNA mutations, large scale chromosomal aberrations and replication errors that promotes aneuploidy/polyploidy within a genome which has been associated with tumorigenesis (EMA, 2012). These tests detect hazards using our understanding on genetic damage biomarkers, which are used to identify carcinogens and mutagens through positive assay tests results that indicate the presence/absence of genotoxic compounds. For biotechnology and pharmaceutical drug discovery/development, regulatory mandates require that *in vitro* and *in vivo* genotoxicity tests assessments are performed in the preclinical development stage, ensuring that both institutions participate in extensive testing well before a drug reaches commercialisation (Mondal et al., 2010; Snyder & Green, 2001).

In vitro tests, including the Ames test and micronucleus (MN) assay are regulatory testing assays utilised at early drug testing as they provide greater relative throughput, reduce the need for animal models when testing non-genotoxicants while enabling high-throughput

screening (Mondal et al., 2010). A thorough understanding behind the mechanisms underlying genotoxic compounds is highly desirable as positive genotoxicity findings may arise due to excessive cytotoxicity – driven by experimental factors such as chemical saturation, precipitation and non-physiological test concentrations – rather than direct genetic damage; therefore, a vigilant approach must be employed when handling positive genotoxic data (Galloway, 2000; Snyder & Green, 2001). The *in vitro* MN test is significant for chromosomal damage detection from clastogenic/aneugenic compounds as it detects any aneugenic/clastogenic genetic defects through the physical manifestation of MN within the cells where clastogens produce MN with chromosomal fragmentation whilst aneugens produce MN containing whole chromosomes (Fenech, 2000: Organisation for Economic Co-operation and Development (OECD), 2016). Modern assays/analyses allow for mechanistic differences between compounds to be thoroughly understood, making them effective tools for genotoxic hazard assessment, as industry-wide and standardised assays, that follows the OECD guidelines effectively, place chemicals into categories to identify them as genotoxic/non-genotoxic. Additional details surrounding the utility and applicability of each assay used will be covered later in this chapter and continues to lay the foundational framework of this thesis.

1.2 – DNA Damage

Genomic stability ensures essential genetic information responsible for protein synthesis, cell function and signalling pathways is operational and desired characteristics are retained and inherited by the offspring to aid a species' survival (Jackson and Bartek, 2009). However, DNA experiences damage from different exogenous and endogenous stimuli that lead to DNA replication errors, DNA mismatches and oxidative damage (Chatterjee & Walker, 2017; Song et al., 2015). Exogenous DNA damage arises from the interaction of DNA with external physical, chemical or environmental factors including ionising radiation, alkylating agents, and genotoxicants (Chatterjee & Walker, 2017). These mechanisms can induce breakages in single/double stranded DNA (ssDNA/dsDNA) and subsequently alter the intramolecular/intermolecular structures, bonding, and charges in the DNA phosphodiester backbone, hydrogen bonds between DNA bases (Chatterjee & Walker, 2017). Meanwhile, endogenous DNA damage arises from normal metabolic processes where chemically active DNA undergoes hydrolytic reactions or oxidative reactions with reactive oxygen species (ROS) to generate replication errors and DNA mismatches, resulting in lesions which are continuously repaired by cellular mechanisms (Chatterjee & Walker, 2017; Tubbs &

Nussenzweig, 2017). Both sources can induce the production of defective genetic information through mutations, which subsequently promotes faulty protein formation, encourage cell death, stimulate unregulated cell replication, and induce carcinogenesis, affecting the organism's survival and reproduction (Wilson & Hunt, 2002; Song, et al., 2015).

DNA damage is repaired through a series of repair pathways, including specialised systems such as non-homologous end joining and homologous recombination to repair dsDNA damage, while direct DNA repair mechanisms repair damage caused by UV/alkylation. However, DNA repair mechanisms exist in 3 primary pathways: base excision repair (BER), nucleotide excision repair (NER) and mismatch repair (MMR). These mechanisms all involve cleaving/excision of incorrect DNA components with endonucleases/UvrD proteins/exonucleases, correcting any mismatches using DNA polymerases, and re-ligating the DNA molecule with DNA ligase (Hakem, 2008). However, single-base mutations caused by oxidation, alkylation, or deamination trigger base excision repair (BER), while DNA lesions resulting from ultraviolet (UV) radiation, environmental mutagens, or chemotherapeutic agents activate nucleotide excision repair (NER). In contrast, mismatch repair (MMR) operates as a post-replication process, correcting errors that escape DNA polymerase proofreading (Chatterjee & Walker, 2017). These mechanisms are activated upon detection of abnormal genetic information at damage checkpoints to reconstruct DNA as these errors are catastrophic to human health if left unrepaired, usually leading to syndromes and cancers (Hakem, 2008). Figure 1.2.1 below displays the interactions between potential DNA-damaging agents or sources that challenge DNA integrity, the specific types of DNA lesions they inflict, and the specialized DNA repair mechanisms deployed to mitigate their effects.

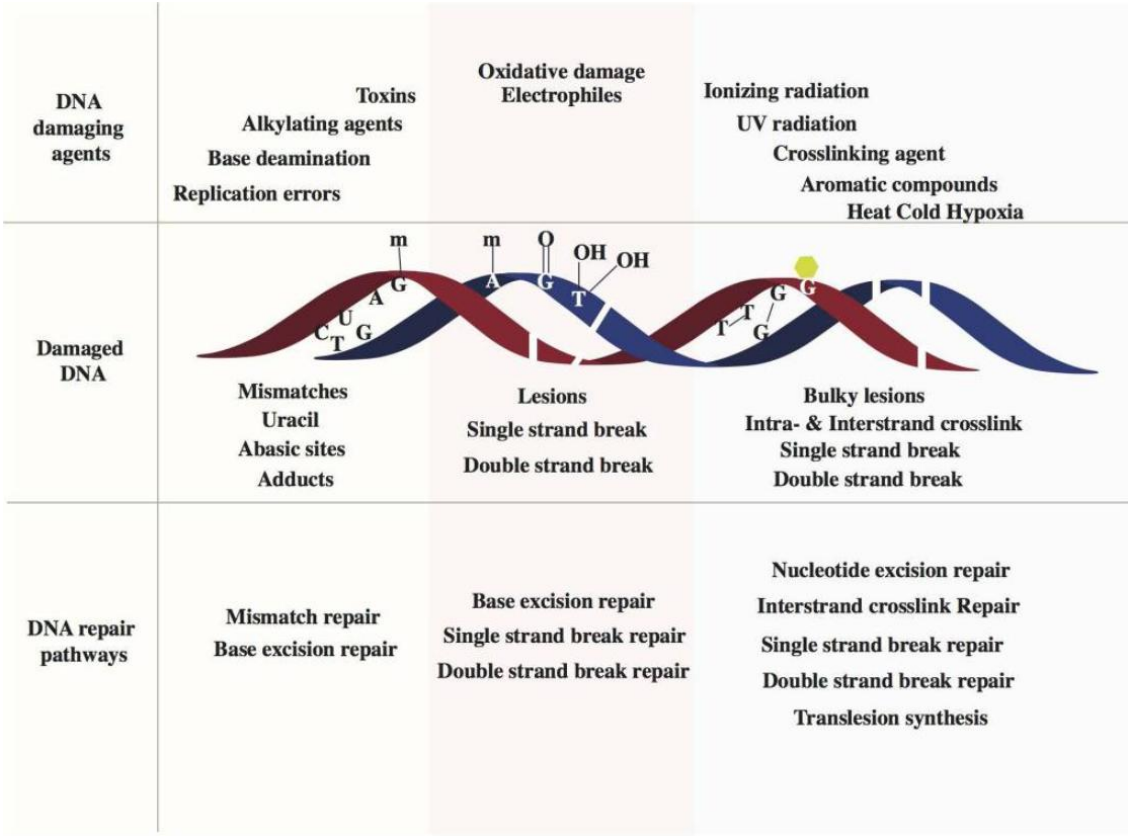


Figure 1.2.1 – An overview of different DNA damaging agents and repair pathways, the type of DNA/genetic material damaged and different agent types that induce each type of DNA damage repair as depicted in the paper by Chatterjee & Walker (2017). The top panel illustrates the DNA damaging agents, the middle panel representing the damaged DNA caused by that agent and subsequently, the lower panel showing the associated DNA repair pathway.

DNA reactive carcinogens, such as alkylating agents and nitrosamines, are genotoxic substances that form electrophilic intermediates that undergo reactions with nucleophilic centres in DNA by forming chemical-specific covalent adducts, where nitrogen and oxygen atoms are targeted within the phosphodiester backbone or bases in the DNA (Barnes et al., 2018). For example, the N7 and O6 positions of guanine, the N3 position of adenine, and the N3 and O2 positions of cytosine are nucleophilic sites where electrophiles form covalent adducts, leading to mutations. The electrophilic strength of a compound is determined by its intrinsic reactivity towards nucleophilic sites. In biological systems, this reactivity determines the nature of the DNA adducts formed which can be measured with sensitive techniques to assess compounds that promote replication arrest and activate DNA repair mechanisms, revealing the nature behind their genotoxicity (Kobets, Iatropoulos & Williams, 2019). DNA reactive carcinogens include polycyclic aromatic hydrocarbons, aromatic amines, and nitrosamines, which undergo nitro-reduction, ring oxidation and conjugation to generate electrophiles that form adducts, resulting in oxidative DNA damage associated with mutations (Carroll, Warnakulasuriyarachchi, Nokhbeh & Lambert, 2002; Ritter et al., 2002; Nagy, Johansson, Zeisig & Möller, 2005). For example, N-nitrosodimethylamine (NDMA), a nitrosamine regularly found in processed foods and angiotensin II drugs, is widely known as a carcinogen to a variety of animal and human tissue as it increased the probability of malignant tumour formation in the stomach, lungs, and liver (Song, et al., 2015). However, DNA reactive carcinogens usually produce more than one type of DNA adduct, making it challenging to identify specific lesions/ mechanisms/ responsible for mutagenesis.

DNA non-reactive carcinogens interact indirectly with DNA by activating other molecular or cellular processes to promote carcinogenesis, including enzyme inhibition, DNA precursor imbalance and reactive oxygen species (ROS) production to disrupt protein synthesis, DNA replication and inducing oxidative stress, producing a genotoxic effect (Kobets, Iatropoulos & Williams, 2019; Dickinson et al., 2004**).** Dichlorodiphenyltrichloroethane, polychlorinated biphenyls, and chlordane are examples of DNA non-reactive compounds that promote liver tumour formation by facilitating the transformation of pre-existing hepatocytes to become cancerous (Williams, 1981; Williams & Numoto, 1984; Anderson et al., 1994). Overall, DNA damage arises through both direct and indirect mechanisms, and the figure below illustrates these mechanisms, highlighting the types of DNA damage they induce as well as base modifications and bulky lesions. Figure 1.2.2 illustrates the differences between DNA

reactive and non-reactive carcinogens highlighting their direct and indirect effects on DNA, respectively, and the types of lesions they produce.

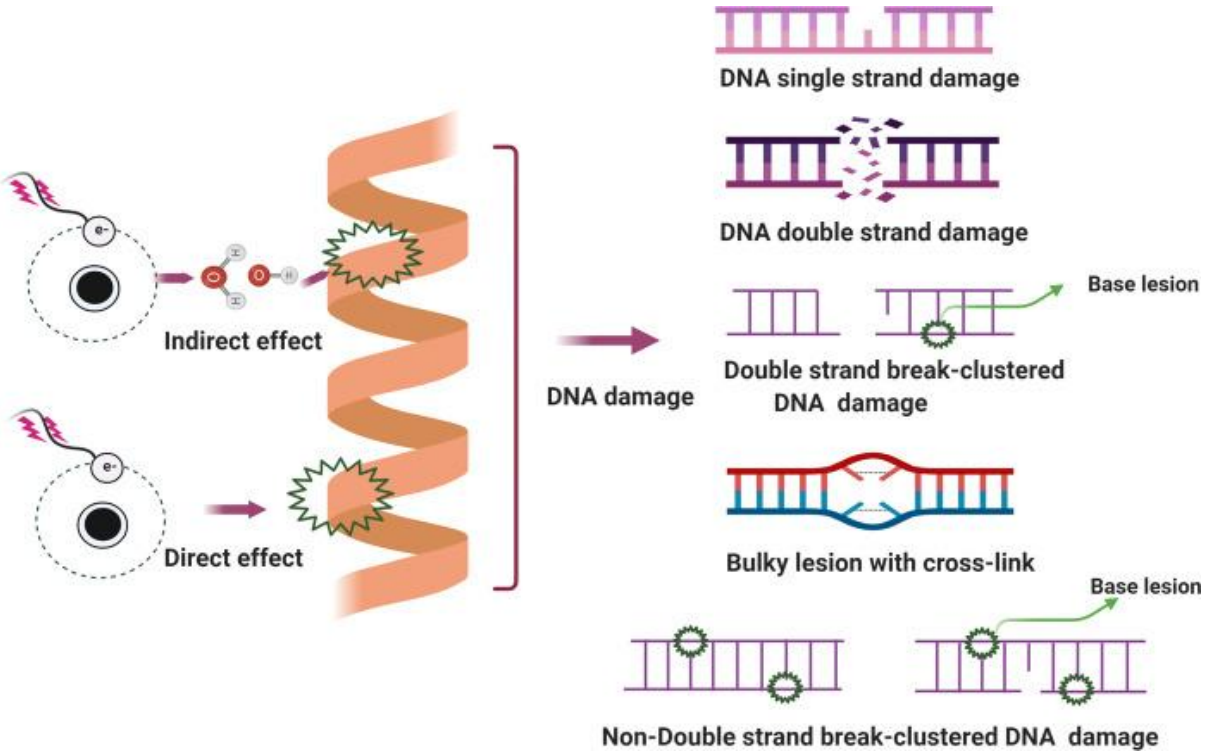


Figure 1.2.2 – An illustration of indirect and direct DNA effect and the types of DNA damage that they can produce, such as ssDNA/dsDNA damage and base/bulky lesions within the DNA molecule (Huang & Zhou, 2021).

1.3 Benchmark Dose-response Model & Dose-response Relationships

Paracelsus quoted in 1538:

"Alle Dinge sind Gift, und nichs ist ohne Gift; allein die Dosis macht, das sein Ding kein Gift ist. All things are poison, and nothing is without poison; the dosage alone make it, so a thing is not a poison."

This adage outlines a basic principle of toxicology, indicating that substances are unable to be exclusively categorised as toxic or non-toxic. Instead, the doses should predominantly determine the lethality of each compound, and they can only be compartmentalised based on the dose required induce toxicity and even chemicals that are deemed to be safe such as oxygen and water are dangerous if an organism is exposed to a high enough dose. The extent of damage can vary based on the dose ingested as the properties and capabilities of all chemicals can change based on the dosage taken.

Toxicology includes intentionally exposing cells, tissue, animals or humans to various chemical agents in controlled concentrations to assess their capacity to cause damage and harm to different cells and organisms. Different toxicological assays provide researchers with key details on toxicity of singular/combinational agents across wide ranges of doses and *in vivo* tests allow for the observation and assessment of chemical exposure. Laboratory exposure levels performed on animals are generally considerably higher than concentrations related to environmental/clinical exposure, therefore, lower dose extrapolation is performed, where the adverse effects of prolonged low-dose exposures are considered. Toxicological studies consider four extrapolatory shapes when it comes to the underlying dose-response curves naturally generated: linear, sublinear, superlinear or sigmoidal shapes as shown in figure 1.3.1 below (Hassenzahl & Finkel, 2008).

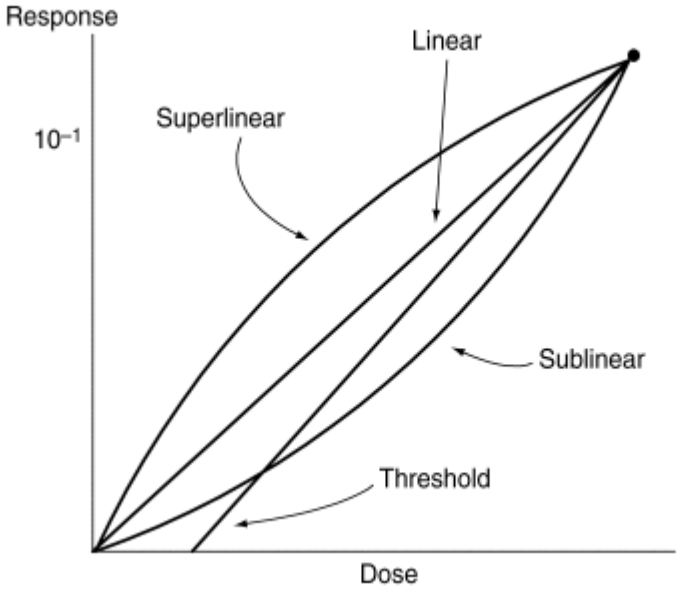


Figure 1.3.1 – Four possible dose-response extrapolations associated with toxicological studies producing linear, sublinear, superlinear or sigmoidal shapes to each curve once the threshold dose has been reached (Hassenzahl & Finkel, 2008).

The four possible dose-response curves illustrate the variability in biological response to toxicants. Linear response indicated a directly proportional relationship, sublinear and superlinear curves suggest that there are changes in dose response deviating from strict proportionality, where sensitivity changes at different doses. The threshold curve reveals a significant rapid increase in dose-response once a particular concentration is reached that rapidly increases response. The general dose-response relationship is normally taken into consideration in genetic toxicological research, especially when regarding to observation of the effects various compounds have on tissues. The dose-response relationship is an important component for pharmacodynamics (the effect and mechanism of drugs to the body) and pharmacokinetics (the drug action upon administration) and it describes the effect drugs have based on the dose provided, generating a curve that follows the Hill equation as it directly reflects the dose-response of different chemicals in biological organisms, which is highly important when assessing response across multiple end-points (Aronson, 2007; Neubig, et al., 2003). The sigmoidal dose-response curve was derived from previous drug potency studies where the log(dose) concentration was plotted against response percentage which directly reflected drug/ligand binding to a specific effect (Weiss, 1997; Neubig, et al., 2003; Prinz, 2009). The hill equation attempts to mathematically model the dose-response relationship in biological systems as a sigmoidal function, which is shown in both figures 1.3.2 and 1.3.3.

$$\frac{E}{E_{MAX}} = \frac{[A]^n}{EC_{50}{}^n + [A]^n} = \frac{1}{1 + \left(\frac{EC_{50}}{[A]}\right)^n}$$

Figure 1.3.2 - The Hill equation formula in terms of tissue response generated from the function, where “E” is the magnitude of response, “Emax” is the maximum response, “EC_{50}“is the Concentration which achieves 50% response, and “A” is the Drug concentration, which is also (Neubig, et al., 2003).

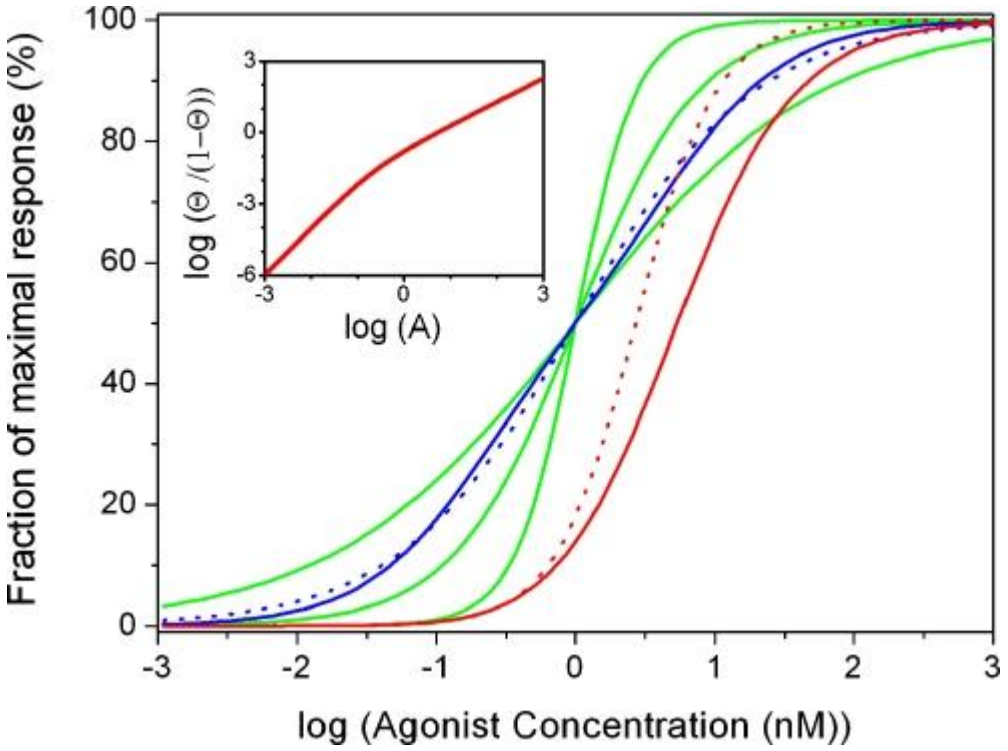


Figure 1.3.3 - The sigmoidal curve generated from the Hill equation where the percentage of maximum response (%) is plotted against an agonist's log drug concentration (nM) to describe dose response behaviour. The Hill coefficient reflects the overall curve shape rather than being a direct measure of the underlying binding mechanism. As a result, similarly shaped curves can arise from systems with multiple independent binding sites. The model also assumes saturable, monotonic relationship between concentration and response with the curves reflecting a steady-state response (Prinz, 2009).

Both models reflect the characteristic relationship between agonist concentration and biological response. This curve highlights the gradual increase in response at low concentrations, followed by a steeper transition near the EC50. At higher concentrations, the response plateaus due to receptor saturation. This provided insight into drug potency, efficacy, and receptor dynamics.

Previous models state that dose-response relationships are linear through the assumption that the doses would be directly proportional to linear DNA damage as it was implied that genotoxic agents can induce DNA damage regardless of dose/concentration (Jenkins, Zaïr, Johnson & Doak, 2010; Bond, Benary, Sondhaus & Feinendegen, 1995). These studies could not obtain threshold values for dose-responses (a point where increasing doses no longer contributes to an increased response), rejecting the idea that dose-response followed a sigmoidal function, as it was originally thought that single point mutations could generate cell proliferation and cancer (Henderson, Albertini & Aardema, 2000; Goldsmith & Kordysh, 1993). However, historically, this linear-no-threshold perspective assumed a binary classification of chemicals as either genotoxic or non-genotoxic, rather than recognising that genotoxic effects may appear at a particular point-of-departure (Jenkins, Zaïr, Johnson & Doak, 2010). Such assumptions ignored humans evolving natural defence systems to manage chronic low-level genotoxicant exposure, meaning DNA damage and dose are not directly proportional (Jenkins, Zaïr, Johnson & Doak, 2010). While very high or single-dose exposure may produce clearly detectable DNA damage, it isn't exactly representative of lifetime/repeated low-dose exposures to chemicals that would the non-linear accumulation of damage due to DNA repair mechanisms, apoptosis and protective barriers. Biological barriers – including epithelial, cellular, and nuclear defences – together with excretion, DNA damage repair pathways (e.g. base excision and nucleotide excision repair), apoptosis, and general DNA redundancy (less than 1% of DNA codes for proteins) are responsible for reducing the likelihood of low-level exposures result in genotoxic outcomes. Combined with variability in genotoxic modes of action (MoAs) and assay detection sensitivity, these factors all contribute to sigmoidal dose-response relationships being observed for genotoxic compounds (Chatterjee & Walker, 2017).

The sigmoidal function was the conventional method used to describe dose-response relationships, measuring asymptotic increases in response where sigmoid curve shape varied due to potency, Emax and EC_{50}; variations resulting from differences in experimental

conditions (Chow, Ong, Dougherty & Simons, 2011). However, establishing and maximising the effectiveness of hazard assessment guidance for chemicals requires the determination of increasingly significant values. Lower concentrations demonstrate low response until the threshold dose (the minimum dose where an observable biological response is produced) has been reached. This demonstrates the importance of the correct responses being mapped at lower doses, as this greatly contributes to the determination of this threshold value.
As concentration increases, more drug molecules bind to more cellular receptors at the site of action, producing a continued exponential increase in response, with higher dose concentrations. This continues until the vMax is met, where response plateaus despite an increasing dose due to all receptors have drug molecules bound to them and have become saturated/activated, so increasing concentration beyond the vMax value cannot produce a higher cellular response. Additionally, a j-shaped dose response describes a biological phenomenon where low dose exposure initially induces a slight reduction in response (a potentially beneficial/stimulatory effect) before a higher dose produces an increase in response leading to harmful/inhibitory effects (Ricci et al., 2011).

The point of departure (PoD) describes the first point/dose where a biological response to a stimulus produces a significant/observable change (Sturla, 2018). Estimating the PoD from dose-response modelling methods required using the No Observed Adverse Effect Level (NOAEL) model, which describes the highest dose that produced no significant statistical and clinical adverse effect, marking the margin of safety (MOS) (U.S. FDA, 2002; Dorato & Engelhardt, 2005). Additionally, the No Observed Genotoxic Effect Level (NOGEL) describes highest dose where no significant genotoxic response is observed , specifically referring to DNA damaging biomarkers. The NOAEL/NOGEL approaches were widely used due to their ease of use and effectiveness in estimating the PoD and can be used when data is unable to suggest a PoD, therefore, it has been used as the standard method for many decades. However, the increase in stronger and more robust PoDs methods have been developed to help locate the true MOS (Dorato & Engelhardt, 2005; Wheeldon et al., 2020). Figure 1.3.4 shows where the MOS is found in relation to the intended therapeutic range and the point where the PoD is crossed, using the NOAEL, and the expected toxicological dose-response range is found, with overt toxicity occurs due to significantly higher doses causing cytotoxicity in cells.

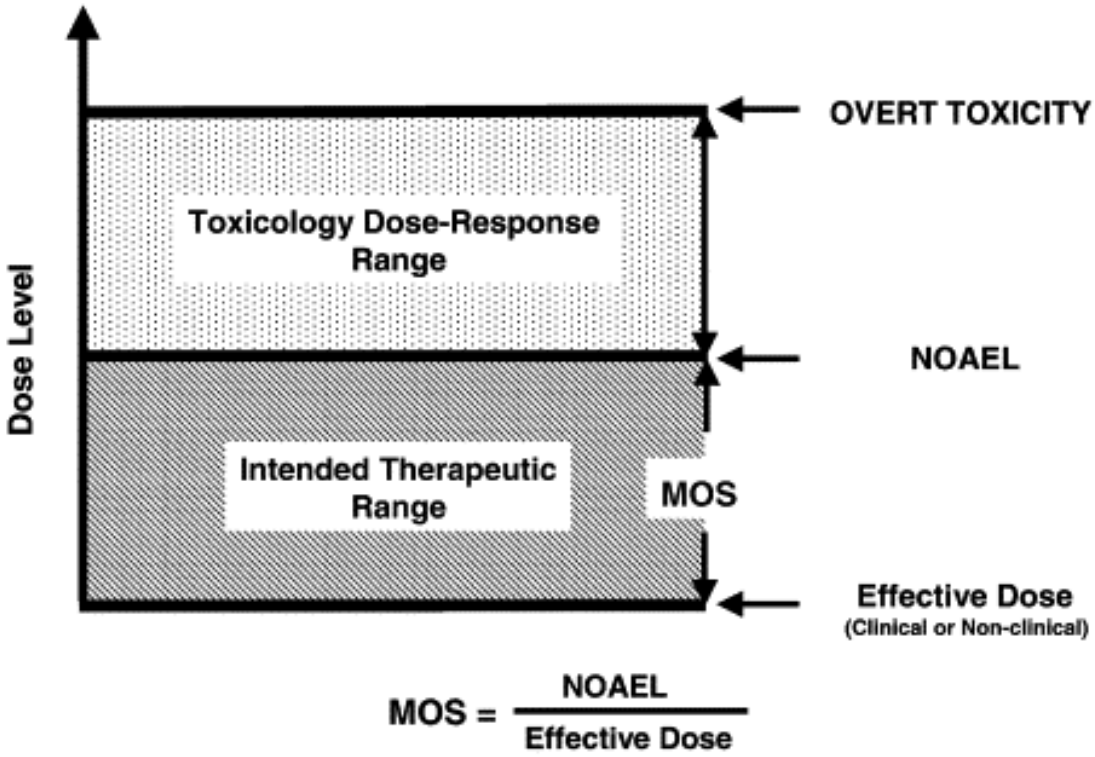


Figure 1.3.4 – A depiction of the intended therapeutic range in relation to the toxicology dose-response range, marking the Margin of Safety (MOS), No Observed Adverse Effect Level (NOAEL), and the effective dose of compounds (Dorato & Engelhardt, 2005).

The NOAEL/NOGEL serves as a robust PoD metric in toxicological potency assessment; however, its applicability is limited due to the traditional assumption that genotoxic effects may occur at any dose instead of occurring at a certain threshold. As a result, alternative models have been developed to better capture human responses across varying compound concentrations. The ability of NOAEL/NOGEL to characterise dose-response relationships heavily relies on a predefined dose selection and spacing, which determine both the shape of the dose-response curve and the identification of the NOAEL/NOGEL. This becomes problematic when experimental design may limit dose ranges/spacing, resulting in weaker dose-response curve modelling, especially at lower concentrations. An alternative to the NOAEL/NOGEL is the benchmark dose (BMD) response model, which describes the dose or concentration that produces a predetermined change in the response rate of an adverse effect caused by a compound (Davis, et al, 2011). It allows for CIs to be estimated which can be used as a guide to estimate a more conservative value for human health guidance as the lower confidence limit (BMDL) is often used when performing drug hazard assessments (Davis, et al., 2011). A small confidence interval (CI) indicates that the BMD is closer to the true value, whereas a larger CI shows that there is increased variation surrounding the true BMD. The shortcomings of the NOAEL/NOGEL models can be mitigated using the BMD approach. For example, the BMD is much less dependent on dose selection, experimental doses/dose spacing compared to NOAEL/NOGEL models. It considers the dose-response curve's shape, has much higher flexibility in determining biologically significant rates, corresponds to a consistent response level and can be utilised to cross compare results across chemicals and studies (Davis, et al., 2011; Rezvanfar, 2014). The goal of dose-response analysis in toxicology is to quantitatively establish exposure thresholds for defined biological or toxicological endpoints to derive a point of departure and determine a MOS, which can be used for hazard characterisation (Ochoa, 2018). The graph below (Figure 1.3.4) demonstrates how the BMD concept is modelled and how the distance between the BMDL and BMDU CIs determines the precision of the BMD, where wider range between the two bounds implies a higher uncertainty surrounding the true location of the BMD.

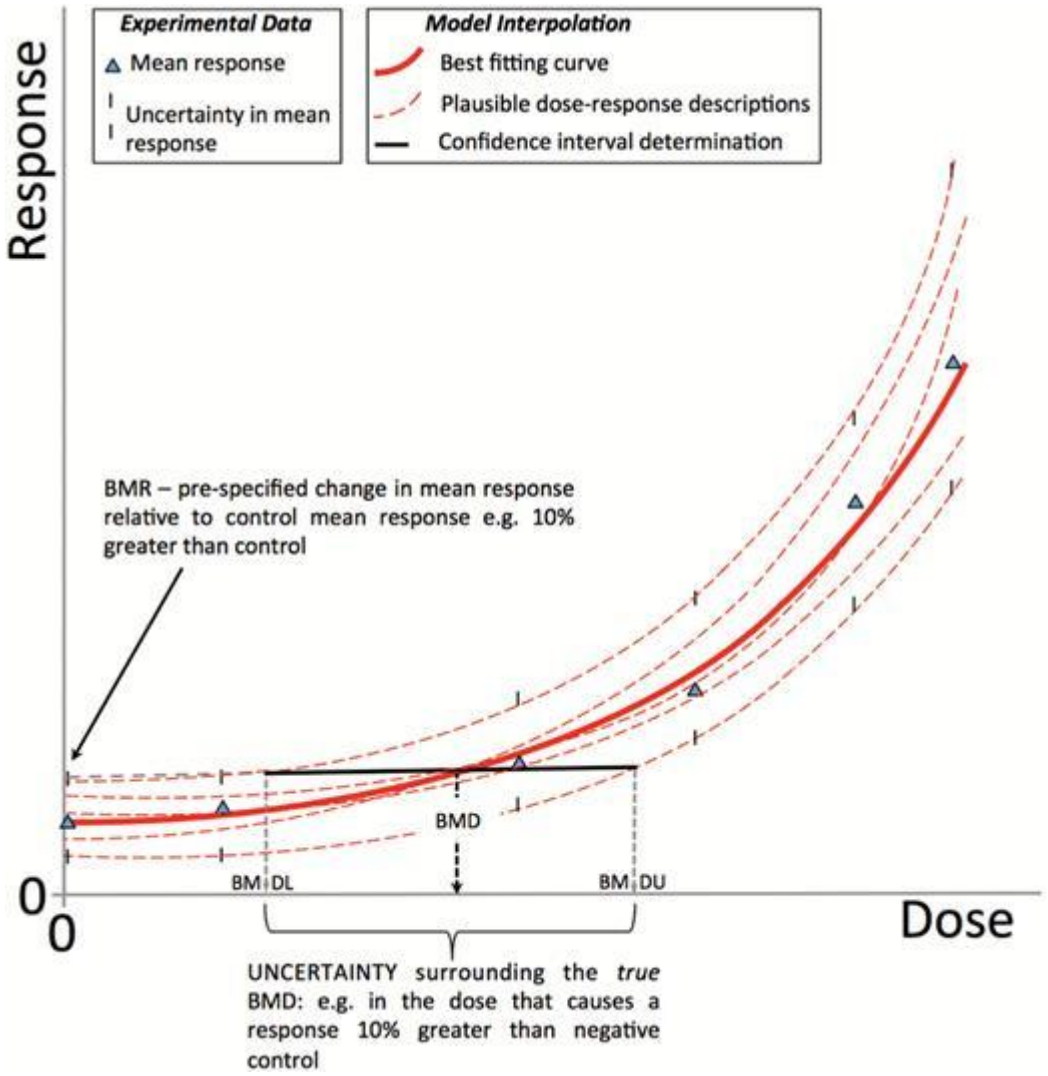


Figure 1.3.5 – A dose-response curve, which illustrates the benchmark dose (BMD) approach, which is positioned between the lower and upper bounds of the benchmark dose (BMDL and BMDU respectively). The BMDL and BMDU represents the uncertainty surrounding the BMD, which is currently at 10% uncertainty, meaning a 90% CI between the BMDL and BMDU. The benchmark response (BMR) represents a set percentage increase in response above the control (i.e. 50%) which is adjustable depending on the analysis required, which is adjusted to 50-100% ($\frac{1}{2}$-fold to 1-fold change respectively) above the control for this investigation (Wills et al., 2015). In Wills et al. (2015), this approach was applied to *in vivo* MN data using a series of fitted models alongside the BMR (relative to the background) to estimate the BMD to derive potency rankings and MoA information where the BMDL and BMDU reflect biological variability/experimental uncertainty.

1.4 Genotoxicity Testing Models for Hazard Assessment

Genotoxicity testing is essential for hazard assessment, as it examines whether chemicals, pharmaceuticals, or environmental pollutants can cause damage to DNA, potentially leading to serious health outcomes, such as cancer, heritable conditions, developmental abnormalities and chronic diseases. An approach to analyse genotoxic hazards involves using *in vitro* assays to assess genetic damage to identify potential genotoxic hazards (Elespuru et al., 2022). *In vivo* testing methods are an alternative method used to assess genotoxic effects of compounds on whole living organisms to get a measure of genetic damage and carcinogenicity to biological systems/environments. These assays take additional variables into consideration by integrating metabolic activation alongside pharmacokinetic and pharmacodynamic processes, including absorption, distribution, metabolism, and excretion (ADME), tissue-specific repair mechanisms and systematic biological responses. As a result, *in vivo* studies capture the biological transparency/complexities of organisms that cannot be fully captured by *in vitro* to support the assessment of genotoxic potential under physiologically relevant conditions (Elespuru et al., 2022). In recent years, *in silico* approaches have gained notoriety for how important they are for genotoxic testing model development. By utilising computer-based/machine learning algorithms these methods can predict genotoxic potential based on chemical structure, reactivity, and molecular interactivity (Benigni, 2021). This aids in prioritisation of substances, ranking, clustering and for further evaluation of compounds and their genotoxic potentials, however, it does not necessarily replace lab-based methods, it instead attempts to provide further information on existing data generated from these empirical approaches. This reduces the quantitative and mechanistic data that is lost in the assessment process whilst increasing reliability and accuracy of the data gathered, particularly during the classification process of chemicals where dose-response information, uncertainty metrics and mechanistic context may be obscured. Combining all these methods develops a refined testing strategy which can bring a deeper comprehensive understanding of potential hazards from chemical compounds and environmental factors that can be potential sources of genotoxicity that would contribute to the development of adverse effects in humans and other organisms.

The Ames, comet and MN assays are well-established genotoxicity testing models that are integral components of toxicological research as these techniques allow for the detection of mutagenic compounds, the establishment toxic doses of chemical agents and the measurements and explanations behind genotoxicity in various chemical agents (Kim et al.,

2013). The MN assay is a popular testing method for genotoxic detection by evaluating the physical manifestations of chromosomal aberrations through micronuclei formation across multiple compounds. Detecting these compounds is a very significant component of medicine as it allows us to understand the compounds that are capable of destabilising and damaging the human genome, which results in the development of various genetic diseases such as cancer and neurodegenerative diseases (Sommer, Buraczewska & Kruszewski, 2020). The identification of the effects in animals that may be predictive of adverse events in humans is the cornerstone of non-clinical safety testing of pharmaceuticals for human therapeutic use.

Previous studies attempted to summarise experimental *in vitro* assay datasets to distinguish between aneugenic and clastogenic compounds. However, these efforts have primarily been constrained through the analysis of individual biomarker readings relevant to clastogenic and aneugenic MoAs, such as analysing direct DNA damage reporters (i.e. yH2AX) and polyploidy. Hence, pharmaceutical industries are currently pushing to using multiple endpoint/mechanistic assays alongside various analytical software programmes to make use of these combined datasets and output diverse MoA assessments surrounding chemicals. Additionally, these assays themselves do not offer very robust dose-response PoD metrics meaning that their currently implemented approaches (i.e. NOAEL, standard deviations, etc) could be significantly improved with a direct dose-response model that simulates human response to genotoxicant exposure and can be applied to multiple datasets to produce reliable PoD figures. This can be done through quantitative dose-response modelling such as BMD modelling approaches, which fit a series of mathematical models to dose-response data to estimate PoDs whilst incorporating uncertainty using BMD confidence intervals (CIs). Applying this across a series of reporters could improve the interpretability and consistency of *in vitro* genotoxicity data, enhancing their utility for potency evaluation, MoA characterisation, chemical categorisation, and read-across applications.

The *in vitro* standard test battery for genotoxicity is a procedure comprised of a set of complementary assays conducted for the regulatory determination of a compound's genotoxicity (Kirkland et al., 2005). These tests include the Ames bacterial reverse mutation assay, the *in vitro* mouse lymphoma assay, and the *in vitro* chromosome aberration assay (MN assay), which all measure chromosomal aberrations, increasing the sensitivity and predictability for carcinogenicity detection in rodents (ICH S2, 2012). The ICH S2 guidelines recommend a combination of the bacterial reverse mutation assay (for detecting rodent

carcinogens) and *in vitro* mammalian cytogenetic tests (such as the mouse lymphoma assay or micronucleus assay) to enhance predictability and sensitivity for identifying carcinogens in rodents. This approach broadens the spectrum of detectable genetic events but may reduce specificity (Baldrick, 2021; ICH S2, 2012). Although the low specificity of these tests may result in misleading false positives, no single test can detect all genotoxic mechanisms, so the battery approach is best by broadening the spectrum of genetic events and endpoints detected (Kirkland et al., 2007; Kirkland, Kasper, Muller, Corvi, & Speit, 2008). At stage 1, a compound testing negative is determined to be non-genotoxic and then each chemical is cross-referenced with existing literature on chemical structure/MoA and the results from other tests (*in vitro* mammalian tests). A negative result at stage 1 would indicate a compound is non-genotoxic whereas a positive result would subsequently trigger stage 2 *in vivo* testing which outlines "Option A" (targeted *in vivo* testing based on endpoints observation) or "Option B" (weight of evidence approach considering structural alerts, MoA and prior data) to confirm genotoxicity.

In vitro and *in vivo* testing models have differences between their kinetic and toxic properties and compounds may be mutagenic *in vitro* but not *in vivo* and testing methods that takes absorption, distribution, metabolism, and excretion into consideration should be assessed to solidify a weight of evidence approach (ICH S2, 2012). The rationale for the use of the second phase of the battery is that mutagenic endpoints were found in the first stage and aneugenicity/clastogenicity needs to be determined (Chemicals in Food, 2011). Expert judgements for tumour/tissue-based evidence and increased/prolonged exposure being required are also justifications commencing *in vivo* testing phase (ICH S2, 2012). Assays used include the rodent bone marrow/peripheral blood chromosomal aberration assay, the comet assay (*in vivo* predictive hazard testing) to measure dsDNA breaks and the transgenic rodent mutation (TGR) assay, which locates and quantifies DNA mutations, detecting a wide range of DNA damage mechanisms (ICH S2, 2012; OECD, 2025). The specificity for *in vivo* assays can be higher than the *in vitro* assays due to their increased biological context and reducing the presence of false positives from indirect toxicity. However, sensitivity can vary as some genotoxic compounds may still be missed if the exposure conditions are insufficient. Indirect toxicity effects can hence influence dsDNA breaks (comet assays), physiological changes to the animal including body temperature and erythropoiesis can increase MN count and bring false positives to unrelated compounds, further complicating the interpretation of genotoxicity datasets (Leist, 2012; ICH S2, 2012; Tweats, et al., 2007; Buick et al., 2021).

This emphasises the need for a comprehensive biologically relevant weight-of-evidence approach which accounts for these confounding factors to ensure accurate genotoxicity assessment. Figure 1.4.1 below displays a flowchart that outlines this two-stage testing battery for the hazard identification process and highlights the key genotoxicity tests used at each stage, including the high-throughput *in vitro* screening methods and the subsequent physiological *in vivo* characterisation testing stages. This system outlined however, only applies to cosmetics as pharmaceuticals, agriculture, foods, and other substances will be tested *in vitro* and *in vivo* regardless of outcomes to obtain a comprehensive understanding of a compound's genotoxicity.

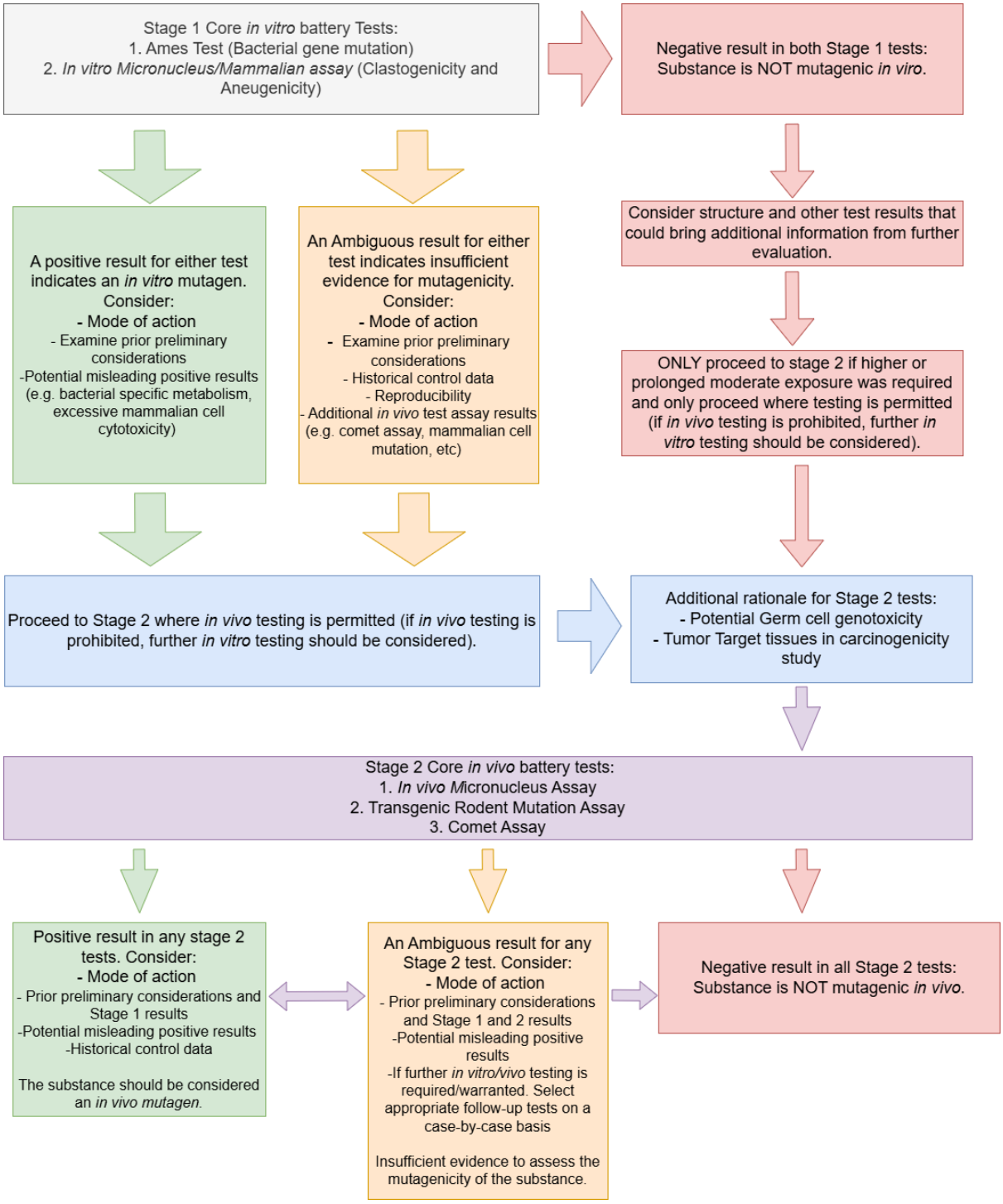


Figure 1.4.1 – A flow chart/other chart showing the first *in vitro* stage and the second *in-vivo* stage of the hazard of the hazard identification process including the genotoxicity tests involved. Adapted from Committee on Mutagenicity of Chemicals in Food (2011) and accepted genotoxicity testing frameworks, including ICH S2(R1), FDA, EFSA and OECD genotoxicity test guidelines.

The *in vitro* battery might have higher sensitivity to genotoxic compounds at the cost of specificity resulting in higher percentages of false positives when compared to the *in vivo* test battery (Kirkland et al., 2007; Kirkland, Kasper, Muller, Corvi, & Speit, 2008). Following up with *in vivo* tests can confirm the genotoxicity of the positive stage 1 compounds, involving the MN assay, comet assay and transgenic rodent mutation assay (Chemicals in food, 2011). Positive results indicate a mutagenic compound's presence and negative results means the compound is non-genotoxic. However, animal welfare concerns have placed regulations on *in vivo* testing meaning a reduction in its use, preventing the maximum compound information from being extracted and the loss of cosmetic products (Allemang et al., 2020). Improvements in *in vitro* genotoxicity predictability tests are required to make up for this lost data, reducing animal use while protecting human safety in compound hazard identification.

The application of a standard test battery has revealed complications in distinguishing MoA. DNA reactive genotoxicants are electrophilic and interact directly with DNA while non-DNA reactive genotoxicants induce genotoxicity through indirect mechanisms (Sasaki et al., 2019). These differences in MoA complicate the use of a standard testing battery for hazard assessment. While many directly acting carcinogens are assumed to follow a linear extrapolation model, where effects can be projected linearly to zero and no safe dose is presumed, some agents, such as MMS and EMS, exhibit threshold dose-responses, indicating a dose below which no significant effect occurs (Kirsch-Volders, Aardema, & Elhajouji, 2000; Lovell, 2000; Nohmi, 2018). Meanwhile, indirectly acting carcinogens act through other mechanisms that inherently hold thresholds, where genotoxicity only presents once that threshold is broken. Below this threshold, cellular tolerance and repair mechanisms set in place to effectively mitigate damage, enabling the establishment of a biologically relevant safe dose (Allemang et al., 2020). However, current *in vitro* OECD guidelines are constrained to detecting mutagenic/clastogenic potential of chemicals but do not provide mechanistic insight into whether genotoxicity arises from direct/indirect DNA interactive processes. As a result, critical information on tolerability, threshold behaviour and MoA, which are essential for hazard identification/risk assessment, remains inaccessible unless assay protocols are modified (e.g. through the inclusion of ROS scavengers).

1.5 Modern Genotoxicity Testing Systems

1.5.1 Multiflow *In vitro* DNA Damage Assay for Modern Genotoxicity Testing

Chromosomal damage arising from DNA damage are critical initiating events in carcinogenesis and represent the mechanistic basis of genotoxicity (Fenech, 2000). The *in vitro* MN assay (following OECD guidelines) is a gold standard testing method developed for the detection of MN within a cell's cytoplasm during interphase (Terradas, Martín and Genescà, 2016). When cells are exposed to mutagens/carcinogens, the presence of MN reveals that an exposure to the tested compound has resulted in the physical manifestation of DNA damage and chromosomal instability due to the presence of MN in cells (Kirsch-Volders et al., 2010). Therefore, it is a simple, accurate and versatile assay for predicting and detecting mutagenic and carcinogenic effects caused by different compounds across various cell types. It is excellent at mapping post-exposure effects induced by compounds, however, to classify them as either clastogenic or aneugenic requires additional techniques such as fluorescence *in situ* hybridisation, to reveal the nature of mutagenicity exhibited by chemicals (Fenech, 2007; Kirsch-Volders et al., 2010). MN visualisation and scoring is performed with analysis techniques, such as flow cytometry, to allow for cytogenetic damage analysis across different chemicals (Bryce, Bemis, Avlasevich and Dertinger, 2007). Flow cytometry allows for the relative nuclei counts (RNCs) to be determined, providing information about treatment cytotoxicity and cell stasis. Cytotoxicity measurement is critical because it ensures that observed genotoxic effects are not confounded by excessive cell death, which could lead to false positives or misinterpretation of results. Additionally, understanding cytotoxicity helps establish dose ranges for further testing, ensuring that concentrations used are biologically relevant and not artificially toxic, distinguishing between true genotoxic effects and secondary effects caused by cell death or stress (Kirkland et al., 2005). This gives the MN assay flexibility and adaptability in its use across various industrial purposes as the use of flow cytometry has a simple methodology that allows for automation for the detection of multiple mutagenic, carcinogenic, and epigenetic endpoints relatively quickly compared to alternative methodologies (Kirsch-Volders et al., 2010).

The MultiFlow assay detect DNA damage response pathways and aneuploidy production through the combination of biomarker information extraction and single flow cytometric analysis for multiplexed datasets (Bryce et al., 2017). The three endpoints used for this include p53, γH2aX and p-H3. The nuclear p53 content activation is typically evaluated as a primary reporter, as p53 represents an early and sensitive marker of DNA damage response. It is measured in free nuclei providing nucleus-specific readout, enhancing the specificity of this endpoint. The γH2aX (serine 139) is an early step in cell DNA damage response by

contributing towards nucleosome formation and chromatin remodelling to repair DNA when double strand DNA damage is detected (2nd endpoint). The phosphorylation of histone H3 (at serine 10) p-H3 endpoint is an indicator of mitotic cell activity and is generated during metaphase when H3 is phosphorylated during mitosis/meiosis to condense chromosomes and progress the cell cycle (3rd endpoint). Aneugenic compounds induce aneuploidy/monosomy by targeting non-DNA components of the mitotic apparatus leading to numerical chromosome aberrations in daughter cells. Through a centromeric MoA, whole chromosomes/chromatids fail to segregate properly, resulting in multinucleated cells and the formation of centromere-positive central micronuclei. These effects are reflected in alterations of the p-H3 biomarker (Beedanagari, 2017; Wilde, Queisser, Holz, Raschke, & Sutter, 2018; More et al., 2021). In contrast, clastogenic compounds induce structural chromosomal damage by generating dsDNA breaks that produce a mixture of chromosome fragments that lack centromeric sequences. This acentric MoA results in centromere-negative (non-central) MN which is associated with the activation of the yH2AX biomarker, a reporter that marks dsDNA damage (Beedanagari, 2017; Wilde, Queisser, Holz, Raschke, & Sutter, 2018; More et al., 2021). This allows the MultiFlow assays to discover new aneugenic/clastogenic compounds by employing this detection system and when incorporated with the MicroFlow assay, allows for MN and polyploidy to be examined, allowing for the physical manifestation of genotoxicity to be observed.

1.5.2 – Microflow assay datasets

MicroFlow assay kit developed by Litron Laboratories US was utilised in data components of chapter 1 when assessing Microflow/Multiflow assay data generated in the Avlasenvich et al., (2021) article, looking at larger datasets. The Microflow assay involves the incubation of mammalian cell cultures to quantitatively score MN formation in human lymphoblastoid (TK6) or mouse lymphoma (L5178Y) cells using flow cytometric techniques. The advantage to using the *in vitro* Microflow method compared to other automated microscopic scoring methods is the use of sequential staining that differentiates MN and chromatin of mid/late stage apoptotic and necrotic cells, meaning a reliable MN measurement can be extracted even in the presence of these apoptotic/necrotic cells (Avlasevich et al., 2006 & Bryce et al., 2007). This means key events linked to MN manifestation, a hallmark of genotoxicity, can be accurately mapped, allowing for reliable MN measurements that distinguish genotoxic effects from background responses and enable precise quantification, improving the reliability and interpretability of the data. Ethidium monoazide (EMA) fluorescently dyes cells by entering

their compromised membranes and covalently bonding to their DNA, following a photoactivation step, to identify both apoptotic/necrotic cells. Once identified, the cells are washed, stripped of their cytoplasmic membranes, and incubated in SYTOX Green (RNase and pan-nucleic acid) dye (Avlasevich et al., 2011). This provides a suspension where free nuclei and MN (EMA-/CYTOX+) that are differentially strained from the chromatin associated with dead/dying cells (EMA+/CYTOX+), allowing us to reliably discriminate between nuclei/MN in normal cells and necrotic/apoptotic chromatin in fragmented cells (Bryce et al., 2007; Bryce et al., 2013 & Avlasevich et al., 2011). 6 µM fluorescent beds were used to acquire cell-to-counting beads and nuclei-to-counting beads ratios through the protocol described in Bryce et al., (2013), as a measure of relative cell survival rate. Flow cytometric scoring of MN allows for SYTOX-associated fluorescence to discriminate between different phases of the cell cycle allowing for histograms and bivariate plots of TK6 cells to be produced to quantify percentages of EMA+ events and highlight these cells cycle events (Bryce et al., 2007). Figure 1.5.2.1 illustrates this.

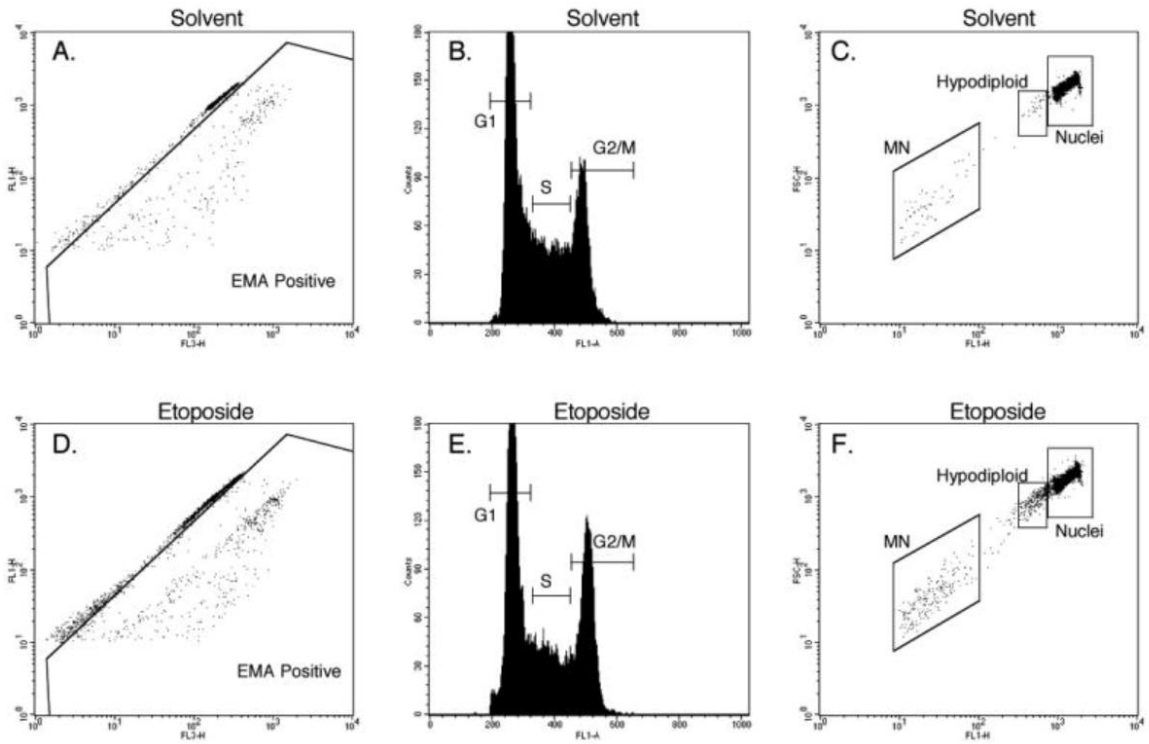


Figure 1.5.2.1 – The Histograms (graphs B and E) generated from SYTOX fluorescence are used to quantitatively evaluate cell cycle disruption. The bivariate plots (graphs A & D and C & F), which that are analysed under flow cytometry to quantify percentage EMA-positive events and to quantify the incidence of nuclei, hypodiploid nuclei and MN using light scatter and fluorescence characteristics respectively (Bryce et al., 2007).

24-hours post treatment exposure, the cells were processed for MN frequencies using flow cytometric analysis using the *in vitro* Microflow® Kit reagents (Litron Laboratories, Rochester, NY, USA) and followed the protocols highlighted in (Bryce et al., 2013; Avlasevich et al., 2011 & Litron Laboratories., 2018). The flow cytometry component was carried out using a BD FASCANTO™ II flow cytometer equipped with a BD High Throughput sampler device, stock photomultiplier tube detectors and optical filters set were used to detect fluorescence emissions associated with the MicroFlow kit provided. SYTOX green detected in the FITC channel and ethidium monoazide in the EMA PerCP-Cy5.5 channel (Avlasevich et al., 2021). The MN analysis is restricted to concentrations which resulted in less than 60% reduction in relative increased nuclei count (lowest precipitation concentration or 10mM, whichever is lowest), and the MN% was converted to fold change values relative to the solvent control (S9± as separate readings) (Avlasevich et al., 2021). Treatment induced MN formation was considered significant under 2 conditions: the fold change was higher than 2.5x over the concurrent mean solvent control, and a mean MN% that exceeded an upper bound tolerance interval was calculated from historical negative control distribution (S9± as separate readings) (Avlasevich et al., 2021). The MN assay was assessed in isolation and in a combinational assay to determine an aneugenic/clastogenic MoA prediction.

1.5.3 Toxys ToxTracker Assay

The Ames test, mammalian mutation test and chromosomal mutation assays have high sensitivity, making them excellent for mutation detection, however, they lack specificity and fail to provide the desired insight into the genetic mechanisms that destabilise DNA. This limits the improvement on safety assessments of novel chemicals, which are developed for a wide range of applications. It is crucial that any adverse effects are detected beforehand so chemicals/materials that negatively affect cells by damaging the DNA and further inducing tissue damage/cancer can be readily identified. Quantitative assessment of specific cell stress response pathways upon xenobiotic/chemical exposure would provide further details to the type/extent of cellular damage induction and biological reactivity of materials. For this assay, a cell system is required where all these response pathways remain unimpaired: the most suited for this being the mouse embryonic stem (mES) cells. In these circumstances, mES cells are generally considered more stable and proficient in the pathways needed for the accurate detection of carcinogenic properties, particularly due to their in-tact cell cycle

checkpoints, functional DNA damage response pathways and DNA repair capacity compared to TK6/HepG2 derived cell lines (Hendriks et al., 2011). Additionally, mES cell lines exhibit a very high proliferation rate and heightened sensitivity to DNA damage as it is theorised that cancer originated within stem cells, which possess long replicative lifespans and self-renewal capacity, allowing the accumulation of propagation of oncogenic mutations (Hendriks et al., 2011). The Toxys B.V. ToxTracker is a single mammalian stem cell-based reporter assay that was designed to readily identify genotoxic/carcinogenic properties of new pharmaceutical compounds for diligent potency assessment and evaluation, where potentially toxic compounds are detected with information about their respective MoA. In combination with MACSQuant X flow cytometer meter, the ToxTracker assay analysed six mES green fluorescence protein (GFP) reporters that predominantly discriminated between oxidative stress induction, heat shock/unfolding protein response and p53 cell stress response. The activity of cancer-associated stress responses was assessed by analysing the activation of a panel of biomarker genes in mES cells exposed to various compounds. These activated biomarker genes were strongly associated with DNA damage, antioxidant and unfolded protein responses when exposed to carcinogens. Increased ROS levels, particularly from mitochondrial dysfunction or a xenobiotic derived reactive metabolite, can subsequently lead to oxidative DNA damage due to faulty SOS response leading to tumor progression. The srxn1 encodes the sulfiredoxin-1 protein, and is a known Nrf2 target, which defends cellular oxidative stress against by reducing oxidised cysteines in peroxiredoxins in peroxisomes (Hendriks et al 2011; Chang et al 2004). Nrf2 is a nuclear transcription factor that remains bound to Keap1 protein under normal conditions, in the cytoplasm, but dissociates with Keap1 when ROS levels rise, where Nrf2 is translocated to the nucleus and associates with antioxidant response elements promoter genes to produce detoxification enzymes and antioxidants (Niture et al., 2010). Biliverdin reductase B is expressed by the Blvrb gene, shows an important role in heme metabolism through its association with antioxidant response, however, the Blvrb promoter sequence does not contain an Nrf2 binding motif (Smith et al 2008). DNA damaging agents that disrupt microtubule formation are also linked with the p53 and Nrf2 cell signalling response (Hendriks et al 2016).

DNA damage response is mediated by several major signalling pathways: ataxia telangiectasia mutated (ATM) and the ataxia telangiectasia and Rad-3 related (ATR) pathways, both members of the phosphatidyl inositol 3- kinase-like family of serine/threonine protein kinases (PIKKs). Cell cycle checkpoint activation and DNA damage

repair systems are activated by both these kinase pathways where ATM is recruited when protomutagenic DNA lesions that are detected, phosphorylating the checkpoint kinase 2 (CHK2) and ATR is recruited when at each replication fork when DNA replication stalls, phosphorylating checkpoint kinase 1 (CHK1) (Reinhardt and Yaffe 2009). Both CHK1 and CHK2 inhibit CDC25 phosphatase and p53 tumour suppression by blocking cyclin-dependent kinases (CDKs) from activating, inhibiting the cell cycle at the G1/S phase, and inducing cellular apoptosis, preventing further development of the genetically damaged cells (Smith et al 2010; Meek, 2009). The Berardinelli-Seip congenital lipodystrophy (Bscl2) gene encodes for the SEIPIN protein, which accumulates in the endoplasmic reticulum, associates with lipids, and is linked to Berardinelli-Seip congenital lipodystrophy, a condition marked by impaired adipocyte differentiation (Magre et al 2001; Szymanski et al 2007). In mES cells, Bscl2 responds to DNA damage associated with the ATR-CHK1 signalling pathway, in response to inhibited DNA synthesis (Hendriks et al 2016). Meanwhile, the Rtkn gene encodes for Rho-GTPase proteins which regulates cell morphology, motility, and cytokinesis (Ito, Morishita, & Nagata, 2018). Rtkn is associated with the ATM-CHK2 signalling pathway, where dsDNA breaks are detected, and links with the NF-kB (nuclear factor kappa light chain enhancer of activated B cells) cytokine signalling pathway to prevent further cellular proliferation and survival, leading to apoptosis in response to the DNA damage (Sabatel et al 2011). Rtkn activation is strongly linked to aneugenic compounds, where Hendriks et al. (2016) explains how aneugenic mitotic spindle fibre poisons (colcemid, nocodazole, vinorelbine and Taxol) activated the NF-kB pathway; linked to the Rtkn gene. Meanwhile, in the same study, Bscl2 hardly induced any response after 24-hours, so it appears that Rtkn and Bscl2 can distinguish between aneugenic and clastogenic MoA.

Upregulation of the unfolded protein response, triggered by misfolded proteins and resulting in reduced protein production, apoptosis, and tumour formation, drives cancer development under cellular stress. The upregulation of heat shock proteins, triggered by stressors like heat, UV, or cold, may promote tumour formation by causing protein misfolding and structural alterations in secondary and tertiary protein interactions. However, heat shock protein expression is generally considered a non-genotoxic stress pathway rather than an indicator of DNA damage. Unfolded protein response alleviates perturbations caused by heat shock proteins/unfolded protein damage resulting in endoplasmic reticulum (ER) stress. This ER stress results in the dissociation of the glucose-related protein 78 (GRP78) protein from PERK, IRE and ATF6, and GRP78 binds to the unfolded protein, degrading the cell through

cell cycle arrest/apoptosis through the PERK, IRE and ATF6 signalling pathways through cell cycle arrest, cell apoptosis and ER stress response. The DNA damage-inducible transcript 3 (DDit3) monitors activation of this unfolded/misfolded protein response. The B-cell translocation gene 2 (Btg2) reporter is associated with p53 response to DNA damage, oxidative stress, or metabolic stress affiliated with cytotoxicity, and apoptosis (Meek 2009). Btg2 is involved in the regulation of the G1/S cell cycle checkpoint (p53 response) and is induced by various cellular stresses (Winkler 2010).

The six ToxTracker reporter genes effectively distinguish between different compounds and whether they elicit distinct DNA damage response and cellular stress pathways, including oxidative stress and protein damage. All cell lines are GFP-tagged, and upon exposure to a genotoxic compound, DNA damage response activation is assessed using MACSQuant X flow cytometry. The 24-hour GFP fold-change reveals the chemical's MoA after fluorescent biomarkers are analysed and scored. As stated earlier, the ToxTracker assay also detects for aneugenic and clastogenic MoA and it can detect clastogens by measuring the DNA damage biomarkers while aneugenic compounds will measure the dsDNA breaks, however for aurora kinase inhibitors, they do not activate any ToxTracker markers, despite their nature to induce aneuploidy (Miltenyi Biotec & Toxys 2019). The ToxTracker Assay servers as an innovative tool for genotoxic testing, assessing multiple endpoints through the integration of reporter lines analysis and computational flow cytometric imaging techniques to encapsulate a potential hazard’s true genotoxicity. This can be used to further safeguard human health and safety by supporting future regulatory decision making on older/newly manufactured pharmaceutical products. It greatly contributes to the refinement of hazard assessment methodologies, and it will be effectively utilised to reliably distinguish between potential genotoxic compounds in this investigation. Developing reliable *in vitro* assays requires a deep understanding of cellular stress responses to toxicants. Figure1.5.3.1 illustrates the system used to establish the ToxTracker assay, highlighting key genes associated with DNA damage, oxidative stress, and unfolded protein responses, which are represented as green fluorescence reporter lines.

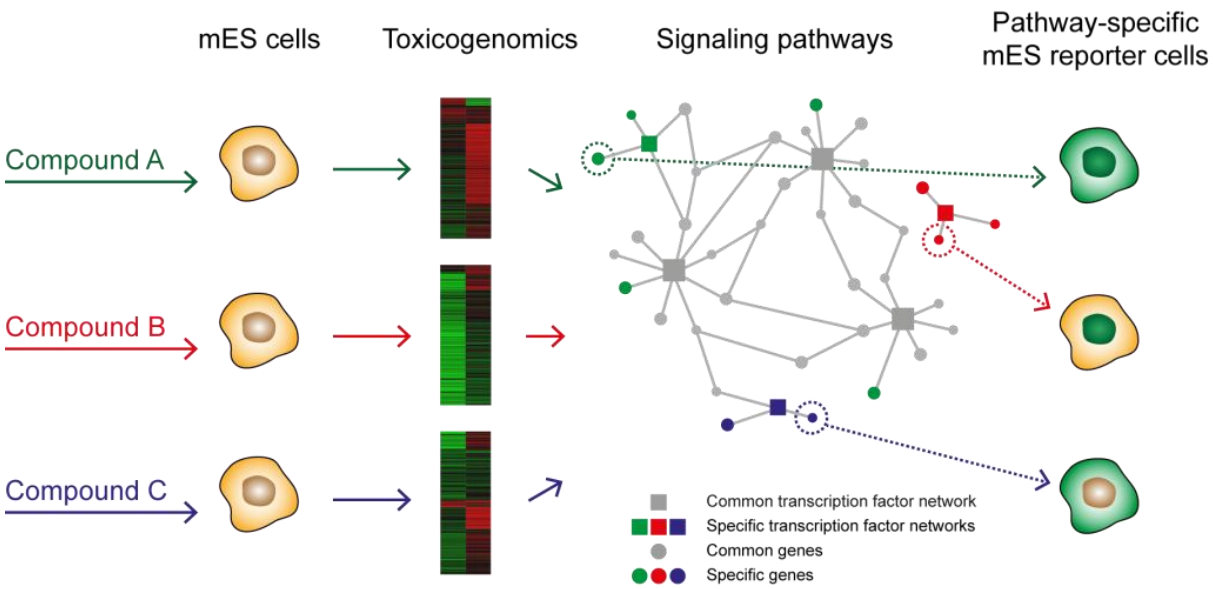


Figure 1.5.3.1 – This diagram reveals the system implemented to generate the Toxtracker assay through extensive whole-genome transcription profiling of mES cells once exposed to 40 different known chemical carcinogens. The genes activated post-exposure were identified to represent DNA damage, oxidative stress and unfolded protein responses being the main pathways identified across six genes that are represented as green fluorescence mES cell reporter lines (Toxys, 2024).

1.5.4 HESI GTTC TGx-DDi Transcriptomic Biomarker Test

Transcriptomic approaches utilise gene expression patterns to infer DNA damage response activation from chemical exposure. Typically, alterations to genetics are an important hallmark of potential DNA damage and drive cancer development in cell systems. Conducting higher-throughput tests offers significant advantages by providing comprehensive insights into gene perturbations caused by chemical exposure. Such approaches are invaluable for regulatory agencies in hazard characterisation, as they facilitate the identification of adverse outcome pathways (AOPs) and a deeper understanding of the genotoxic MoA of a chemical. The mechanistic association between chemicals, their ability to induce direct DNA damage and its carcinogenic effects on humans is complex due to non-genotoxic mechanisms of carcinogenesis and limitations of the *vitro* chromosomal damage assays (Li et al., 2019). Transcriptomic approaches address these challenges by enabling the identification characteristic gene expression signatures triggered by DNA damage, which can be used to detect and mechanistically classify genotoxicants beyond traditional reporter assays.

The *in vitro* chromosomal damage assays detect structural chromosomal abnormalities induced by test compounds, however, compounds that test positive may not show any carcinogenic toxicity to humans and animals (Li et al., 2019). Two methods include characterising the relevance of the positive results followed by additional *in vitro* and *in vivo* studies that detect similar endpoints in combination or using two *in vitro* tests as an alternative to *in vitro* chromosomal damage assays (Rendon et al., 2011). However, there are calls on the reduction of animal use in these testing methods, therefore, *in vitro* testing methods that use relevant human cell systems/genes that model tissue-like behaviour are very important for the development of hazard assessment strategies for human and animal safety (Li et al., 2017). Applying transcriptomic testing methods to identify DNA damage response would alleviate the issue by allowing *in vitro* methods to effectively assess genotoxic compounds. DNA damage network is initiated by DNA damage sensors and signalling pathways to appropriately manage any potential cell damage/cancer development brought on by DNA damage inducing (DDI) agents (Buick et al., 2021; Li et al., 2019). These signalling pathways activate transcription of DDI response genes, which are involved in cell cycle control, DNA damage repair and cell apoptosis. Therefore, the Health and Environmental Science Institute Genetic Toxicology Technical Committee (HESI GTTC) group had

developed and validated the TGx-DDI transcriptomic biomarker test for emerging systems toxicology for the assessment of risk (eSTAR).

The HESI GTTC group had developed the TGx-DDI biomarker model from a training set of global gene profiles from human TK6 cells exposed to 28 prototype DNA reactive agents that are known to be DDI or non DDI chemicals. The model contained 13 known DDI and 15 known non-DDI compounds (aneugens were classified as non-DDI compounds due to their ability to disrupt spindle apparatus formation) (Li et al., 2015).

Post-exposure, a machine learning approach was used to pull out genes that were consistently altered. A dose optimisation protocol was performed using machine learning and quantitative real time-PCR (qRT-PCR) of stress response genes (which would show a 1.7-fold change) followed by DNA microarrays (Agilent) analysis, which used a nearest shrunken centroid method, to locate 64-gene DDI signature probes (Li et al., 2015; Li et al., 2019). These procedures identified 64 DDI genes by assessing consistencies in gene activation data across the 28 known DDI compounds. The nearest shrunken centroid method determines the probability based on the signature that the gene/chemical is either DDI or non-DDI (P (0.9)). Principal component analysis (PCA) analyses with the original training set and 2D clustering was used in the three-pronged method approach and 42 additional chemicals were tested for verification and the TGx-DDI protocol was developed and the 64 DDI genes validated (Li et al., 2015; Li et al., 2019). The biomarker comprises of these 64 genes where changes in the expression of these genes classify these chemicals as DDI and non-DDI using gene expression technologies (DNA microarrays, qRT-PCR, high throughput Nanostring nCounter methods and RNA sequencing (RNA-seq)) in human TK6 cells (Li et al., 2019). The TGx-DDI biomarker has strengths in detecting clastogenic MoA *in vitro* as DDI compounds could inflict DNA damage through deletion, insertion, and substitution of chromosomal sections (Li et al., 2019).

1.5.5 The iScreen High-Content Imaging Platform

The industry standard assessment of aneugenic and clastogenic compounds can be performed using the *in vitro* MN test (following OECD, 2016 guidelines). To effectively perform the *in vitro* MN test, a battery approach is required to maintain ability of predictability and specificity. Although this approach increases sensitivity, it is labour-intensive as microscopic

scoring does not discriminate between different MoAs and have higher costs than a hypothetical single chromosomal aberration assay. The manual scoring methodologies were traditionally used; however, it was more time consuming and did not meet assessment the demands for larger compound lists and extracting mechanistic information from chromosomes. Automated flow cytometry and image analysis systems can be implemented circumvent these issues and provide increased chromosomal structural and numerical information, allowing for higher throughput approaches for quantitative dose-response assessment. Cells exposed to compounds can make use of fluorescently tagged multiplexed biomarkers to accurately classify large lists of chemicals across different genotoxic MoA, without directly assessing chromosomal damage or mutations.

Pharmaceutical companies are currently developing comprehensive, high content multiplexed imaging platform alternatives to flow cytometry, one of these being iScreen currently undergoing development by Pfizer. It incorporates MN identification, centromere detection in MN and multiple DNA damage response biomarkers based on existing AOP (Sun et al., 2022). A centromere protein A (CENPA) antibody-stained kinetochore the regions so that MN containing full/fragmented chromosomes could be identified. Next, the dsDNA breaks endpoints used to detect and confirm any clastogenic MoA were the γH2aX high population induction and spots/cell biomarkers. In addition, the MPM-2 and phosphorhistone-H3 endpoints were used as mitotic biomarkers to measure increases in the mitotic index to observe tubulin binding compounds (Gollapudi et al., 2014). Meanwhile, polyploidy induction and phosphorhistone-H3 dephosphorylation, often markers used to detect aurora kinase inhibitors, were also measured as normally aurora kinases phosphorylate histone H3 during mitosis (Gollapudi et al., 2014). The cell line selected for iScreen is the A375 melenoma cell line as its origin, wild type p53, contains a stable MN background frequency and shows rapid growth. The combination these endpoints takes inspiration from the Multiflow assay and allows for the comprehensive assessment of compounds allowing for aneugenic and clastogenic compounds to have their MoA identified. The new platform is an alternative approach for genotoxicity screening and providing MoA information that can be used to clinically assess drug developments in the future, although its applicability, relevance and sensitivity may vary across therapeutic classes. For example, targeted agents such as MEK kinase inhibitors as they are less likely to produce strong enough signal for reliable detection.

1.5.6 *In vitro In vivo* Extrapolatory modelling

In vitro methodologies for assessing genotoxic potential are invaluable for advancing drug hazard assessment. The procedures discussed exemplify how *in vitro* studies enhance our understanding of compound toxicity and expand toxicological knowledge. Although *in vitro* methods are simple, cost-effective, and high throughput, their validity is limited as they fail to replicate the complexities of organisms, including metabolism, biochemical processes, and cellular interactions (Bhogal et al., 2005; OECD 2018). This makes them an imprecise representation of real-world compound toxicity and potency, particularly for predicting *in vivo* outcomes, assessing systemic effects, or identifying organ-specific toxicity. *In vitro* studies, while cost-effective, reliable, and efficient for drug analysis, have limitations in translating directly to human-equivalent values (Beal et al., 2022). These methods often fail to account for the complexity of human biology, including the coexistence of healthy and diseased cells, the limited metabolic complexity and systematic interactions between cells/organs/tissues both acutely and chronically, and the lack of genetic/epigenetic diversity, which can all influence drug and dose response behaviour (Bhogal et al., 2005; OECD 2018). This can lead to underdosing or overdosing when relying solely on *in vitro* data, as they lack the physiological and pathological variability present in humans (Hartung, 2009).

In vivo studies address these shortcomings, so extrapolatory methods combining both *in vitro* and *in vivo* methodologies would provide the best system for compound potency assessment (Wetmore et al., 2018). However, *in vivo* methodologies bring ethical concerns regarding testing of compounds on animals so alternative methods that directly extrapolate from *in vitro* to *in vivo* data figures would allow for these concerns to be addressed whilst maximising the output of effective data generated (Beal et al., 2022). The utilisation of *in vitro in vivo* extrapolation (IVIVE) modelling and data generation to obtain human equivalent extrapolated figures, known as administered equivalent doses (AED) (Beal et al., 2022). IVIVE is an alternative to animal testing for genotoxicity which removes the need for any *in vivo* animal testing, allowing *in vitro* data to be extrapolated to human dose values for human hazard assessment. IVIVE is also referred to as reverse dosimetry (dose reconstruction/estimation of environmental exposures consistent with biological monitoring data) and uses pharmacokinetic (drug action upon administration/the body's effect on the drug), toxicodynamic (the effect and mechanism of drugs/drug's effect on the body) and high throughput toxicokinetic (HTTK) (estimates daily dose required in humans to produce a level in the body equal to the concentration that is deemed toxic *in vitro*) modelling to predict what

in vivo doses are required to achieve an *in vitro* concentration in the blood and target tissue (Rotroff et al., 2010).

Specific toxicokinetic modelling remains unavailable for IVIVE application therefore generic HTTK modelling provides a broader chemical space and greater model implementation confidence surrounding ADME, despite the increased uncertainty. The generic HTTK model provides good estimates of human/rat *in vivo* measurements for most chemicals, however, for HTTK predictions not on the same order of magnitude (within $\sqrt{10}$ fold) as the *in vivo* figure, they will have a higher steady-state concentration in the plasma (Css) value, generating a conservative value for human hazard evaluation purposes (with very few underestimates) (Wambaugh et al., 2015). Toxicodynamically, the AC50 (concentration at 50% maximum activity) is measured using 1400 *in vitro* ToxCast assay data and systemic bioactivity/assay specific concentration is extracted from each assay (Rotroff et al., 2010). Toxicokinetic parameters are measured using the *in vitro* toxicokinetic data from the plasma protein binding assay and hepatic clearance assay with physico-chemical parameters (molecular weight and logP) to obtain IVIVE figures (Beal 2022). Toxicokinetic and toxicodynamic data is then combined to estimate the oral/dermal/inhalation AED, which is a more relevant figure for human health.

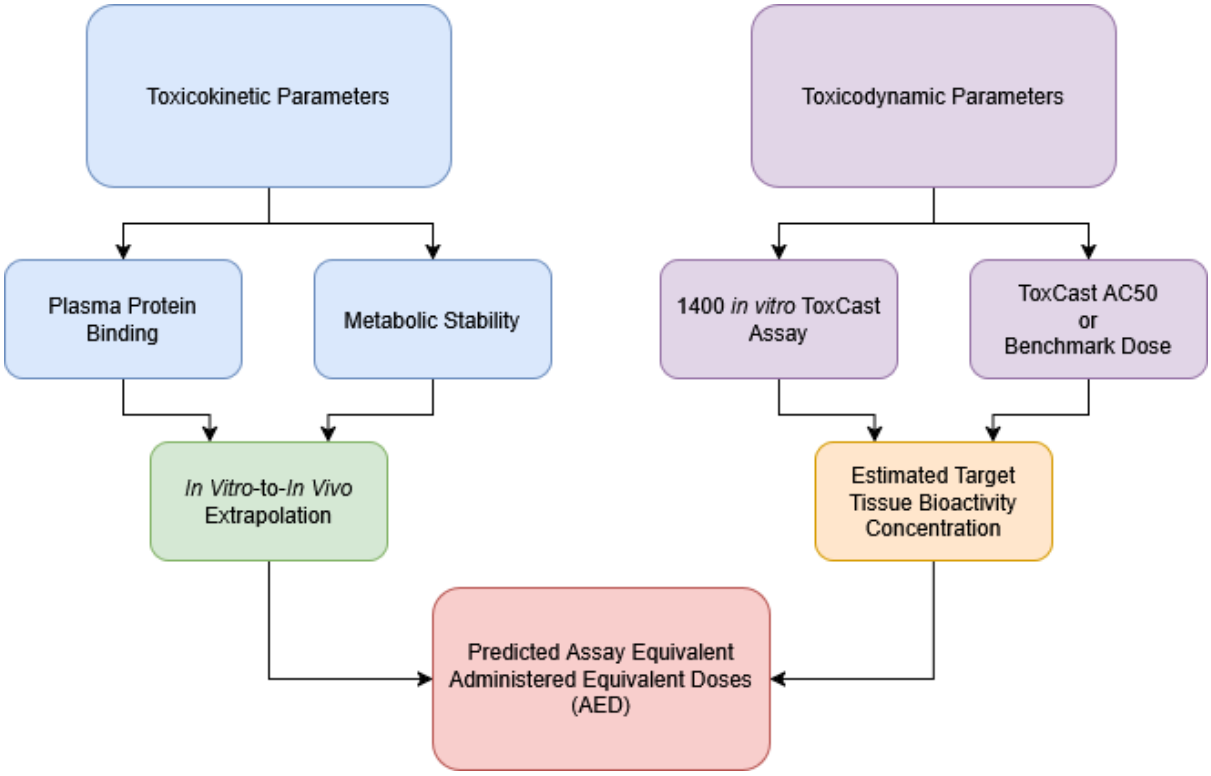


Figure 1.5.6.1 – This diagram illustrates how the integration of toxicokinetic (Plasma Protein Binding with Metabolic Stability) and toxicodynamic parameters enables the generation *in vitro in vivo* extrapolated (IVIVE) figures and estimated target tissue bioactivity concentration values, which contribute to the prediction of the administered equivalent dose (AED) in humans adapted from Rotroff et al. (2010).

Figure 1.5.6.1 shows that the integration of toxicokinetic and toxicodynamic parameters provides a more comprehensive framework for translating *in vitro* findings to *in vivo* equivalent figures. By estimating target tissue bioactivity concentrations through IVIVE, this approach enhances the predictive accuracy of the AED, supporting more informed chemical safety assessments and regulatory decision-making.

1.6 Investigation Objectives, Genotoxicity Hazard Assessment Paradigms and Precautionary Considerations

Data generated from automated *in vitro* genotoxicity multiplex assays can quantitatively measure the genotoxic effects of various mutagenic compounds on DNA damage response biomarkers at 4/24-hour analysis and +S9/-S9. Current information extraction from datasets remains limited, as pharmaceutical companies generate data-rich outputs during safety potency assessments of chemical toxicity and potency. This could potentially lead up to detailed dose-response and mechanistic biomarker data being overlooked resulting in the loss of promising drug candidates. Additional valuable information must be efficiently presented to facilitate further analysis and comprehensive dataset representation. This investigation further analysed multiplexed data generated by HESI GTTC members (Litron US, Procter and Gamble, Pfizer and Health Canada) for chemical potency determination, chemical grouping, and the efficient assessment of this information to alleviate the limitations in quantitative analysis for genotoxic studies. This underscores the importance of condensing information-rich datasets on genotoxic compounds into a suitable format for effective comparative hazard assessment studies. PROAST calculates BMD CIs that provide quantitative information on the potency and toxicity of these compounds. The Toxicological Prioritization Index (Version 2.3: Marvel et al., (2018)) (ToxPi) ranks and compares the genotoxic potency of various compounds for drug potency assessment and analyses the BMD CI across each endpoint and chemical, using approaches such as hierarchical clustering and PCA.

Chapter 2 of this investigation aimed to investigate the genotoxic effects of eight topoisomerase II inhibitors, as reported by Wheeldon et al. (2020), with a focus on how diverse mechanisms of action influenced genotoxicity scores and clustering patterns. Using QSAR, PROAST, and ToxPi, the analysis explored the relationship between chemical groupings and mechanistic differences, evaluating the ability of these tools to distinguish compounds based on potency, toxicity, and mechanistic effects. Additionally, data from Tian

et al., (2021) on 10 clastogens with DNA-reactive metabolites were analysed to compare parent compounds and their metabolites, utilising BMD confidence intervals and ToxPi profiles. The chapter also assessed the scalability of PROAST and ToxPi using a larger dataset of 30 aneugens and clastogens from the Avlasevich et al., (2021) journal, examining their effectiveness in handling complex datasets with diverse biomarkers and pathways. The overarching goal was to determine how these tools could improve the classification and understanding of genotoxic compounds with varying mechanisms of action.

Chapter 3 evaluated the effectiveness of PROAST and ToxPi in compartmentalising compounds based on their mechanisms of genetic disorder, using data from Allemang et al., (2020). The analysis compared *in vitro* and *in vivo* genotoxicity calls from a combination of assays, including the *in vitro* MN test, Toxtracker, and TGx-DDI, alongside whole genome clustering. The focus was on distinguishing compounds based on their direct and indirect acting properties and mechanisms of action (aneugenic, clastogenic, mutagenic). By integrating multiplexed genotoxic assays into a singular ToxPi output, this chapter aimed to develop an *in vitro* testing strategy that minimised the need for *in vivo* follow-up experiments. The overall goal was to provide a robust framework for predicting genotoxic potential using *in vitro* data, reducing reliance on animal testing while maintaining regulatory relevance.

Chapter 4 assessed the ability of the iScreen imaging platform to discriminate compounds by their MoAs, using data from Sun et al. (2022). The analysis focused on 10 compounds divided into four classes: clastogens, tubulin-binding aneugens, aurora kinase aneugens, and non-genotoxicants. Image analysis algorithms were employed to quantify MN presence, while biomarkers such as CENPA, γH2aX, p-H3, and MPM-2 were used to confirm clastogenicity, mitotic index changes, and polyploidy patterns. The chapter also aimed to develop a ToxPi weighting system for biological endpoints that retained regulatory relevance and minimised false positives and negatives. By demonstrating iScreen rapid scoring capabilities and alignment with regulatory frameworks, this chapter evaluated its potential as a high-throughput tool for genotoxicity screening and MoA determination.

Chapter 5 attempted to develop IVIVE (*In vitro to In vivo* Extrapolation) methods for potency prediction, using data from the Beal et al., (2022) publication. The analysis generated human

equivalent extrapolated doses (AEDs) through pharmacokinetic and high throughput toxicokinetic modelling, testing the accuracy of *in vitro* benchmark dose (BMD) assays in predicting real-world toxicities. A key focus was addressing the limitations of *in vitro* systems, such as their inability to replicate the coexistence of healthy and diseased cells in humans. The chapter also established a potency and risk factor system to reduce reliance on animal testing for genotoxicity assessment. By directly linking *in vitro* results to *in vivo* potency metrics, this chapter aimed to provide a robust framework for potency prediction and risk assessment, offering an alternative to traditional animal-based testing methods.

Chapter 6 is an in-depth review of all the acquired methodologies for benchmark dose analysis and their effectiveness across all *in vivo/in vitro* studies to evaluate their capacity at generating chemical potency information and overall compound genotoxic capacity. A potency assessment and safety summary of all information and methodologies in all previous chapters is combined, and evaluations/conclusions are drawn from all resources, figures and results gathered. The robustness of the methods employed would also need to be demonstrated regarding the relevance of the data analysed, what limitations outside of the scope of this project requires consideration (i.e. is benchmark dose a relevant metric to be used in dose response analysis), and whether the reduction of complex multilayered data into a single ToxPi metric would feasibly function or if it would present further problems.

Chapter 2 – Benchmark Dose Analysis on the *in vitro* MultiFlow and MicroFlow Assays for Quantitative DNA Damage Biomarker Assessment for Aneugens or Clastogens Across Three Genotoxic Studies

2.1 Introduction

This chapter delves into the intricacies behind the benchmark dose (BMD) approach implemented to evaluate automated *In vitro* Multiflow genotoxicity assay datasets. This study was conducted in collaboration with Litron US, who performed the Multiflow experiments and primary data acquisition, and HESI GTTC, who provided the reference chemical panel and financial support for the study. This analysis provides a quantitative assessment of compound genotoxicity through the relative potency by applying ToxPi-derived potency rankings which integrate multiple biomarker BMDL and BMDU figures to enable comparative ranking of compounds overall from most to least potent. Additionally, unsupervised hierarchical and k-means clustering are also performed chemical clustering patterns across the same DNA damage biomarkers: p53, yH2AX, p-H3, MN and polyploidy. These endpoints encompass a distinct array of DNA damaging mechanisms/pathways to undergo examination, allowing for a diverse set of compounds to be comparatively analysed. These biomarkers cover diverse DNA damage pathways, enabling comparative analysis of various compounds. The chapter examines eight topoisomerase II inhibitors across 4/24-hour time points Wheeldon et al., (2020) and evaluates parent compounds and DNA-reactive metabolites across ten clastogens, considering S9 presence (Tian et al., 2020). Additionally, it assesses 30 aneugenic/clastogenic compounds tested by Litron US published in the Avlasevich et al., (2021) paper, which incorporated biomarkers for polyploidy and the MN assay to explore a range of pathways, with analysis of ranking scores and clustering changes as the compound spectrum and MoAs expand. The original study found that combining both MultiFlow DNA damage response biomarkers with MN scoring improved the specificity of genotoxicity calls relative to the MN assay alone, and that BMD CIs provided additional mechanistic MoA and genotoxicity information across chemicals.

The broader significance of investigating multiple end point techniques would be to enhance the effectiveness of hazard assessment approaches to provide a more nuanced evaluation of their robustness in identifying potential chemical and environmental hazards. This chapter serves as a demonstration of the complexities between these analytical techniques and their ability to decipher characteristics of genotoxic and non-genotoxic compounds, aiding

decision-making in the distribution of potential new pharmaceutical drugs and the identification of hazardous chemical compounds for human safety.

2.1.1 – Multiflow Assay Datasets Handling

Litron Laboratories and HESI GTTC DNA damage Multiflow datasets were used to achieve the objectives set out by each investigation and test each hypothesis laid out in this chapter. The introduction section of this thesis describes the MultiFlow assay as a single flow cytometric analysis that combines information of several relevant DNA damage response and aneuploidy induction biomarkers against different compounds for genotoxic hazard identification and differentiating aneugenic/clastogenic compounds (Bryce, Bernacki, Bemis and Dertinger, 2016). The analysed biomarkers include γH2AX generated at Serine-139 for double strand DNA breaks assessment, p-H3 at Serine-10 for mitotic progression/cell cycle activity assessment and p53 as a nuclei-specific measurement for DNA damage detection. Each biomarker at 4/24-hour time points in the presence/absence of S9 (metabolic activation) represented an endpoint and the nuclei count was performed through a ratio of nuclei (those with 2n and higher) to counted beads calculated RNCs and cytotoxicity percentage as 100% - %RNC at 24-hours (Dertinger et al., 2019). The MultiFlow endpoints were generated for each objective/hypothesis, where human lymphoblastoid (TK6) cells was exposed to each tested compound following the protocol provided by (Bryce et al., 2016). The γ-H2aX and p53 endpoints were based on the shift in median channel fluorescence intensity relative to the solvent controls and the p-H3 biomarker measurements were based on their frequencies among other nuclei (Dertinger et al., 2019).

TK6 cells were prepared for Flow cytometry analysis using the MultiFlow® DNA damage kit developed by Litron Laboratories, (Rochester, NY) each investigation tested for various genotoxic compounds provided by HESI GTTC/Litron Laboratories, each of which will be explained in their respective chapters. MultiFlow flow cytometric data containing compound names, concentrations (μM), 24-hour RNC, the 4/24-hour γ-H2AX and p53, median fluorescence intensities and 4/24-hour p-H3 nuclei frequencies were used to obtain BMD values, and the BMD data was placed into ToxPi for further analysis.

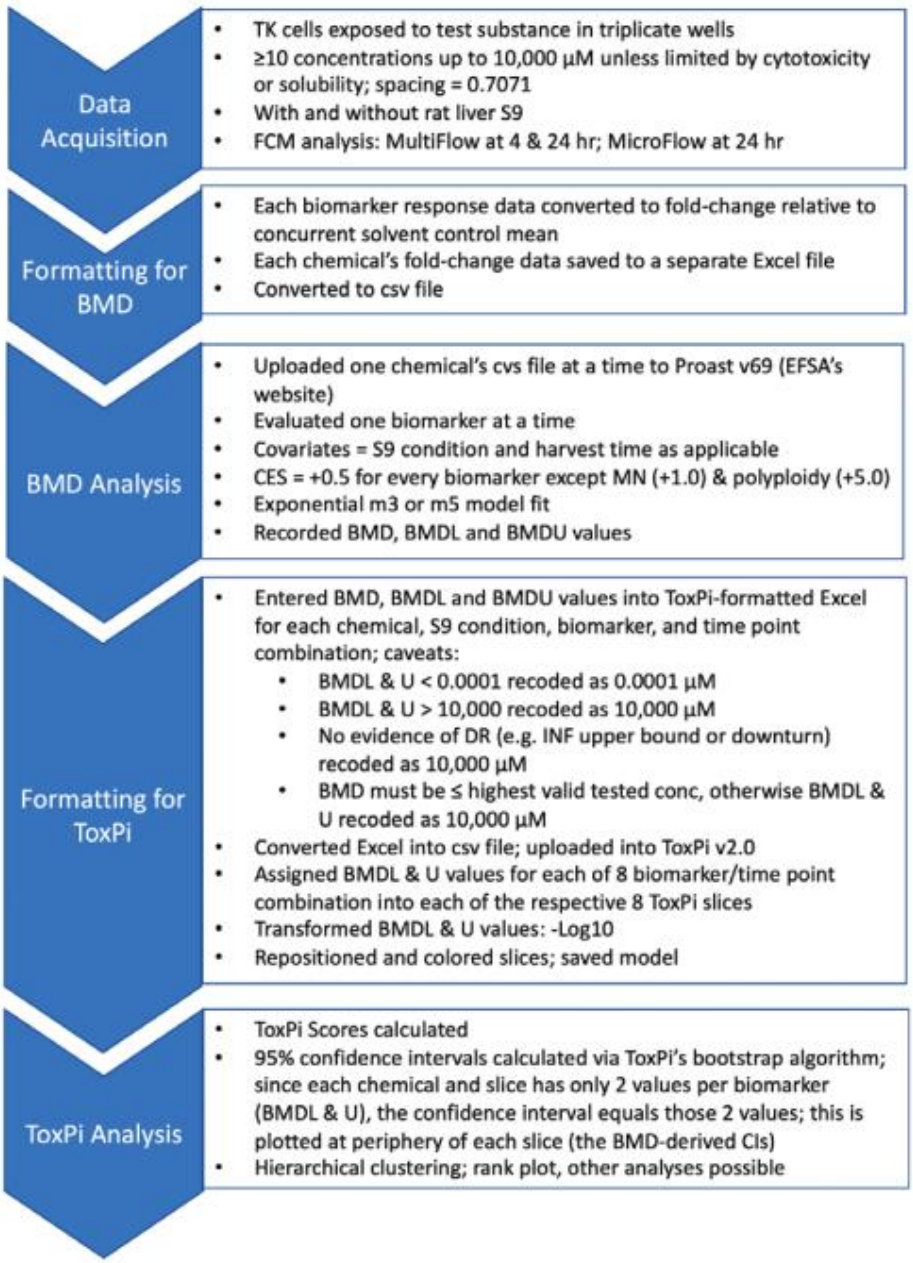


Figure 2.2.1.1 – A schematic overview of the analytical workflow adapted from Avlasevich et al., (2021), illustrating the sequential processing steps from the MultiFlow assay data acquisition, the formatting and processing of BMD modelling, multivariate integration of ToxPi for the ranking profiling, and clustering of compounds.

2.1.2 – MicroFlow Assay Datasets Handling

The Microflow Assay further tested the objectives and aims of the second chapter. The flow cytometric approach for scoring *in vitro* MN is an effective marker of genotoxicity, used to assess MN formation in TK6 cells treated with chemicals to investigate genotoxicity and dose-response (Bryce et al., 2011). The Microflow assay protocol involves the incubation of mammalian cells with a test compound and sequential staining with ethidium monoazide to segregate between MN of interest and chromatin fragments from necrotic/apoptotic cells which generates an EMA-/CYTOX+ (free nuclei and MN in normal cells) suspension segregated from the EMA+/CYTOX+ (ethidium monoazide covalently bound to chromatin fragments from apoptotic cells) suspension after washing, cell membrane lysis and incubation in RNase/SYTOX Green solution (Bryce et al., 2013). Following this, counting beads, where a relative cell counts and nuclei count is obtained from the cell-to-counting beads and nuclei-to-counting beads ratios and the flow cytometric scoring of MN allowed for the discrimination between aneugenic and clastogenic compounds to be found once all compounds were exposed to TK6 cells for 24-hours following the MicroFlow protocol using the FACSCANTO II flow cytometer equipped with a BD high throughput sampler, which detected fluorescence using a photomultiplier tube and optical filters (Bryce et al 2016).

This investigation will examine the MicroFlow data generated in Avlasenvich et al., (2021) to critically assess the ability of 30 reference compounds that induce genotoxicity through MN formation in TK6 cells. Litron Laboratories provided raw induction data for the MN endpoint at 24-hours from a mix of weak clastogens, clastogens, aneugenic tubulin binders, aneugenic aurora kinase inhibitor and non-genotoxic chemicals providing a mixture of MoA to effectively assess the classification capabilities of the microflow assay. The BMD dataset of MN induction was formatted in excel and converted to a txt file to process through PROAST and ToxPi for further assessment.

2.1.3 – DNA damage biomarkers analysis to generate compound potency assessment and providing adverse outcome pathway information

The assessment of genotoxic potency requires careful examination of complex datasets, which include relevant biomarkers to effectively evaluate the abilities of compounds to induce adverse effects on cells, tissues and organisms. Each assay assessed their own selection of biomarkers and that are related to a form of genetic damage that is likely to induce mutations/anomalies to the DNA that cause disease, most notably cancer. Chapter 1

summarises the endpoints from each assay with their respective MoAs assessed which are used to determine potential overlapping/distinct DNA damage mechanisms that produce genetic malformations, usually detecting response markers for general DNA damage, DNA breaks endpoints, MN induction, mitotic index changes, etc. For this chapter, both *in vitro* MultifFlow and MicroFlow assay biomarkers are analysed across all investigations to discover which compounds contribute towards DNA/chromosomal damage and by which MoAs based on compound ranking, profiling and clustering.

2.1.4 – The use of Benchmark Dose model to mark the point-of-departure and margin of safety to determine genotoxic potency

As discussed in Chapter 1, the BMD model will be used to determine the point-of-departure of compounds and quantitatively estimate their genotoxic potency. As previously stated, it provides an estimate of the dose required to produce a predefined change though the BMR and is preferred to highlight the MOS of a compound over the NOAEL method (Soeteman-Hernández et al., 2015). It utilises an entire dataset through modelling dose-response assessment, can be consistently applied across many assays and studies and allows for the estimation of CIs addressing any variability and uncertainty in datasets, proving its highly advantageous approach to measuring the adverse effects of compounds (Wills et al., 2016; EFSA 2022). The BMD provides a flexible approach for determining genotoxicity of substances by incorporating the entire dose-response analysis and allowing for a more reliable predictive forms of statistical analysis through the measurement of the BMD CI (EFSA, 2022). Continuous BMD modelling is applied across all chapters and will continue to be used as the modelling metric for genotoxic potency to uncover the main objectives of this investigation. Robust dose-response metrics including median fluorescence intensities, fold change and direct assay outputs were applied to retain dataset integrity and ensure that the valid dose-response figures with minimal transformations are applied to ensure the model retains its predictive accuracy.

Normally when BMD is graphically plotted, the graph contains the log concentration on the base, each compound being placed on the y-axis and the CIs represented as horizontal projections on the covariate plot. This has enabled researchers to comparatively rank compound potency and assess the uncertainty around these values. However, these covariate plots only represent the CI for comparative ranking purposes of compounds due to datasets being condensed. To obtain detailed dose response curves or apply clustering algorithms,

further tests would need to be performed to maximise information extraction from chemical datasets. This investigation aims to condense the process of running multiple analyses into a singular protocol to obtain highly detailed information from raw compound dose-response metrics.

2.1.5 Flow Cytometry as a scoring basis for *in vitro* MicroFlow and MultiFlow assays

Both MicroFlow and MultiFlow assays measure compound dose-response, however, quantification of the endpoints associated with both assays prefers flow-based advanced imaging techniques to quantify raw response data. The MicroFlow and MultiFlow assay both assess human TK6 cells through similar flow cytometric means to score compounds and for automated dose-response data acquisition to obtain raw induction figures or fold-change values as robust metrics of compound potency. The BMD modelling applied to both assays make use of latex microspheres as counting beads to obtain a nuclei-to-bead ratio, used to extract RNCs for raw response and any median channel fluorescence intensity shifts to extract the fold-change. The BMD CIs extracted can adequately highlight potency differences across compounds which has been demonstrated in previous studies (Bemis et al., 2016; Allemang et al., 2018).

2.1.6 The Three Components

This section introduces the three MicroFlow and MultiFlow investigations that were assessed in this chapter to address the importance of mechanistic assays as a drug hazard assessment tool.

2.1.6.1 Topoisomerase II inhibitors covering an array of mechanisms

Topoisomerase II is a crucial enzyme found in all prokaryotic and eukaryotic cells, maintaining DNA topology during replication, transcription, and repair by managing supercoiling and DNA tangling. The enzyme's catalytic cycle is driven by metal ions and ATP, which activate it to cleave double-stranded DNA (dsDNA), creating a transient break. Using tyrosine as the catalytic residue, it unlinks and unknots the DNA. Another dsDNA strand passes through the break, and the enzyme reseals it, relieving torsional stress and preserving the relaxed DNA structure (Nitiss, 2009). Topoisomerase II inhibitors prevent the enzyme from binding to DNA, either by disrupting enzyme activation through ATP interference or by blocking its cleaving action (Larsen et al., 2003). Inhibitors do this through covalently/non-covalently interacting with the active site of the topoisomerase II enzyme.

hese inhibitors interact with the enzyme's active site, contrasting the earlier hypothesis that the drugs primarily target DNA itself (Wilstermann and Osheroff, 2003). This disruption of the catalytic cycle prevents proper dsDNA re-ligation and stabilizes the normally transient topoisomerase II-DNA cleavage complexes. As a result, single and double-stranded DNA breaks accumulate when replication/transcription machinery collide with these trapped complexes. Therefore, topoisomerase II enzymes are important anticancer targets as inhibiting their activity can lead to replication fork arrest, trapping covalent enzyme-DNA complexes, and preventing DNA cleavage and ligation (Delgado et al., 2018). This results in DNA stabilization, leading to the accumulation of dsDNA breaks that promote replication fork collapse, chromosomal instability and ultimately cell death. This component will assess 8 topoisomerase II inhibitors: Emodin, Teniposide, Etoposide, Mitoxantrone, Doxorubicin, Flumequine, Ciprofloxacin and Genistein from the Wheeldon et al., (2020) study to observe how PROAST/ToxPi displayed the differences in their genotoxic response profiles to an identical drug target. This is through the comparison of ranking/clustering differences found in the final ToxPi result to the BMD CI chart and Derek Nexus-derived structural information quantitative structural activity relationship (QSAR) found in the original study the datasets were derived from.

The QSAR analysis utilises strong knowledge base of compound structural information to summarise and correlate relationships between compound structures and toxicity, especially when referring to chromosomal damage (Herrmann et al., 2020). Mathematical models and experimental data covering endpoints, such as the Ames bacterial mutagenicity test, are used to generate algorithms that predict compound toxicity using statistical correlations to derive physiochemical, environmental and structure-based information on compounds. It includes an increased level of depth compared to qualitative compound structure methods; however, compound classification is done based on bacterial mutagenicity assay/chromosomal damage/MN data. Therefore, a robust alternative model for QSAR needs to be developed to assist in predicting the results for alternative assays, as depicting similarities between QSAR data with unrelated endpoints, would not necessitate correlation between the two, especially considering the dichotomous nature of QSAR and its potential uncertainty surrounding the comparisons being made. There is also an argument that conventionally similar compounds in shape may lack clear identified relevant structural analogues, further emphasising this need for QSAR's evolution by including a "fingerprinting" technique that considers both key chemical bonds/structures and structural fragments – such as functional groups, ring systems

and atom connectivity patterns (Chakravarti, Radha & Alla, 2024). To enable characterisation, compounds are matched based on relevance to a query compound to determine QSAR models as shown in figure 2.1.6.1.1.

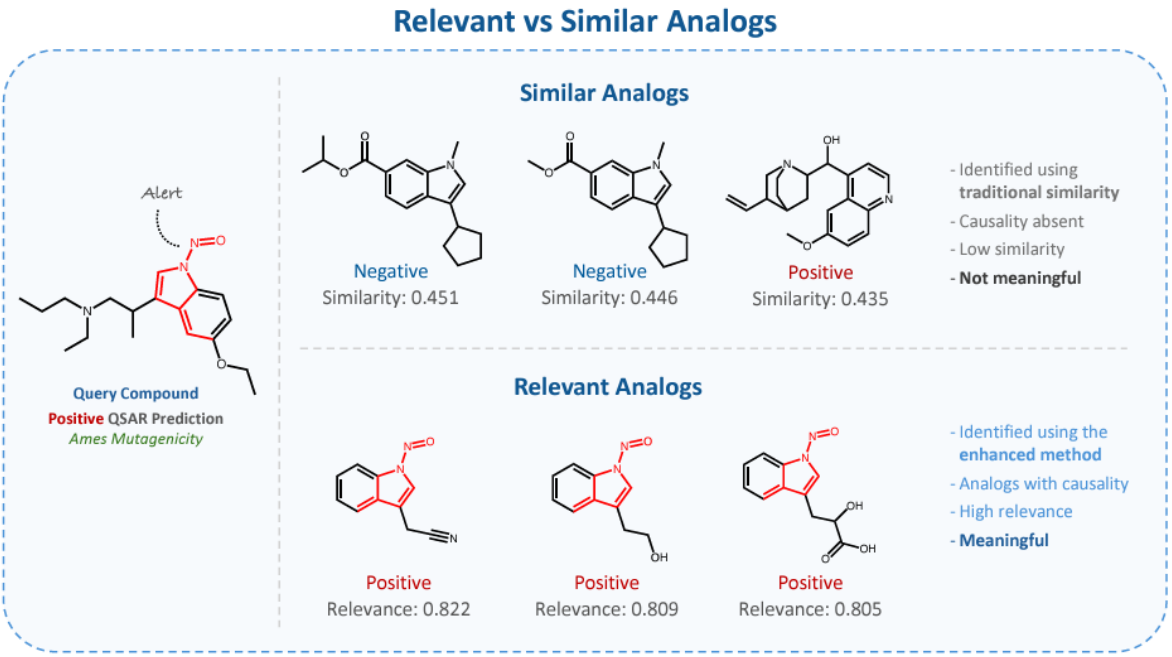


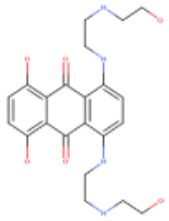
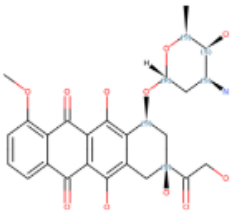
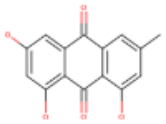
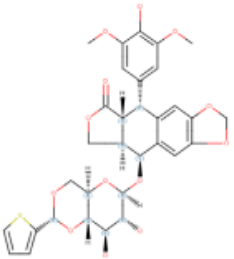
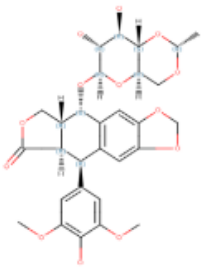
Figure 2.1.6.1.1 – Chakravarti, Radha & Alla (2024) depicting how relevant analogues have increased positive QSAR prediction results and higher relevance values to the original query compound as opposed to the traditional Similar Analogues.

Despite this, the Derek Nexus programme from Lhasa appears to provide useful structural information to further assist in sub-categorisation process of the 8 topoisomerase II compounds on a chemical-structure basis; demonstrating a quantitative approach on structure-potency based chemical groupings as shown in table 2.1.6.1.2 and table 2.1.6.1.3 (Wheeldon et al., 2020). These eight compounds are classified into one of four groups, including:

Group	**Classification**	**Compounds Tested (Topoisomerase II inhibitors)**
1	Anthraquinones	Mitoxantrone, Doxorubicin, Emodin
2	Podophyllotoxins	Teniposide. Etoposide
3	Fluoroquinolones	Ciprofloxacin, Flumequine
4	Isoflavones	Genistein

Table 2.1.6.1.2 – A table depicting each topoisomerase II inhibitor assessed and their respective structural classifications.

TABLE II. Topo II poison compound Derek Nexus derived structural information

	Compound	Alert description	Structure
1	Mitoxantrone	Hydroxylated anthraquinone/Hydroquinone derivative Ethanolamine	
	Doxorubicin	Hydroxylated anthraquinone and/or derivative Anthracycline	
	Emodin	Hydroxylated anthraquinone or precursor Anthraquinone derivative	
2	Teniposide	Podophyllotoxin or analogue Thiophene	
	Etoposide	Podophyllotoxin or analogue	

(Continues)

TABLE II. (Continued)

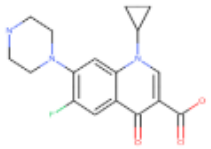
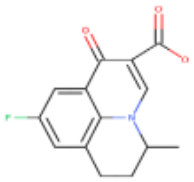
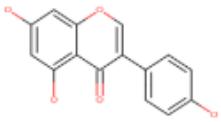
	Compound	Alert description	Structure
3	Ciprofloxacin	Halogenated quinolone-3-carboxylic acid	
	Flumequine	Quinolone-3-carboxylic acid	
4	Genistein	4′,7-Dihydroxyisoflavone or derivative Enol ether	

Table 2.1.6.1.3 – The Derek Nexus – an *in-silico* platform for QSAR – derived structural information derived from Wheeldon et al., (2020) for the 8 topoisomerase II poisons covered in this investigation, showing each compound, a description regarding each of their structural information and their respective skeletal formulae.

2.1.6.2 Clastogens that form DNA-reactive metabolites through metabolic (S9) activation

Drug metabolism converts lipophilic drugs/xenobiotics to hydrophilic metabolites in the liver before the body excretes the byproduct and, during this process, xenobiotics are bio transformed and may activate to generate chemical reactive species; a process known as metabolic activation (Attia., 2010). Compounds that were initially non-reactive have bio-transformed to highly reactive metabolites through metabolic activation, which can induce cytotoxicity/genotoxicity. The damage caused by these metabolically activated xenobiotics can target various cellular processes/mechanisms or the DNA strand directly or indirectly, however, some direct acting drug intoxicants are capable of directly acting without requiring metabolic activation, therefore, adequate testing needs to be performed to assess these differences. Figure 2.1.6.2.1 illustrates the metabolic pathways of xenobiotics, emphasising the enzymatic transformations that dictate their fate within the body. It highlights the relationship between detoxification and bioactivation, providing insight on mechanisms that may contribute to xenobiotic toxicity.

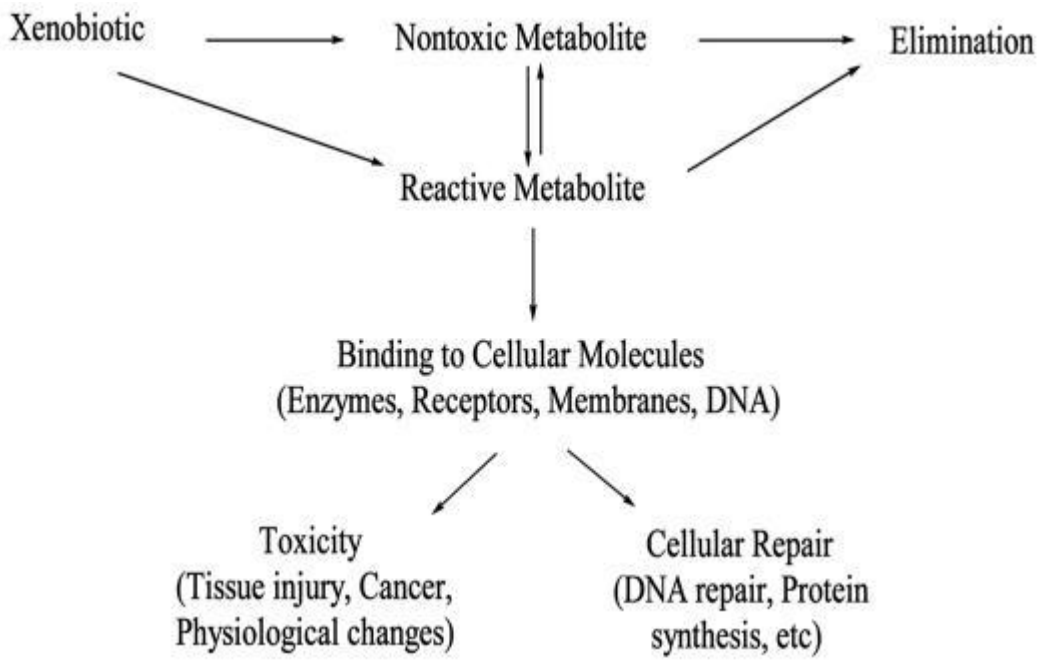


Figure 2.1.6.2.1 – A diagram demonstration the metabolic processing and potential metabolic activation, detoxification and toxicity produced by xenobiotics (Attia., 2010).

DNA reactive metabolites, including electrophilic agents such as epoxides, quinones, reactive oxygen species and free radicals can modify DNA molecules, leading to the deterioration of chromosomes/genetic material, despite their higher reactivity results being shorter half-lives due to their instability (Attia., 2010). These electrophiles or free radicals can be divided into categories based on their electrophilic strength, which determines the DNA adducts formed and can be measured to assess a compound's abilities to promote replication arrest and activate DNA repair systems to reveal their genotoxicity (Kobets, Iatropoulos & Williams, 2019). Electrophiles can form covalent DNA adducts by the interaction of electrophilic reactants with nucleophilic DNA centres, generating DNA breaks by targeting oxygen and nitrogen atoms in the phosphodiester backbone (Abdelhameed et al., 2020; Barnes et al., 2018). Free radicals (atoms with unpaired electrons) covalently bond to other free radicals (i.e. H+ atoms) to generate new radicals or remove electrons from anions to generate radical cations, which disturbs DNA structure by producing breaks, for example reactive oxygen species.

In genotoxicity testing, evaluating drugs with and without metabolic activation using S9 rat liver fraction is crucial as it simulates metabolic processes occurring in the liver due to a rich presence of xenobiotic metabolising enzymes and cytochrome P450, which can convert pro-mutagens (inactive compounds) into their active mutagenic forms. The S9 fraction, derived from rat liver enzymes, mimics these metabolic reactions, thereby providing a comprehensive assessment of a substance's genotoxic potential. This dual approach ensures that both the parent compound and any metabolites are tested for genotoxic effects, leading to a more accurate and thorough evaluation of the compound's potential genotoxic effects. By incorporating metabolic activation, researchers can identify genotoxic agents that might otherwise be missed in non-activated tests, thereby enhancing drug safety profiles and regulatory compliance.

This chapter assesses MutliFlow data generated on 10 known clastogens that form DNA reactive metabolites through metabolic activation using S9 rat liver at 4 and 24-hour time points: Mitomycin C, Dibenzol[a,l]pyrene, 2-Aminoathracene, Benzo[a]pyrene, 7,12-Dimethylbenzanthracene (DMBA), 2-amino-1-methyl-6-phenylimidazo [4,5-b]pyridine (PhIP), Cyclophosphamide Monohydrate, 2-Actylaminofluorene, Diethylnitrosamine and Resorcinol (generally non-genotoxic, although weak clastogenic properties effects have been reported *in vitro* at high concentrations) (Wheeldon et al., 2020). This was to examine the

effectiveness of PROAST/ToxPi in comparing parent compound and metabolite to observe any significant differences between BMD CIs, ToxPi profiles, ToxPi scoring, clustering and overall compound potency and ranking. However, as only clastogens were analysed, only the p53 and γH2aX were assessed as p-H3 was an irrelevant biomarker that would not provide relevant response figures and would skew any potency and clustering findings generated. This means that p53 and γH2aX results were analysed across both 4-hour and 24-hour timepoints.

2.1.6.3 Extended assessment of and expanded dataset on compound potency assessment

There has been a continued effort to supplement *in vitro* MN endpoint with biomarkers that provide an in-depth analysis of the MoAs of potentially genotoxic compounds.

Although some tests are more sensitive to some compounds compared to others (e.g. the chromosome aberration test detect clastogens), combining endpoints has shown great improvement in mechanistic discrimination, such as differentiating between clastogens and aneugens when the chromosomal aberration assay is used alongside MN analysis (Parry et al., 2002; Kirkland et al., 2011). However, as more testing assays emerge, some concern has been placed on which tests are chosen and how to categories between different testing combinations to detect different MoAs across different investigations/study designs. Therefore, investigations into the expanded use of combined MultiFlow and MicroFlow assay were initiated by Avlasecvich el al., (2021) building on the evidence from Smart et al., (2020) supports the use of combinational flow cytometry-based endpoint assays on TK6 cells.

Upon extension of any dataset, questions arise surrounding how results/data points would shift or be influenced by the assays performed, endpoints measured, and the list of compounds included. The primary aim of this component is to evaluate whether the integrated PROAST/ToxPi methodology can maintain a robust assessment of compound potency, ranking, clustering and visualisation when applied to a larger and more diverse dataset. These include the assessment of 30 substances comprising of aneugens, clastogens and non-genotoxic compounds and the number of endpoints assessed (p53, γH2aX, phosphorhistone-H3, MN, and polyploidy) at 4-hours at 24-hours timepoints, both in the presence and absence of S9 rat liver fraction (Avlasevich et al., 2021). By doing so, the predictive utility of PROAST/ToxPi to be applied to realistic compound screening panel while enhancing mechanistic resolution across diverse pathways and biomarkers.

2.2 – Materials and Methods

This chapter constructs the backbone surrounding the methodologies used for all 3 components highlighting the design surrounding both MultiFlow and MicroFlow assays.

2.2.1 – Dataset Acquisition

The selection of testing systems in this chapter are purely *in vitro* based system and consist of the endpoints generated by both the MultiFlow and MicroFlow DNA damage assays, including γH2aX, p53, p-H3, MN and polyploidy. These results were gathered and placed into tables separating out their concentrations in (μM) from the median fluorescence intensity using propidium iodide for and γH2aX and p53 endpoints, microbead counting sphere RNCs and relative population doubling for p-H3 and polyploidy, and finally, MN induction was found using RNC with fluorescence emissions from sequential staining of SYTOX green and ethidium monoazide at G1 phase, S phase, G2M cells and polyploidy cell phases. Raw data for both the 8 topoisomerase II of varying mechanisms with an identical drug target and 10 clastogen metabolic activation comparison between the parent compound and the metabolite components were provided directly by Wheeldon et al., (2021), while the 30 aneugen/clastogens collection data was provided by Litron Laboratories. Cell lines across all 3 components involved TK6 cells. For both assays, all measurements were performed at either 4/24-hour timepoints depending on the component of this chapter and all spreadsheets were generated on Excel with their known *a priori* MoAs highlighted for each chemical; this will be further elaborated on later in this chapter.

2.2.2 – Cell treatments and MultiFlow processing for the assessment of identical targets with varying mechanisms of 8 topoisomerase II inhibitors

The benchtop experiments were performed by Derek Bernacki, Stephen Dertinger, Steven Bryce and Jeffery Bemis from Litron laboratories using their MultiFlow DNA damage kit – p53, γH2aX and p-H3 endpoints. ATCC® purchased TK6 cells were grown in a humid environment at 37C with 5% CO2, maintained at 1x10^6 cells/mL (Wheeldon et al., 2020). The culture medium used consists of RPMI 1640, 200 μg/mL sodium pyruvate (both from Sigma-Aldrich St. Louis MO), 200 μM L-glutamine, 50 units/mL penicillin, 50 μg streptomycin (all 3 purchased from Mediatech, Manassas VA), and 10% v/v heat inactivated horse serum (Gibco Thermo Fisher Scientific, Waltham MA) (Wheeldon et al., 2020). Cells

were exposed to 20 concentrations of each test chemical with root √2 serial dilutions (70.71% factorial difference between consecutive concentrations (1÷√2 application)) and each concentration was evaluated in duplicate wells (n=2) (except for the solvent dimethylsulfoxide (DMSO), which was replicated across 4 wells). Once cell count was adjusted to 2x10^5 cells/mL, 198 µL of TK6 cell suspension was added to 96-well plates along with 2 µL of each test chemical dissolved in DMSO to generate 200 µL in total per well. At 4/24-hour timepoints, the cells were resuspended by pipetting and 25 µL is removed from each well and subsequently placed into a new 96-well plate containing 50 µL of the MultiFlow reagent mixture in each well and the test chemical is mixed via pipetting. After incubation for 30 minutes, the sample is analysed via flow cytometry. The chemicals used in this investigation are listed in the table below:

Table 2.2.2.1 – A table highlighting the chemicals used in this component of the Wheeldon et al., (2020) study.

Chemical (Abbreviation)	**CAS Registry Number ID, Source – Sigma Aldrich**	**Mode of Action**
Mitoxantrone	65271-80-9	Topoisomerase II inhibitor (Clastogen)
Doxorubicin	23214-92-8	Topoisomerase II inhibitor (Clastogen)
Teniposide	29767-20-2	Topoisomerase II inhibitor (Clastogen)
Etoposide	33419-42-0	Topoisomerase II inhibitor (Clastogen)
Genistein	446-72-0	Topoisomerase II inhibitor (Clastogen)
Emodin	518-82-1	Topoisomerase II inhibitor (Clastogen)
Ciprofloxacin	85721-33-1	Topoisomerase II inhibitor (Clastogen)
Flumequine	42835-25-6	Topoisomerase II inhibitor (Clastogen)

Flow cytometry is performed using FACSCanto™ II flow cytometer, equipped with a BD™ High Thoroughput Sampler or a Miltenyi MACSQuant® analyser 10 flow cytometer, integrated with a 96-well MiniSampler device. Photomultiplier tubes and optical filters measured any fluorescence emitted and the fluidics parameters were user defined and are specified for each:

Table 2.2.2.2 – This table highlights each flow cytometer device used and their respective fluidics settings (Bryce et al., 2016).

Device	**Fluidics Settings**
FACSCanto™ II Flow Cytometer	40 µL sample mixed (x4) Mixing speed of 180 µL/sec 20 µL volume analysed at a flow rate of 1 µL/sec (until 10000 events acquired of 20 µL exhausted) 400 µL sheath fluid used to wash probe
MACSQuant® Analyser	40 µL sample mixed 20 µL volume analysed at a flow rate of 50 µL/min 208.4 µL sheath fluid used to wash probe

The flow cytometer was used to assess p53 and γH2AX levels through median channel fluorescence intensity. Propidium iodide fluorescence was employed to determine DNA content and establish fold change values relative to the solvent control, calculated by dividing the fluorescence measurements by the mean solvent control value (Wheeldon et al., 2020). Gating criteria required propidium iodide fluorescence to correspond to 2n–4n DNA content. To minimize the influence of mitotic and apoptotic cells on the measurements, γH2AX-positive and p-H3-positive events were excluded from the analysis (Bryce et al., 2016; Bryce et al., 2017). Phosphorylation of histone-H3 (ser10) is a well-established marker of chromatin condensation during mitosis and increased levels indicate dysfunction in mitotic progression, arising from disruptions in mitotic spindle fibre assembly or chromosome segregation. Extended mitotic arrest may result in mitotic slippage or chromosomal mis-segregation, this results in the accumulation of cells exhibiting polyploidy/aneuploidy thereby providing mechanistic evidence of aneugenic genotoxicity. Data on p-H3 and polyploidy data was also collected through flow cytometry, however, instead of median channel fluorescence intensity,

p-H3 measurements is found through the relative percentage of p-H3 positive nuclei and increased DNA content for polyploidy. The mitotic index was calculated by comparing the proportion of p-H3-positive events with propidium iodide fluorescence corresponding to ≥4n DNA content relative to events with 2n or greater DNA content (Bryce et al., 2016; Bryce et al., 2017). Polyploidy was similarly quantified, but the proportion of events with ≥8n DNA content was compared to those with ≥ 2n DNA content (Bryce et al., 2016; Bryce et al., 2017). Additionally, p-H3 data were converted to fold change values by dividing by the mean value of the solvent control, analogous to the normalization procedure for γH2AX and p53 (Wheeldon et al., 2020).

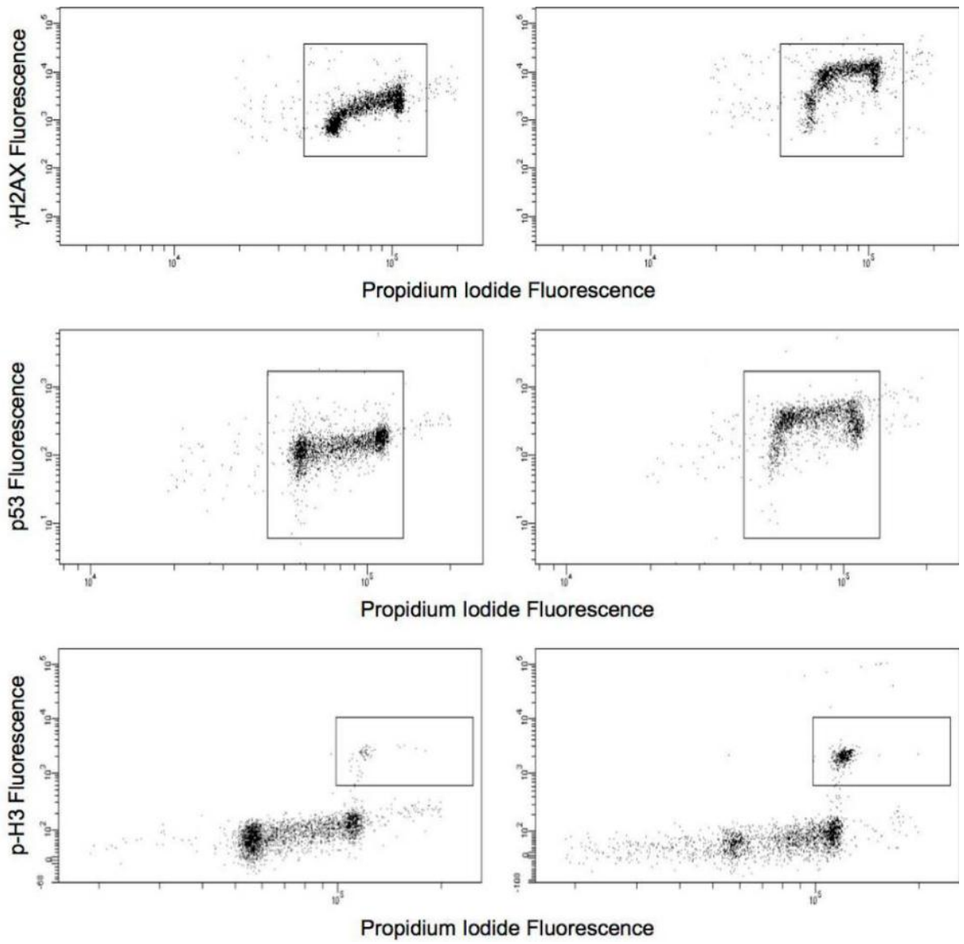


Figure 2.2.2.1 – Demonstration of gating process performed during flow cytometric analysis of p53, γH2aX and p-H3 in the MultiFlow assay, where cells treated with solvent controls are on the left and cells exposed to each test chemical is on the right. The X-axis represents cell populations through propidium iodide fluorescence, and the Y-axis shows fluorescence intensity of each reporter assessed in the MultiFlow assay with boxed regions on each graph indicating biologically relevant cell population used for analysis.

- Top panels: yH2AX versus propidium iodide induction. Solvent control (left) shows baseline yH2AX induction while Mitomycin C (right) exposure increases yH2AX fluorescence, consistent with dsDNA breaks and clastogenic activity.
- Middle panels: p53 versus propidium iodide induction. Solvent control (left) shows baseline p53, while etoposide (right) exposure increases nuclear p53, reflecting DNA damage response activation due to Topoisomerase II inhibition activity.
- Bottom panels: p-H3 versus propidium iodide induction. Solvent control (left) shows baseline p-H3 while griseofulvin (right) increases p-H3 levels, reflecting the accumulation of cells in mitosis due to spindle fibre disruption indicative of aneugenic activity (Bryce et al., 2016).

2.2.3 – Cell treatments and MultiFlow processing to analyse metabolite and parent compound across 10 clastogens

The procedures and methods described in section 2.2.2 directly correlate with the methodologies applied to this component in the investigation with three key differences. Firstly, the p-H3 endpoint was never tested for as this study was primarily designed to characterise clastogenic responses. For example, although some topoisomerase II inhibitors have been reported to exhibit secondary aneugenic effects at certain concentrations, the current experimental framework prioritised DNA damage biomarkers mostly aligned with strand break mediated mechanisms (Attia et al., 2003; Attia, 2011). Therefore, p53 and γH2aX are still tested for across both 4-hour and 24-hour timepoints to assess early DNA damage signalling alongside downstream cell cycle dependent responses following prolonged exposure (to allow one cell cycle progression of TK6 cells). Secondly, the compounds in this component are tested in both the presence and absence of low dose 0.25% vol/vol S9 rat liver as described in (Tian et al., 2020) to assess the effects of metabolic activation on the 10 clastogens assessed. Therefore, the endpoints assessed in this component are:

- 4hr-γH2aX S9-
- 4hr-γH2aX S9+
- 24hr-γH2aX S9-
- 24hr-γH2aX S9+
- 4hr-p53 S9-
- 4hr-p53S9+
- 24hr-p53 S9-
- 24hr- p53 S9+

Finally, the list of compounds tested have changed and are summarised in the table below:

Table 2.2.3.1 – A summary of all 10 clastogens assessed, their respective CAS reference numbers, each MoA and metabolic properties, adapted from the Tian et al., (2020) paper.

Clastogenic Chemical (Abbreviation)	**CAS Number, Source – Sigma Aldrich**	**10 Clastogens Mode of Action, Metabolic Properties and Citations**

Mitomycin C	50-07-7	Clastogen, DNA cross-linking (dsDNA breaks), alkylating properties and oxidative stress (Kirkland et al., 2016)
Dibenzol[a,l]pyrene	191-30-0	Clastogen, polycyclic aromatic hydrocarbon, requires metabolic activation (thought to be primarily activated by CYP1A1) (Arif and Gupta., 1997)
Resorcinol	108-46-3	Non-genotoxic, *in vivo* (Negative in mouse MN), weak clastogenicity *in vitro* mammalian cell positive (MLA assay with and without metabolic activation positive/*in vitro* human lymphocyte MN positive in absence of metabolic activation) (European Food Safety Authority, 2010).
2-Aminoathracene	613-13-8	Clastogen, aromatic amine, requires metabolic activation (CYP1B1, 2A family) (Carriére et al., 1992).
Benzo[a]pyrene	50-32-8,	Mutagenic (with clastogenic properties), polycyclic aromatic hydrocarbon, requires metabolic activation (CYP1B1, 2A family) forming DNA-reactive metabolites that induce Guanine to Thymine transversions (Kirkland et al., 2016).
7,12-Dimethylbenzanthracene (DMBA)	57-97-6	Clastogen, requires metabolic activation (CYP1B1), forms bulky adducts (Kirkland et al., 2016).
2-amino-1-methyl-6-phenylimidazo [4,5-b] pyridine (PhIP)	105650-23-5	Clastogen, heterocyclic amine, requires metabolic activation (CYP1A family) (Kirkland et al., 2016; Krais et al., 2016).
Cyclophosphamide Monohydrate	6055-19-2	Clastogen, nitrogen mustard, requires metabolic activation (CYP2B6, CYP2C19, CYP2C9, and CYP3A4/5) (Kirkland et al., 2016; Rodriguez-Antona and Ingelman-Sundberg., 2006)
2-Acetylaminofluorene	53-96-3	Mutagen (with secondary clastogenic effects), requires metabolic activation (CYP1A2), that undergoes cytochrome P450–mediated N-hydroxylation that form

		DNA reactive metabolites that generates C8 adduct on guanine (Otteneder and Lutz, 1999; Kirkland et al., 2016).
Diethylnitrosamine	55-18-5	Clastogen, Polycyclic aromatic hydrocarbon, requires metabolic activation (Yamazaki et al., 1992).

The main aim of this component was to demonstrate the capacity that ToxPi/PROAST can differentiate between parent compounds and their respective metabolites. Their effectiveness for the comparison between parent compound and the metabolite can be examined when assessing significant differences in Benchmark Dose CIs, ToxPi profiles, ToxPi scores, clustering, and overall genotoxicity.

2.2.4 – Cell treatments and MultiFlow/MicroFlow processing to investigate the effects of data expansion on compound potency assessment and MoA grouping using 30 aneugens/clastogens

The procedures described in the Bryce et al., (2010; 2016; 2017) and Tian et al., (2020) publications reflected the MultiFlow process and parent compounds against metabolite assessment of across all compounds at 4-hour and 24-hour timepoints, hence, section 2.2.2 and 2.2.3 can be referred to for these procedures. After 4/24-hours of exposure to each chemical, cells are processed through the *In vitro* MicroFlow® kit reagents (Litron Laboratories, Rochester, NY, USA) using the remaining 2 X 25 µL samples from the MutliFlow assay. The MicroFlow assay reagents included buffer solution, nucleic acid dye A solution (ethidium monoazide), nucleic acid dye B solution (SYTOX® green dye), lysis solution I and II and RNase solution.

Harvested TK6 cells are stained with 50 µL of complete nucleic acid dye A solution (6.5 mL buffer solution and 65 µL of nucleic acid dye A) and resuspended with 100 µL of complete lysis solution 1 (11.0 mL incomplete lysis solution 1, 44 µL nucleic acid dye B, 55 µL RNase solution and 1 drop of counting beads) with the forceful addition of 100 µL of complete lysis solution 2 (11.0 mL incomplete lysis solution 2 and 44 µL nucleic acid dye B) across all 96-wells. Cells are incubated for 24-hours, and flow cytometric analyses were carried out using a BD FASCANTO™ II flow cytometer equipped with a BD High Throughput sampler, stock photomultiplier tubes and optical filters set were used to detect fluorescence emissions associated with the MicroFlow kit provided: SYTOX Green (FITC

channel) and ethidium monoazide (EMA, PerCP-Cy5.5 channel). 10,000 EMA- nuclei per replicate well was the stop mode setting and MN concentration was restricted to ≤60% reduction to the relative increased nuclei count, lowest precipitating concentration evaluated was either the lowest concentration or 10mM (whichever was lowest) and MN% was converted to fold change figures relative to the mean solvent control (S9± separate). MN significance was set at either >2.5x over the solvent control and a mean MN% exceeding the upper bound tolerance interval from historical negative control data as expressing responses as fold change relative to the solvent control normalises for inherent background variability and facilitates the identification of biologically meaningful increases in genotoxicity with adequate statistical sensitivity. The procedures here are described in Avlasevich et al., (2006), Bryce et al., (2013), Bryce et al., (2016), Litron Laboratories, (2018), Avlasevich et al., (2021) and the compounds assessed in this component are included in the table below:

Table 2.2.4.1 – The list of 30 compounds tested for the extended assessment of MultiFlow and MicroFlow assay datasets observed in Avlasevich et al., (2021), including their CASRN and MoA.

Chemical (Abbreviation)	**CAS Registry Number ID, Source**	**Mode of Action**
Vinblastine Sulphate	143-67-9, Sigma-Aldrich	Tubulin Binder (Aneugen) (Kirkland et al., 2016)
Ixabepilone	219989-84-1, Selleckchem	Tubulin Binder (Aneugen) (Lee et al., 2001)
Etoposide	33419-42-0, Sigma-Aldrich	Topoisomerase II inhibitor (Clastogen) (Kirkland et al., 2016)
Paclitaxel	33069-62-4, Sigma-Aldrich	Tubulin binder (Aneugen) (Kirkland et al., 2016)
Colchicine	64-86-8, Sigma-Aldrich	Tubulin binder (Aneugen) (Kirkland et al., 2016)

AMG900	945595-80-2, Selleckchem	Aurora kinase inhibitor (Aneugen) (Payton et al., 2010)
Melphalan	142-82-3, Sigma-Aldrich	Nitrogen mustard-type alkylator (Clastogen) (Dertinger et al., 2012)
Temozolomide	85622-93-1, Sigma-Aldrich	DNA alkylating agent (Clastogen) (Chinnasamy et al., 1997)
VX680	639089-54-6, Selleckchem	Aurora kinase inhibitor (aneugen) Harrington et al., 2004
PF03814735	942487-16-3, Selleckchem	Aurora kinase inhibitor (aneugen) (Jani et al., 2010)
Chlorambucil	305-03-3, Sigma-Aldrich	Nitrogen mustard-type alkylator (Clastogen) (Dertinger et al., 2012)
Cisplatin	15663-27-1, Sigma-Aldrich	Atypical alkylator/cross-linking (Clastogens) (Kirkland et al., 2016)
Genistein	446-72-0, Sigma-Aldrich	Topoisomerase II inhibitor (Clastogen) (Klein and King., 2007)
Methyl Methanesulfonate (MMS)	66-27-3, Sigma-Aldrich	DNA alkylating agent (Clastogen/mutagen) (Kirkland et al., 2016)
Dibenzol[a,l]pyrene	191-30-0, Sigma-Aldrich	Polycyclic aromatic hydrocarbon, requires metabolic activation (thought to be primarily activated by CYP1A1) (Clastogen) (Arif and Gupta., 1997)
Methotrexate	59-05-2, Sigma-Aldrich	Anti-metabolite (Clastogen) (Keshava et al., 1998)
2-Aminoanthracene	613-13-8, Sigma-Aldrich	Aromatic amine, requires metabolic activation (CYP1B1, 2A family) (Clastogen) (Carriére et al., 1992)

Thiotepa	52-24-4, Sigma-Aldrich	DNA alkylating agent (Clastogen/mutagen) (Dertinger et al., 2012)
2-amino-1-methyl-6-phenylimidazo [4,5-b] pyridine (PhIP)	105650-23-5, Toronto Research Chemicals	Heterocyclic amine, requires metabolic activation (CYP1A family) (Clastogen) (Kirkland et al., 2016; Krais et al., 2016)
Ethyl Methanesulfonate	62-50-0, Sigma-Aldrich	DNA ethylating/alkylating agent (Clastogen) (Gocke, Bürgin, Müller and Pfister. 2009)
Brefeldin A	20350-15-6, Sigma-Aldrich	ER-golgi transporter inhibitor, ER stress-induced apoptosis (Non-genotoxic) (Moon, Kim, Shin and Park, 2012)
Hydroxyurea	127-07-1, Sigma-Aldrich	Ribonucleotide reductase inhibitor; (non-mutagenic clastogens) (Ware and Dertinger, 2021)
Cyclophosphamide Monohydrate	66-81-9, Sigma-Aldrich	Nitrogen mustard, requires metabolic activation (CYP2B6, CYP2C19, CYP2C9 and CYP3A4/5) (Clastogen) (Kirkland et al., 2016; Rodriguez-Antona and Ingelman-Sundberg., 2006)
Thapsigargin	67526-95-8, Sigma-Aldrich	Cytotoxicant, ER stress-induced apoptosis (Non-Genotoxic) (Futami, Miyagishi and Taira, 2005)
Dexamethasone	50-02-2, Sigma-Aldrich	Glucocorticoid receptor agonist (Non-Genotoxic) (Krishna et al., 1995)
Alosetron Hydrochloride	122852-69-1, Sigma-Aldrich	5-HT3 receptor antagonist (Non-Genotoxic) (Kirkland et al., 2016)
Cycloheximide	66-81-9, Sigma-Aldrich	Protein synthesis inhibitor (Non-Genotoxic) (Youngblom, Wiencke, Wolff, 1989)
D-Mannitol	69-65-8, Sigma-Aldrich	Polyol (Non-Genotoxic) (Kirkland et al., 2016)

Tolterodine Tartrate	124937-52-6, Cayman Chemical.	Holinergic Muscarinic Antagonist (Non-Genotoxic) (Avlasevich et al., 2021)
1, 2-Dimethylhydrazine	540-73-8, Sigma Aldrich	DNA methylating agent, metabolic activation needed (IARC., 1999)

2.2.5 – PROAST

2.2.5.1 – PROAST for Benchmark Dose Analysis

In line with genotoxicity testing guidance, the range of concentrations taken forwards were chosen based on the cytotoxicity (calculated by 100% - RNC) percentages from the results generated after the flow cytometric counting beads process, where top concentrations of each chemical exhibited ≤80% reduction to the RNC at 24-hours, as described in Bryce et al., (2016) and Dertinger et al., (2019). Any concentration where cell survival <20% is observed at 24-hours compromises mechanistic interpretability and increases the likelihood that observed responses reflect secondary toxicity rather than primary genotoxic mechanisms. Therefore, any concentrations where cytotoxicity exceeded 80% was excluded from any further BMD processing and further analysis as anything exceeding this figure would indicate excessive cell cytotoxicity and death. Concentrations showing excessive cytotoxicity are not considered reliable data points because they cause non-specific effects, off-target interactions, and experimental noise, making the results unreliable and irrelevant to therapeutic levels (Stine and Brown, 2015; Chang and Atterwill, 1998). Furthermore, no more than two concentrations within the cytotoxicity range of 70-80% were permitted. These limits would ease MN scoring and prevent external factors outside of drug exposure to contribute to the overall drug dose-response effect demonstrated, generating false positive results (Kirkland et al., 2007). For example:

1. Whether reactive oxygen species were generated in the culture medium or by the test chemicals themselves.
2. If any degradation products were present culture medium prior to testing.
3. Metabolic poisoning/overload leading to excessive cell death.
4. The loss of cellular homeostasis increasing RNC cytotoxicity.
5. Further genetic factors/external cellular contributions producing misleading figures.

Chapter 1.3 discussed the BMD model as dose/concentration that produces a predetermined change in the response rate of an adverse effect caused by a compound and BMD analysis is performed using PROAST, which uses assay generated dose-response data with statistical modelling algorithms to provide quantitative information on compound potency and toxicity (Davis, Gift & Zhao, 2011; RIVM, 2021). The BMD is the estimated dose level generated from dose-response curve fitted to nested exponential/hill models. PROAST software package (version 61.2-70.3 operating in R) utilises Exponential and Hill coefficient models (as recommended by European Food Safety Authority) to analyse concentration-response data for each chemical's biomarker and combines the datasets to generate BMDL and BMDU (90% CI), a quantitative figure regarding compound potency and toxicity, which can be further visualised to understand the dose response curve of each compound tested (Wills et al., 2017 & RIVM, 2021). RIVM have provided a web browser version of PROAST, however, the R Gui version (v4.2.2 - 4.4.1) was preferred as it contained the complex functionality required for this investigation.

The BMD is mathematically modelled after the implementation of nested exponential/hill models are applied to the original concentration and induction values to map out the dose-response curve using this family of models known as the four-parameter model, thereby estimating the BMDL and BMDU using the CES. The Slob (2002) paper describes that for toxicological continuous endpoints, no single dose-response model can perfectly encapsulate the underlying biological mechanisms, instead, a collection of descriptive models can be devised to fit the dose-response data. Despite this, not all models are equally suitable for mapping out this true relationship for hazard assessment, so the Slob (2002) study outlines important points for developing a generic family of dose-response models that fulfils this purpose. (1) The continuous toxicological measurements are non-negatives and any modelling that allows for negative responses are unrealistic, which excludes polynomial functions ($y= a + bx + cx^{2 +}...$) as an option. (2) Threshold models are fundamentally excluded as their strictness is hard to defend when regarding dose-response relationships as it suggests that all individual animals have the same dose-thresholds for each continuous endpoint observed (which is highly unlikely even in a controlled environment). (3) Relative sensitivity across different populations is an important concern (i.e. inter-species, inter-sex, etc) so it's important that the relative sensitivity is reflected by one of the model's parameters. Two populations with identical background values of a particular endpoint would demand an equal response across all doses, however, for populations that have these

differences, this background value needs to be considered. For example, bodyweight between males and females are usually different, hence, sensitivities would differ. The Slob (2002) paper defines equal responses in both sexes as:

$$\frac{maleBW(d)}{maleBW\ (0)} = \frac{femaleBW(d)}{femaleBW\ (0)}\text{, for every dose (d)}$$

Slob (2002) explains that if the relative change compared to the background is the same for all doses, then sensitivity across two groups would be identical, which is expressed in the formula:

$$y = af(x) \text{ – equation 1}$$

Next, Slob (2002) describes how to express differences in sensitivity across two populations. Potency factors are consistent with the concept of uncertainty factors and if suitable, these factors can be directly estimated, hence the dose-response model for different background values follows:

$$y = af(\frac{x}{b}) \text{ – equation 2}$$

Furthermore, Slob (2002) continues to elaborate on (4) equation 2 following f0=1, which is achieved by following exponential equations, (5) nesting models should be applied so that (log) likelihood ratio testing can be performed and (6) the family-models should include functions that plateau/level off at higher doses to reflect/mimic the threshold-like responses. This generates the following:

Model 1: y=a with $a > 0$

Model 2: y=aexp (x/b) with $a > 0$

Model 3: y=a exp($\pm$b x^d) with $a > 0$, b > 0, $d \geq 1$

Model 4: y = a [c – (c – 1) exp (-x/b)] with $a > 0$, $b > 0$

Model 5: y = a [c - (c - 1) exp (- x/b)^d] with $a > 0$, $b > 0$, $c > 0$, $d \geq 1$

For each model, y is any continuous endpoint, the x represents the dose, *a* represents the level of the endpoint at dose 0 (background response), *b* reflects the efficacy of the chemical/sensitivity of a subject and is directly correlates with the potency in the exponential model, *c* is the maximum relative change and *d* behaves like is constrained to larger values to prevent an infinite slope value at dose 0 and reducing the impact dose has on response (Slob, 2002; Slob and Setzer, 2014). All parameters contributing towards the shape of the dose-response curve. The likelihood ratio tests applied to test whether additional parameters to a model result in a statistically significant improvement of the fitting of the results to the model in question (Slob, 2002). This therefore tests whether restrictions/additional parameters on a model are necessary. The same four-parameters (*a,b,c,d*) described earlier can also be applied to the hill models which follow which EFSA introduced as another suitable dose-response family model:

Hill model: $y = a\left(1 + \frac{(c-1)\,x^d}{b^d + x^d}\right)$ with $a > 0$, $b > 0$, $c > 0$, $d \geq 1$

From the conclusions of the Slob and Setzer., (2014) study, which determined that the goodness of fit for the exponential and hill models were difficult to define, they found that exponential and hill models both similarly fit dose response datasets**.** Following EFSA guideline recommendations, both hill and exponential models 3/5 were implemented for BMD analysis. Model selection was guided using the Akaike information criterion enabling the comparison of relative model fit, to apply the appropriate models while balancing explanatory power and model complexity. The chosen models are placed detailed in the result section.

PROAST v70.3 was implemented into the work carried out in this thesis and the results will highlight the settings and parameters selected for each of the 3 components. The original xlsx files containing raw dose-response needed to be converted into a txt file with strict formatting of the compound names, doses/concentrations, induction/response figures, metabolic activation state, testing conditions and covariate values before being imported into PROAST. Once in PROAST, the program had started, a variable selection was performed where user defined values are placed in for the type of data considered (continuous, summary, individual, binary, ordinal, quantal, clustered), critical effect size (CES)/BMR (predefined change over control), exponential models 3 and/or 5 can be selected, each parameter for dose/response/covariates defined and the number of bootstrapping runs for model averaging

performed. This BMD modelling methodology was consistently used across chapters 2, 3, 4 and 5 where each of these parameters/conditions are defined in the results section of each chapter alongside each figure. An explanation for the parameter setting chosen for this thesis is outlined below.

2.2.5.2 – PROAST Parameters Settings

This section outlines the data type, models, variable, covariates, CES, bootstraps and other settings chosen for BMD analysis. Every endpoint for each individual chemical were kept as separate csv/txt files (chemical + endpoint) and this was repeated for all dose-response datasets. Variation in the dose-response data curation process for each assay (outside of the MultiFlow/MicroFlow assays) are explained in each of their subsequent chapters. Upon launching PROAST, each txt file was individually selected and loaded into the program and response data options were provided where continuous individual data was selected for the purposes of this thesis. Subsequently, options for single models or nested family models were provided, where the "exponential model 3 or 5 from various families of models" was chosen. The formulae applied when this option was selected is:

$$\text{Model 3: } y = a\, exp\left(\pm\left(\frac{x}{b}\right)^{d}\right) \text{ with } a > 0,\ b > 0,\ d \geq 1$$

$$\text{Model 5: } y = a\,[c - (c - 1)\, exp\,(-\tfrac{x}{b})^{d}\,] \text{ with } a > 0,\ b > 0,\ c > 0,\ d \geq 1$$

where each variable/parameter was previously annotated and described in Chapter 2.2.5. For the variable section, the independent variable selected was the dose/concentration column on the original dose-response txt file and the dependent variable selected was the response column (fluorescence induction, fold-change, etc) ensuring that robust and raw experimental figures were selected to assist in modelling the BMD. Covariates that directly contributes to the dependent variable were selected depending on their availability (i.e. study design, compounds, etc), otherwise covariates were not selected. For comparative purposes of assessing different compounds and providing additional information, an *in vitro* medium CES (equivalent to the BMR) of +0.5 was chosen (50% increase above the control point of the dose response curve), with exception to the MN endpoint which was recorded at +1.0 and polyploidy at +5.0 (100% and 500% increase above the control respectively) as they are biologically relevant set values for each marker of cytotoxicity, however, the CES for each assay will be listed in their respective sections (Zeller et al., 2017; Bemis et al., 2016). Upon

reaching the model averaging section, a default choice of 200 bootstraps is applied to the existing data where the resamples of the same datasets are used to generate BMDs and their respective CI and the full model is fitted unless the average group size is 1. These parameters allow for the lower and upper bounds of the BMD CI (BMDL and BMDU), with a 90% degree of confidence, to be generated and model averaging is visualised graphically as individual dose-response curves with $\log_{10}$(dose) and the $\log_{10}$(response) were mapped out. The shape of the curve, POD, and extrapolated figures were assessed for positive/negative BMD generation results and BMDL/BMDU figures were carried forward for ToxPi visualisation and analysis, however, if a dataset was unable to generate a BMD/model averaging, no BMD was recorded. The figure below illustrates how PROAST depicted these dose-response curves:

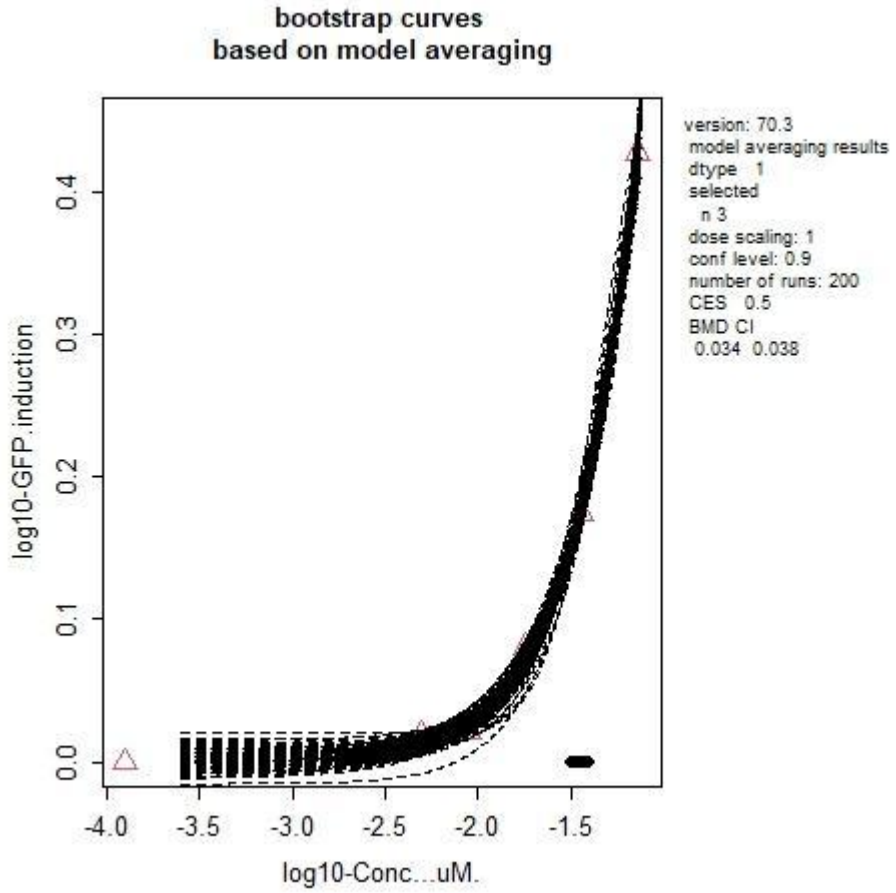


Figure 2.2.5.2.1 – An individual dose response curve generated by PROAST v70.3 for Cisplatin S9- Blvrb endpoint in the ToxTracker assay. The raw concentration and GFP induction figures had the exponential models 3 and 5 applied to them, and a user selected CES of 0.5 with 200 model averaging runs. This generated 90% CI BMD values as the BMDL (0.034) and BMDU (0.038).

To summarise how PROAST is used to perform BMD analysis, raw dose-response datasets were generated, and compounds/endpoints segregated into individual txt files. Next, the dose-response figures for each compound/endpoint were fitted with their appropriate four-parameter exponential models (model 3 and 5). The independent and dependent variables (dose and response respectively) were selected with appropriate log scaling, and compounds/study design were used as covariates (unless unavailable). Next, the CES of +0.5 was selected for both the analysis topoisomerase II inhibitors and extended aneugen/clastogens assessments and +0.3 for the metabolic activation of clastogens section. Meanwhile, CES values of +1.0/5.0 remained consistent for MN/polyploidy respectively across all sections in this chapter. Model averaging was applied through 200 bootstrap runs to generate the BMD CIs for each compound/endpoint. Finally, group size determined whether the full models were fit with the dose-response data and the dose-response curve was mapped to find each compound/endpoint's BMDL/BMDU, which were curated in an excel file ready for analysis using ToxPi.

2.2.6 – Significance of model averaging to generate Confidence Intervals

In traditional statistics, no computer-simulated data processing was used and therefore tests had to be performed before assumptions are made from smaller sample sizes before drawing final conclusions from experiments. An example of this would be T-testing, an inferential method used to verify whether any statistical significance could be drawn from the means of two samples/groups that assume a normal distribution curve with unknown variance. However, when managing statistical significance figures for BMDs, a T-test may not be sufficient unless 100s or 1000s or replicates are run and more data is collected, even if the data collected continues to follow a normal distribution curve. Running more replicates prevents too many assumptions from being drawn from data, however, biased or imprecise results may arise when requirements for replicate counts are not met, which would also be expensive and time-consuming if 1000s of samples are individually taken (Zhu et al., 2007). PROAST supports model averaging as an alternative to singular hill/exponential modelling to produce BMD figures to incorporate uncertainty associated with whichever models are implemented. This allows for increased robustness as it considers variability, reducing modelling biases, considers how well the dataset fits with the applied models using Akaike/Bayesian information criterion information, and quantifies uncertainty surrounding the BMD, which directly reflects the variation surrounding the fitted models (RVIM, 2021).

Bootstrapping is a statistical procedure where sampling with replacement is applied to simulate distributions based on a smaller subset of already gathered data to generate replicate estimate distribution values for a sample dataset (Zhu et al., 2007). This is used to obtain multiple resamples of the same dataset and therefore is used to obtain CI BMDL/BMDU values from dose-response datasets. BMD values can be obtained from BMDL and BMDU combined metrics, resampling the data provides increased efficiency as opposed to manually obtaining data as the results produced are statistically alike to the resampled values and it can distinguish between likely and unlikely sample data by showing where 95% of the data lies in ToxPi and PROAST (Zhu et al., 2007; Marvel et al., 2018). Values are randomly chosen and sampled with replacement (where duplicated values for the raw data are included) and each sample calculates a BMD until an 95% estimate BMD CI can be found from 1000 bootstrap runs; ToxPi automatically performs this task to find the upper and lower bounds of the ToxPi/slice scores (Marvel et al., 2018; Zhu et al., 2007). However, as PROAST is utilised earlier for BMD analysis (to produce 90% BMD CIs), ToxPi would also calculate 90% BMD CIs.

2.2.7 – ToxPi for visualisation and data dimensionality reduction and filtering

An analytical framework is required to convert the PROAST generated BMD figures into appropriate visual profiles. This framework must ensure that the procedures it follows must be flexible and versatile so it can be customised and applied to different tasks/assays/analyses to assist in drug safety assessments in genotoxic research. Eventually, assessment decisions that prioritize/rank chemicals against each other and their respective groupings will need to be made (Committee on the Design and Evaluation of Safer Chemical Substitutions: A Framework to Inform Government and Industry Decision, 2014).ToxPi: a JavaScript program that allows visualisation and interaction of the quantitative data relating to mutagenic compounds, to allow for the mutagenicity of various compounds and the quantitative data to be translated into a single ToxPi score, which ranks and scales the mutagenicity on a continuous spectrum instead of fixed groups to discriminate between potency groups (Reif et al., 2012). K-means and Hierarchical clustering techniques can be performed on ToxPi for vector quantization using automated PCA techniques and to reduce dataset dimensionality and using divisive/agglomerative nesting (AGNES) algorithms to place chemicals into grouped clusters based on similarity in ToxPi scoring and overall potency across the endpoints, allowing us to effectively interpret these datasets (Marvel et al., 2018). It is a flexible program that allows various datasets across assays to be accurately visualised and

provide insight on toxicological potency across a series of compounds. For all generated data across all chapters, any BMDs generated all assays had been transformed so that they would fit within the concentration range of 0.0001 – 10000 μM, hence any BMD below 0.0001 μM are converted to 0.0001 μM and any value above 10000 μM is converted to 10000 μM. These are operational guidelines set in place as:

A) No living cells were found at concentrations exceeding 10000 μM, so a degree of biological transparency requires consideration.

B) PROAST calculated BMD on a mathematical basis through modelling algorithms and formulas so biological relevance/transparency is required when translating the PROAST values into realistic figures that can be used for comparative biological drug toxicity analysis (a combination of mathematical modelling and biological considerations).

C) For regulatory and safety reasons, compounds are not tested below 0.0001 μM as BMDs estimates below this concentration indicate extremely high biological potency. Consequently, additional testing at lower concentrations is unlikely to be meaningfully refine compound potency characterisation while approaching the practical limits of *in vitro* dosing precision and biological interpretability.

D) Datasets with large disparities in BMD figures could introduce heavy skewing during the visualisation process, thereby reducing comparative capabilities of ToxPi, especially between compounds within the same assay exhibiting relatively weaker or stronger genotoxic properties.

E) Uncapped ranged formed indistinguishable projections when assessing endpoints, therefore recoding to set limits was deemed appropriate.

During this investigation, a collaboration led alongside Litron had introduced the opportunity for further transformations to be applied. For example, the utilisation of minimum and maximum values being added as extra columns to the original BMD txt files generated by PROAST before processing through ToxPi and using integral calculations (i.e. trapezium rule) generated from the original dose-response figures to evaluating area under the curve (AUC) values before ToxPi processing commences.

When visualising the dose-response BMD datasets, the appropriate information on compounds, endpoints and each respective BMDL/BMDU figure must be appropriately spaced columns. Once, spaced, the excel file must be converted to a CSV file for processing through ToxPi and can be imported into the program. ToxPi displays all chemicals alongside all endpoints that are assessed in both the “available chemicals” and “included metrics” ToxPi columns where compounds may be individually selected (or all may be selected using the “select all” function), and their endpoints are imported into the “overlapping metrics” column showing all endpoints. Compounds may be added or removed; however, endpoints require removal from the original file if undesired for analysis. Once the “create slices” (or “slices” tab) option had been selected, individual endpoints can be customised and for each slice, the BMDL and BMDU are added from the “available metrics” into the “selected metrics” column using the “add” function for each metric and more slices may be introduced using the “add” function by the slices section. The colour coordination and “show preview” checkbox reveals how the ToxPi profiles would appear as a finalised visual. Next, individual weightings can be user set in integer folds (with shown percentages) for each slice/projection and mathematical transformations can be applied using the available subheadings in the “y=” function, including functions:

1. $y = linear(x)$
2. $y = hitcount$
3. $y = -\log_{10}(x)$
4. $y = -\log_{10}(x) + \log_{10}(max(x))$
5. $y = -\log_{10}(x) + 6$
6. $y = -ln(x)$
7. $y = \log_{10}(x)$
8. $y = sqrt(x)$
9. $y = zscore(x)$
10. $y = uniform(x)$

where the $y = -\log_{10}(x)$ transformation is applied. This is to ensure logarithmic scaling is applied, to prevent unnecessary skewing and visualisation while covering a wide BMD range. Also, as compounds with higher potencies would present with lower BMD figures, a negative function requires application, to ensure that compounds with low BMDs appear as having higher potency values as lower concentrations of that chemical increase toxicity incidence

and higher ToxPi scores/endpoint projections. Furthermore, chemicals with higher BMDs would show as having lower potency figures indicating reduced compound toxicity and lower ToxPi scores/endpoint projections. Finally, the "calculate ToxPi scores" option allows for the BMD dataset to be visualised through the program.

Visualisation of the data runs immediately and the "compute bootstrap CI" option allows for the CI to be visualised. Firstly, ToxPi calculates its own "ToxPi score" by the sum of all input values across all assays being subsequently transformed into a slice/component score for each compound (Auerbach et al., 2016). Next, the combined input values were normalised to values between 0-1 by subtracting the minimum value and dividing by the range (Auerbach et al., 2016). The values were then multiplied by the user selected (integer) fold-weight value (or by 1 ÷ the number slices if the weighting applied to each slice were all equal) to give an overall component score (Auerbach et al., 2016). Finally, all component scores are added together and ranged between 0-1 for the final ToxPi score for that compound, where compounds closer to 1 are relatively more potent and compounds closer to 0 are relatively less potent at inducing toxicity. Each dataset retained their respective slice scores for individual endpoints for each compound, which are displayed as values ranging between 0-1. The formulae below explain the ToxPi score production process (where "*xi*" represents individual input values for the assay, "min(x)" and "max(x)" represent the minimum and maximum values respectively, "*wi*" represents the weight, "*n*" represents the number of slices, and "*i*" stands for the component score for each slice):

Normalisation:

$$xi' = \frac{x - \min(x)}{\max(x) - \min(x)}$$

Weighting:

$$i = xi' \times wi \; or \; \frac{1}{n} (if \; weights \; are \; equal)$$

Summing the component scores:

$$ToxPi \; Score = \frac{\sum_{i=1}^{n} (xi' \times wi)}{\sum wi}$$

Figure 2.2.7.1 – The formula capturing the steps involved in calculating the ToxPi score adapted from (Auerbach et al., 2016).

Finally, 90% CIs were generated from 1000 automated bootstrap runs (with replacement) for both ToxPi scores and slice scores to display the region of uncertainty around each figure to yield compound potency distribution and highlight any significant differences in overall potency. Simultaneously, ToxPi visualises the potency for each dataset and translates them into a transparent ranking chart comparing ToxPi score to ToxPi ranks for all compounds to facilitate decision making and hazard assessment by allowing the direct comparison of chemical toxicity in datasets (Committee on the Design and Evaluation of Safer Chemical Substitutions: A Framework to Inform Government and Industry Decision, 2014).

ToxPi also effectively generates visual profiles depicting both the ToxPi scores and slice scores as ToxPi profiles: an automatically generated pie chart that displays all characteristics of each dataset as different components and features of these profiles. Firstly, the magnitude of each slice is representative of the relative potency for an individual chemical at a specific endpoint (with the $-\log_{10}(x)$ transformation being applied). Paler sections within each ToxPi slice are representative of the 90% CI and the larger that section, the larger the CI where the white line drawn at the centre it the actual slice score. The inside of the CI section represents the lower CI, depicting the position of the BMDU, and the outer section of the CI represents the upper CI, depicting the position of the BMDL (due to $-\log_{10}(x)$ transformation (compounds with lower BMDs have higher potency/toxicity)). Next, the width of each slice is representative of the weighing applied, hence the wider the slice, the more emphasis is placed on that endpoint in terms of importance and the data there is considered more compared to other endpoints. Finally, the grey section at the centre of the ToxPi profile is representative of the amount of missing data for that endpoint, where the amount greying is directly proportional to the amount of absent data. However, numerical determination of missing data requires assessment of the original CSV file. These features may all be observed with colour coordinating/zooming/etc features and numerical figures are shown for each slice score/CI in ToxPi itself and can be saved as PNG/PDF/SVG files.

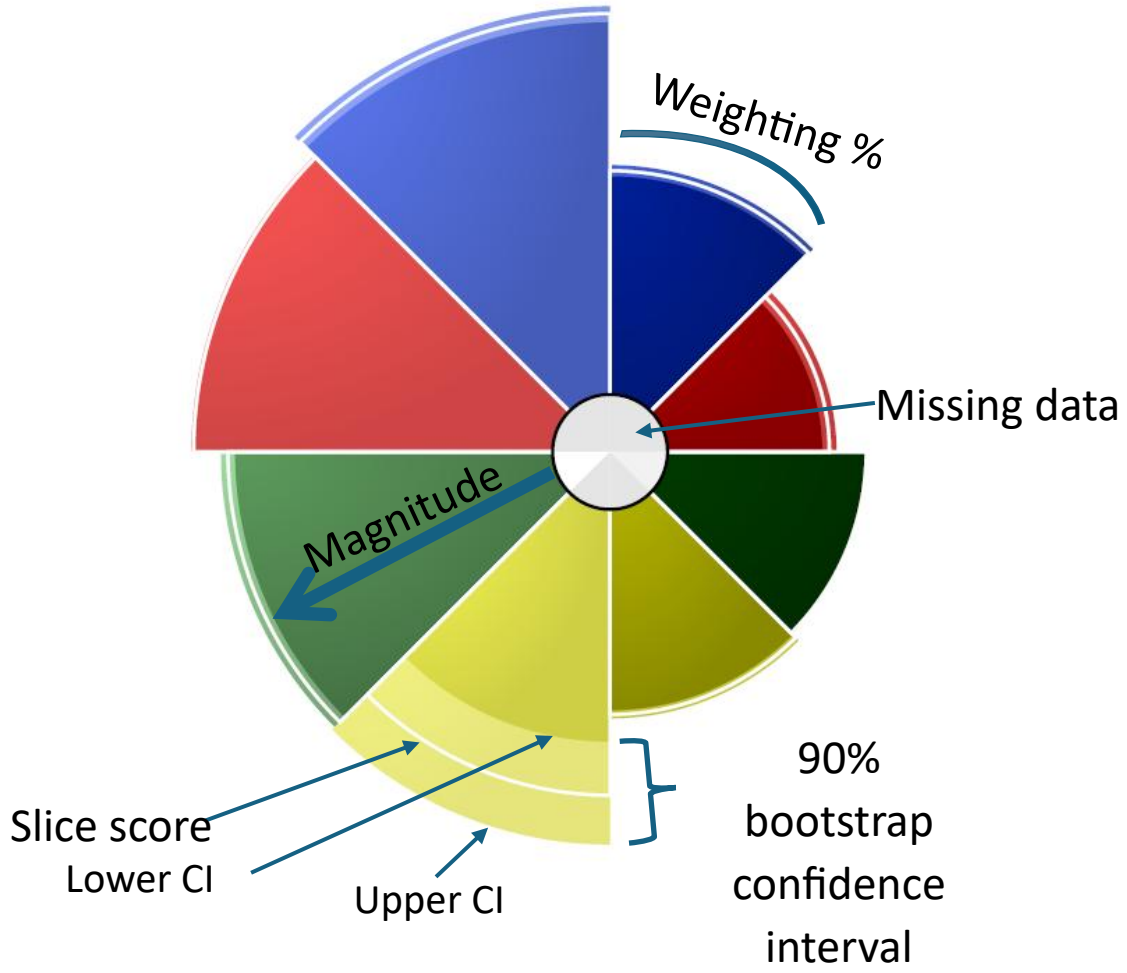


Figure 2.2.7.2 – ToxPi Profile key showing multiple slices each representing a biomarker for a compound. Features include the magnitude which represents the BMD CIs where the pale region represents the 90% bootstrap confidence interval surrounding the BMDL and BMDU highlighting uncertainty with the white line within that region representing the slice score. Additionally, the width of each slice represents the weighting applied to each biomarker and greying the centre of the profile illustrates if there is any missing data, where more absent data is directly proportional to darker shading of this area.

2.2.8 – Unsupervised Machine Learning Algorithms: Hierarchical Clustering

ToxPi has also been used and developed alongside machine learning algorithms to perform hierarchical clustering: a method of clustering that seeks to build hierarchies of clusters using AGNES methods to determine dataset groupings without requiring human intervention or training data to build on these clusters (Marvel et al., 2018). These clusters are formed through data patterns that display similarities and differences as each object is a single element cluster and at each step, two similar clusters combine into a larger one which results in the formation of a dendrogram (Hirano et al., 2004). Hierarchical clustering is a widely used for grouping biomarker genes and drug toxicity (Hasan et al., 2019). AGNES is a bottom-up approach where each object has an independent cluster assigned to it and seeks the most similar pair of clusters before merging into one and it repeats until all initial clusters are merged into single clusters (Hirano et al., 2004). Meanwhile, the divisive is a top-down approach where a single cluster containing all objects are divided into smaller clusters (the opposite to the AGNES approach) (Hirano et al., 2004). Hierarchical clustering module organises ToxPi profiles into clusters based on similarities and differences using the bottom-up AGNES approach here and the clustering method choices available for Hierarchical clustering in ToxPi include those available in the R hcust function: complete, average, single, mcquitty, ward.D and ward.D2 (Marvel et al., 2018). Hierarchical clustering defines a distance metric between objects using Euclidian distance and by repeatedly joining and merging objects, a cluster is formed between two objects with the shortest distance between them (Hasan et al., 2019). The Euclidian distance is determined by the number of objects available and the squared distance between each which follows the formula (adopted from (Hasan et al., 2019)):

$$dGi, Gi' = \sqrt{\sum_{j=1}^{m} (\mathrm{F(Gi, Cj)}^2 - (\mathrm{F(Gi', Cj')}^2)}$$

This is the formula for two datapoints that form a cluster. For further clustering across more datapoints/objects, the Euclidian distance follows (adapted from (Gao et al., 2023)):

$$d(p, q) = \sqrt{\sum_{i=1}^{n} (p_1 - q_1)^2 + (p_1 - q_2)^2 + \cdots + (p_i - q_i)^2 + \cdots + (p_n - q_n)^2}$$

The hierarchical clustering algorithm selected contain multiple methods with differing formulae and approaches, each making changes based to the clustering. One-way which toxicogenomic data is clustered based on similarities/differences is Single linkage hierarchical clustering, where the smallest distance between two objects merge and form a cluster (Marvel et al., 2018). Another approach used is Complete linkage, which instead clusters two toxicogenomic objects when their distance is at its largest (Hasan et al., 2019). In Average linkage, the distance between two clusters is defined as the average between all object pairs where each pair is made of one object from each group and in (Hasan et al., 2019). In Mcquitty clustering, the average of cluster distances is taken without considering the number of points in that cluster, meaning that once two clusters merge, then the distance between them would be the arithmetic average of each cluster with any other cluster (Mcquitty 1966; Charrad et al., 2014). Ward's clustering, it calculates the increase in the error sum of squares after fusing two clusters to minimise information loss (Clifford et al., 2011; Hasan et al., 2019). For the following formulae, $A \cup B$ representing newly merged clusters, and another cluster noted as C. Furthermore, the $d(\cdot,\cdot)$ function represents distance and the number of observations in clusters is denoted as X (Fisher, 1958; Jenks, 1977; Murtagh & Legandre, 2014).

Single-Linkage Hierarchical Clustering – The minimum distance between clusters to be merged found and merge the corresponding objects to form a cluster to produce long chains (Hirano et al., 2004; Hasan et al., 2019).

$$d(\mathrm{A} \cup B, C) = min_{x \in A \cup B, y \in C}\, d(x, y)$$

Complete Hierarchical Clustering – The maximum distance is found between clusters as the AGNES algorithm first merges based on the minimum distance between corresponding objects and finally merge based on the maximum computed distance, forming spherical clusters with consistent diameters (Hirano et al., 2004; Hasan et al., 2019):

$$d(\mathrm{A} \cup B, C) = max_{x \in A \cup B, y \in C}\, d(x, y)$$

Average-Linkage Hierarchical Clustering – Treats distance between two clusters as the average between all pairs of objects where one member of a pair belongs to each cluster, which is affected less by outliers (Hasan et al., 2019):

$$d(\mathrm{A} \cup B, C) = \frac{\sum_{x \in A \cup B} \sum_{y \in C} d(x, y)}{\mathrm{A} \cup B| \cdot |C}$$

Mcquitty Hierarchical Clustering – The distances between two clusters C1-C2 are the weighted average distance between each individual cluster and another cluster (Mcquitty 1966; Charrad et al., 2014):

$$d(\mathrm{A} \cup B, C) = \frac{d(A, C) + d(B, C)}{2}$$

Ward D Hierarchical Clustering – Based on the minimum increase of the sum of squares within a cluster to reduce the total variance within a cluster and at each step, the pair of clusters with the minimum distance is merged (Ward, 1963; Hirano et al., 2004; Murtagh & Legandre, 2014):

$$d(\mathrm{A} \cup B, C) = 2 \cdot \frac{|\mathrm{A} \cup B| \cdot |C|}{|\mathrm{A} \cup B \cup C|} \cdot \left\| \frac{\sum_{x \in A \cup B} x}{|\mathrm{A} \cup B|} - \frac{\sum_{y \in C} y}{|C|} \right\|^2$$

Ward D2 Hierarchical Clustering – Works similarly to Ward D method, however, considers the unsquared Euclidian distance between the centroids of each cluster by taking the square root of the mean relevant clusters, while Ward D considers the squared Euclidian distance between individual datapoints and the average of the combined cluster (Ward, 1963; Murtagh & Legandre, 2014):

$$d(\mathrm{A} \cup B, C) = \sqrt{2 \cdot \frac{|\mathrm{A} \cup B| \cdot |C|}{|\mathrm{A} \cup B \cup C|}} \cdot \left\| \frac{\sum_{x \in A \cup B} x}{|\mathrm{A} \cup B|} - \frac{\sum_{y \in C} y}{|C|} \right\|^2$$

ToxPi allows for these algorithms to be applied to dose-response translated datasets to visualise the automated Hierarchical clustering produced, using the Java port of the R function *hclust* to achieve this functionality (Marvel et al., 2018).

2.2.9 – Unsupervised Machine Learning Algorithms: K-Means Clustering

ToxPi also programmed with its own automated system of implementing K-means clustering through the combination of divisive and PCA algorithms based on similarities and differences of ToxPi scores/profiles (Reif et al., 2012). Unlike hierarchical clustering, the divisive

algorithm used in K-means is a clustering method of vector quantisation that utilises a top-down approach to distinctly group chemicals based on a user defined number of colour and shape coordinated clusters using the "Hartigan-Wong" algorithm from the Java port of the *kmeans* R function (Marvel et al., 2018). Unsupervised K-means clustering reduces dataset dimensionality and uses PCA algorithms to cluster compounds by placing them into a simple graphical plot with principal component 1 (PC1) on the x-axis and principal component 2 (PC2) on the y-axis to visualise the data. Cluster centroids are determined by the arithmetic mean of the data surrounding said cluster using Euclidian distance and the closest data points surrounding each cluster are assigned to each cluster respectively (Zhu et al., 2021). PCA measures the amount of variation among datasets reducing the attributes and features of a dataset whilst maximising information retention. Data here is divided into two principal components, where PC1 is the component from the dataset that shows the greatest variance across the samples by some scalar projection and PC2 is the component that shows the second greatest variance across the samples (Maugeri et al., 2021). K-means clustering will be used extensively through this thesis and provides further insight into dose-response datasets through data noise reduction and comparing compound clustering based on ToxPi score for PC1 and biomarkers analysis for PC2 across all assays.

2.3 - Results

The results section covers the data generated in this chapter, including:

1. The genotoxic assessment of 8 topoisomerase II inhibitors by compared to their ToxPi potency ranking/clustering QSAR chemical structures.
2. The comparison of parent compound to metabolite across 10 clastogens with varying MoAs
3. The analysis of 30 test aneugenic/clastogenic test chemicals to evaluate capability of PROAST/ToxPi to analyse larger datasets of parent compounds and metabolites datasets, to determine whether increasing dataset diversity influences MoA groupings and the identification of genotoxic/non-genotoxic compounds.

The results figures generated from ToxPi are highlighted in this chapter's subsections across all 3 components, highlighting the ToxPi ranking, scores, profiles and both hierarchical/K-means clustering. Alongside this, any original figures included from each respective component are appropriately chosen and have their sources cited.

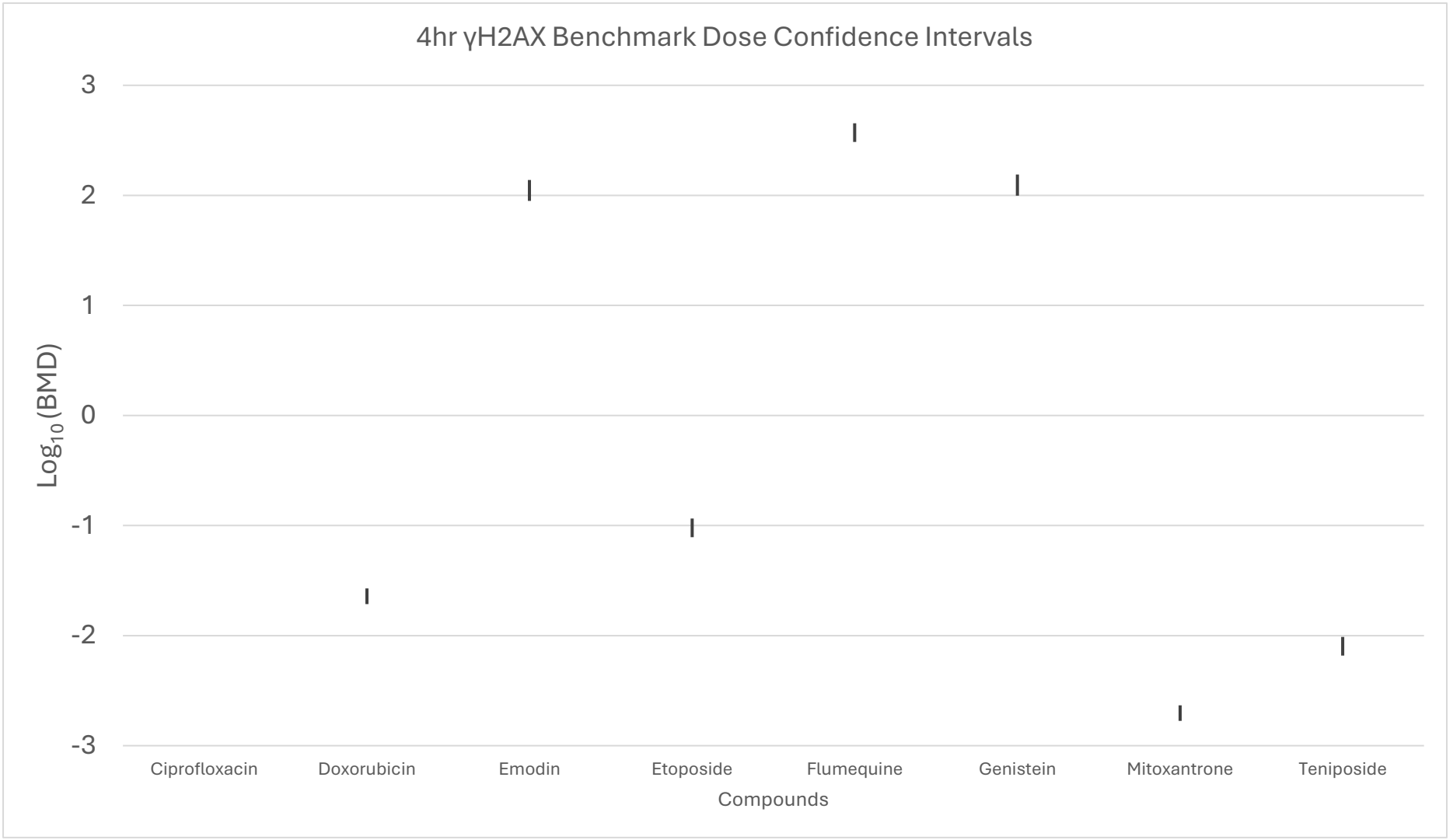


Figure 2.3.1.1 – A 90% BMD (μM) CI plot for 4-hour γH2AX response in TK6 cells exposed to eight topoisomerase II inhibitors. Ciprofloxacin did not generate a dose-response.

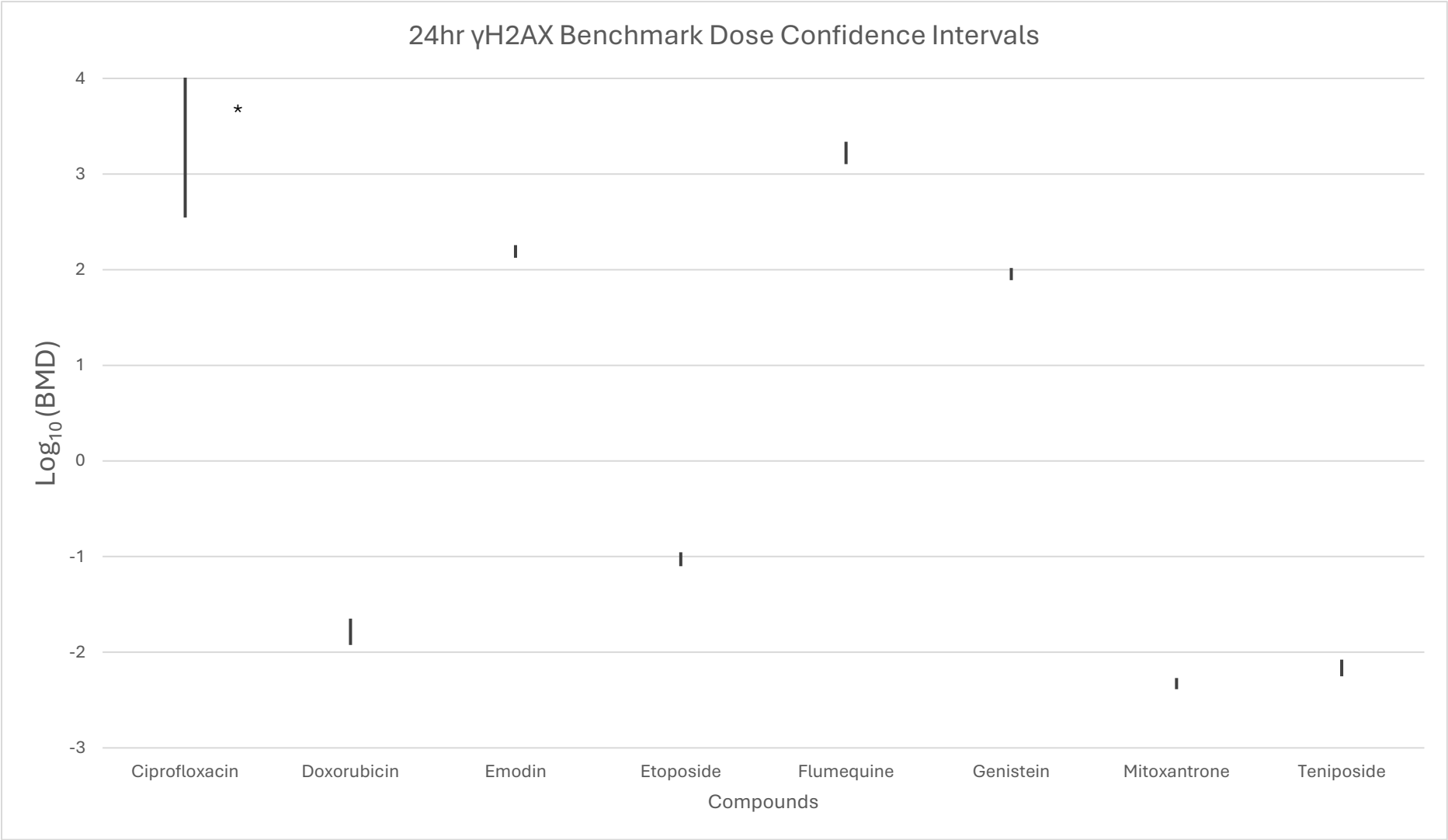


Figure 2.3.1.2 – A 90% BMD (µM) CI plot for 24-hour γH2AX response. Despite a generated BMD for ciprofloxacin at the 24-hour time point, the upper confidence interval is infinite, represented by an asterisk.

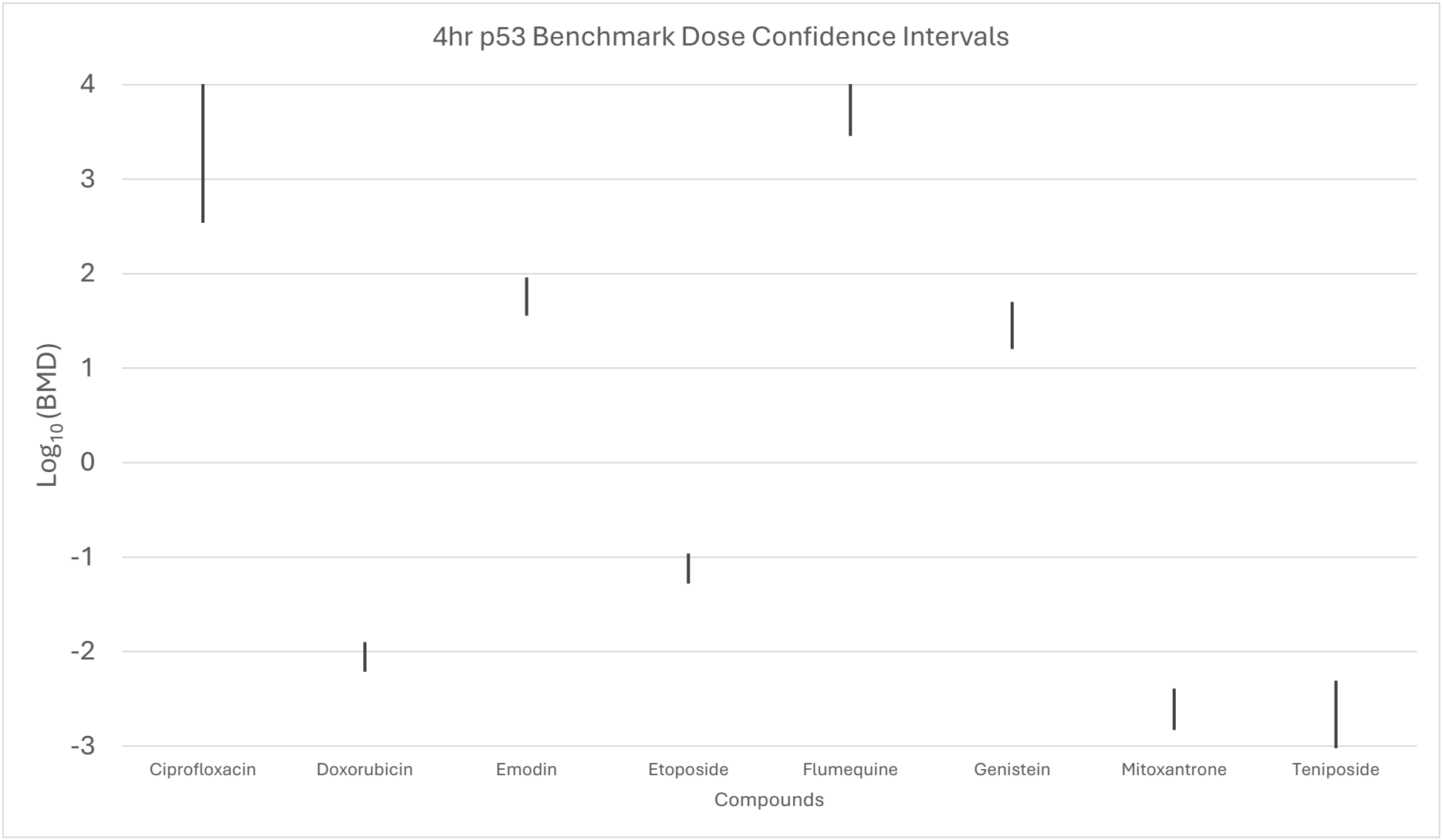


Figure 2.3.1.3 – A 90% BMD (μM) CI plot for 4-hour p53 response, measuring dose-response in the MultiFlow assay.

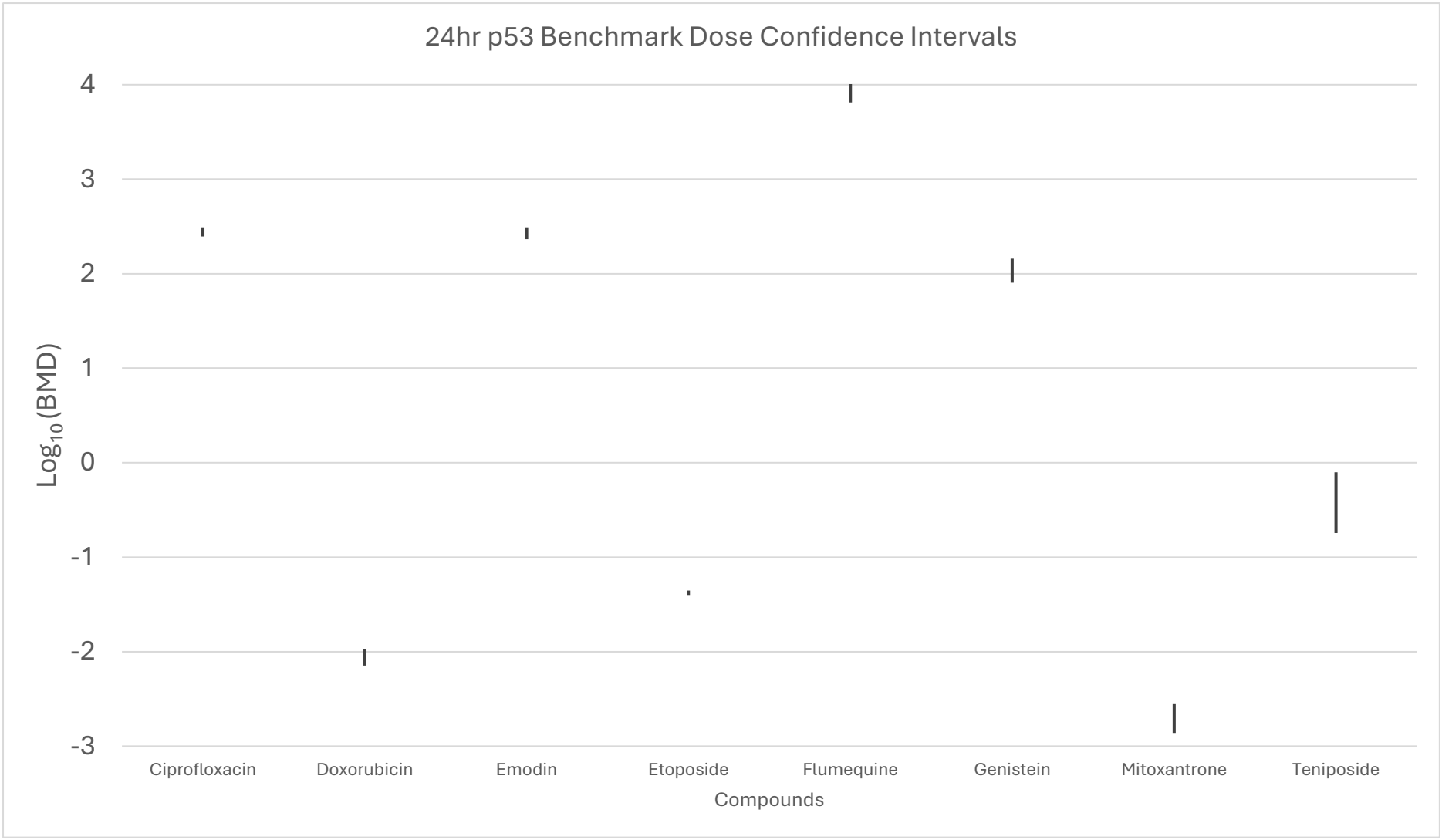


Figure 2.3.1.4 - A 90% BMD (μM) CI plot for 24-hour p53 response, measuring dose-response in the MultiFlow assay.

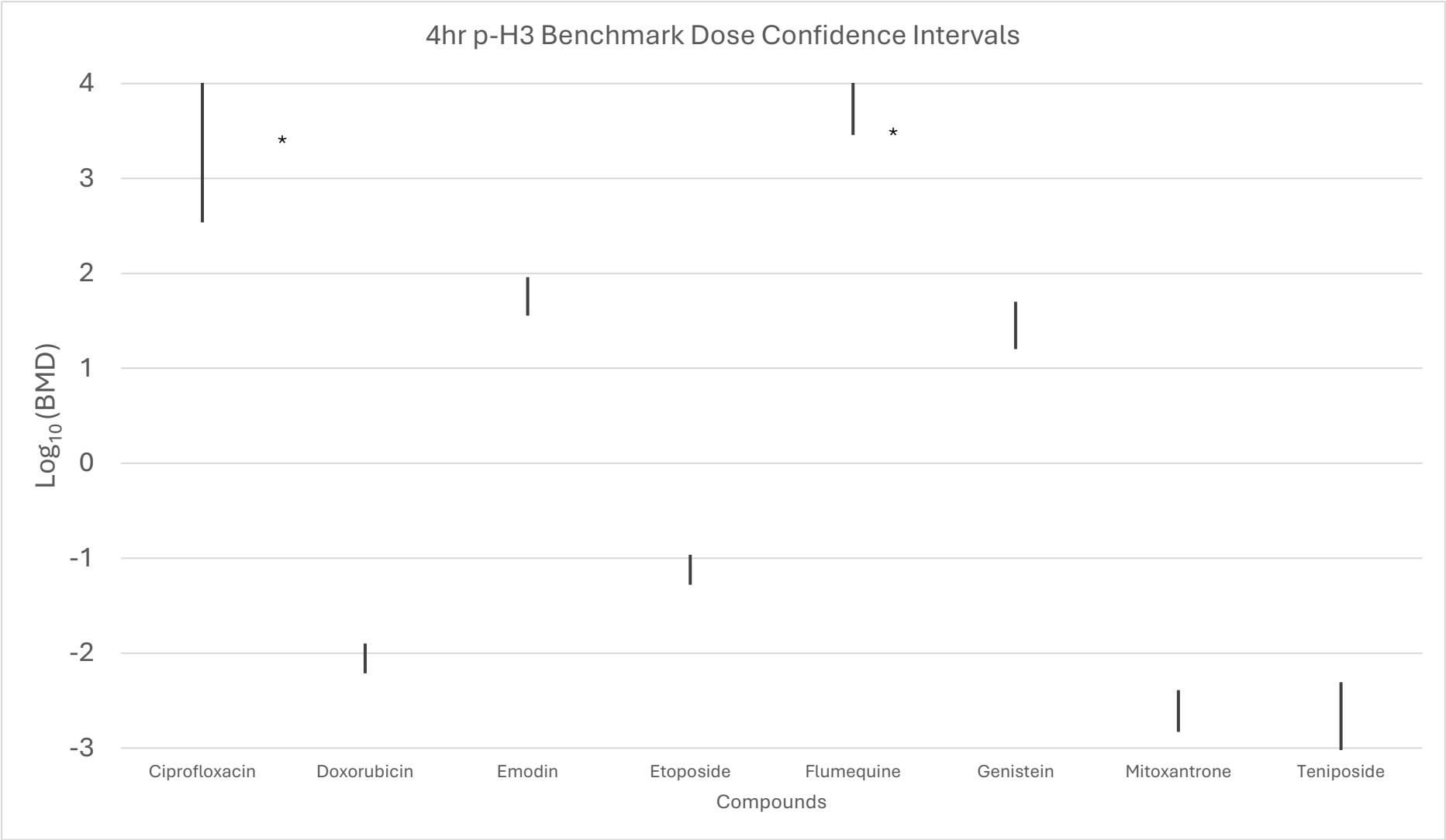


Figure 2.3.1.5 - A 90% BMD (μM) CI plot for 4-hour p-H3 response, indicates infinite upper bound values, marked by asterisks, for both flumequine and ciprofloxacin at the 4hr p53 response mark.

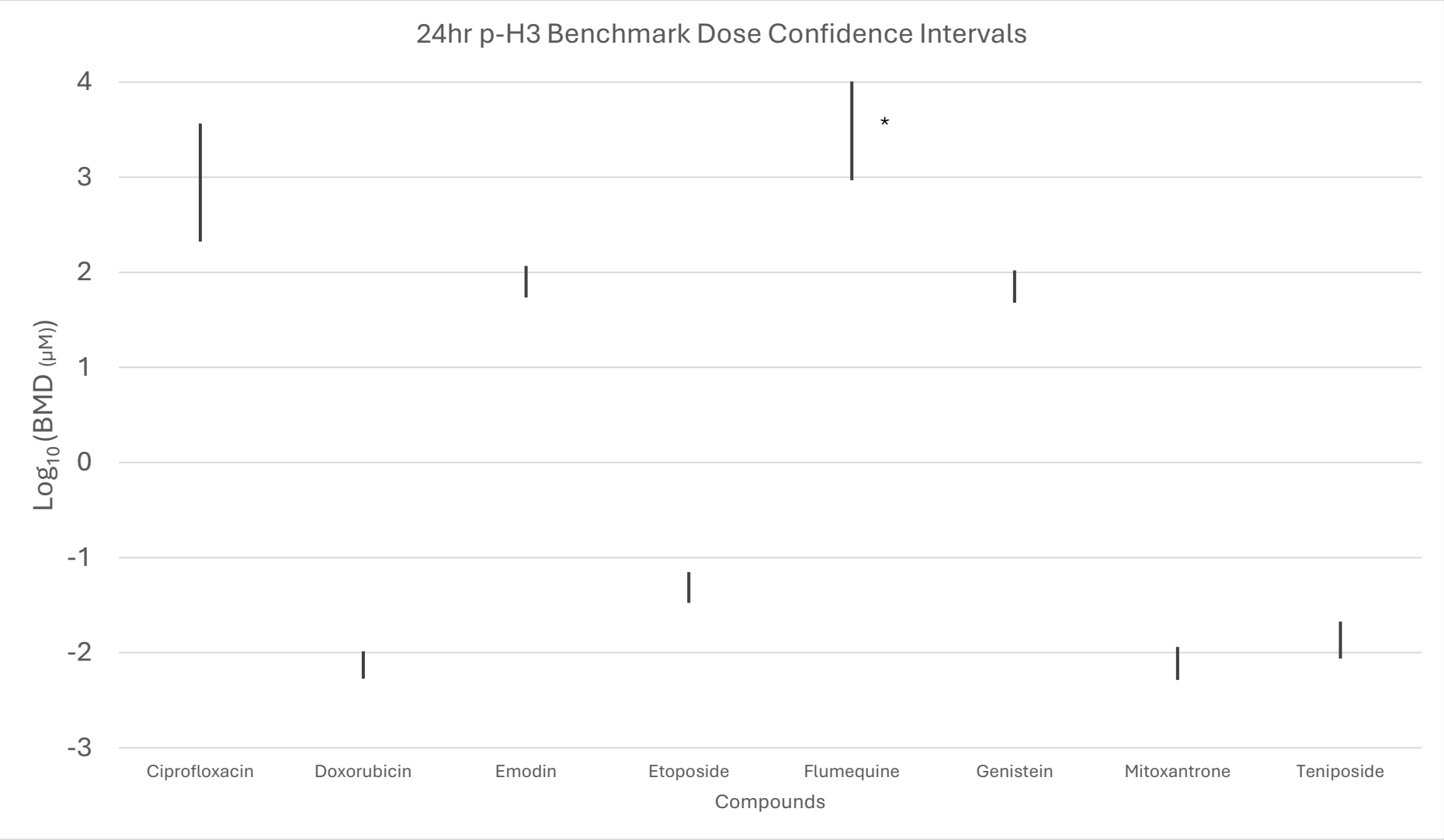


Figure 2.3.1.6 – A 90% BMD (μM) CI plot for 24-hour p-H3 response, measuring dose-response in the MultiFlow assay, where flumequine's upper bound was considered infinite, highlighted by an asterisk

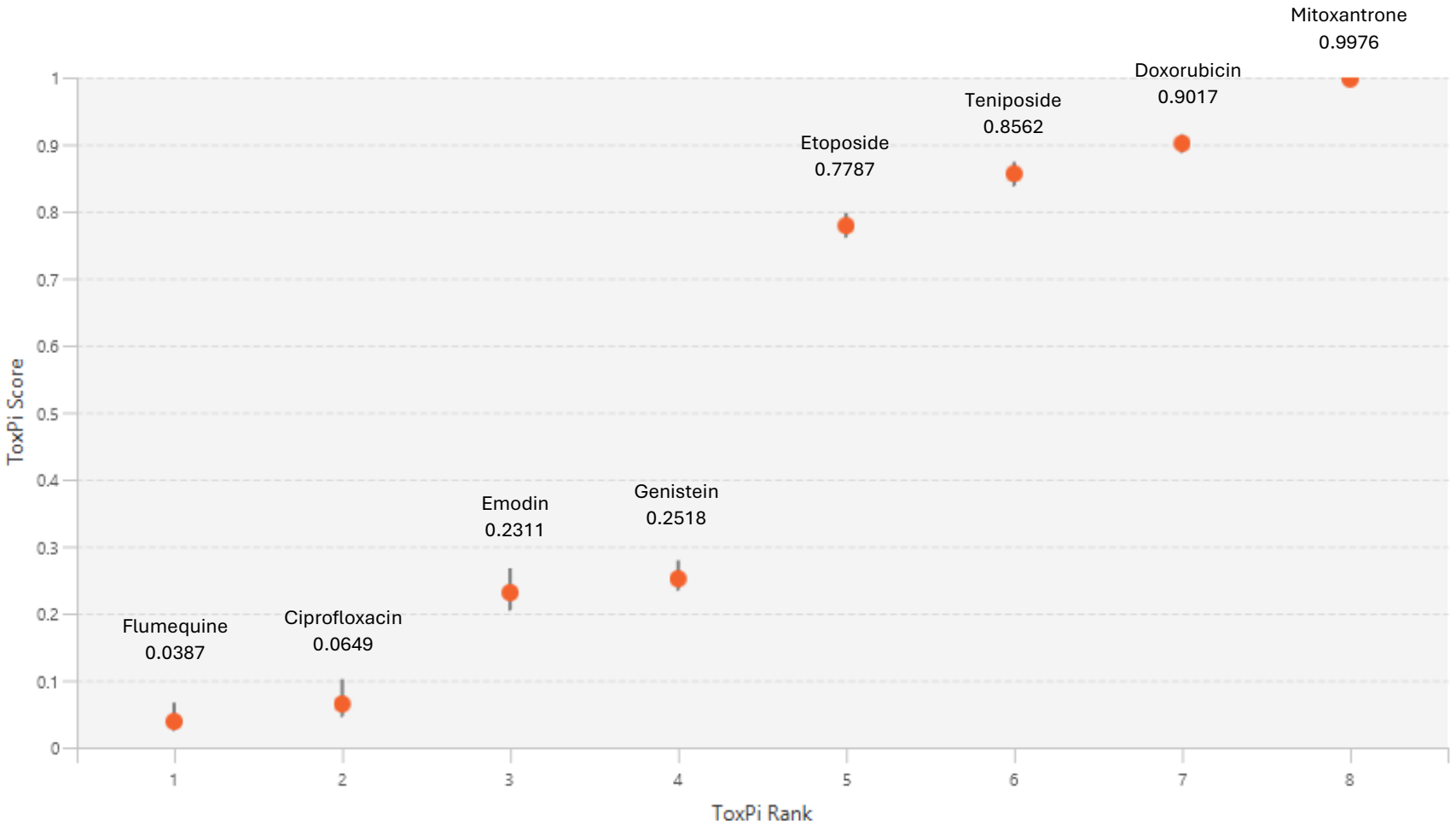


Figure 2.3.1.7 – The ToxPi potency ranking results for component/objective 1 illustrating the response of TK6 cells to topoisomerase II inhibitor exposure across 8 compounds. As described in section 2.2.7, the ToxPi scores on the y-axis ranges from 0-1 and represent a normalised set of dataset-dependent values from integrated genotoxicity endpoints. Higher ToxPi scores corresponds to greater relative genotoxic potency within a compound rather than an absolute measure of potency, where the error bars represent the CI of each compound. Therefore, in this dataset, Flumequine exhibited the lowest relative potency and Mitoxantrone has the highest.

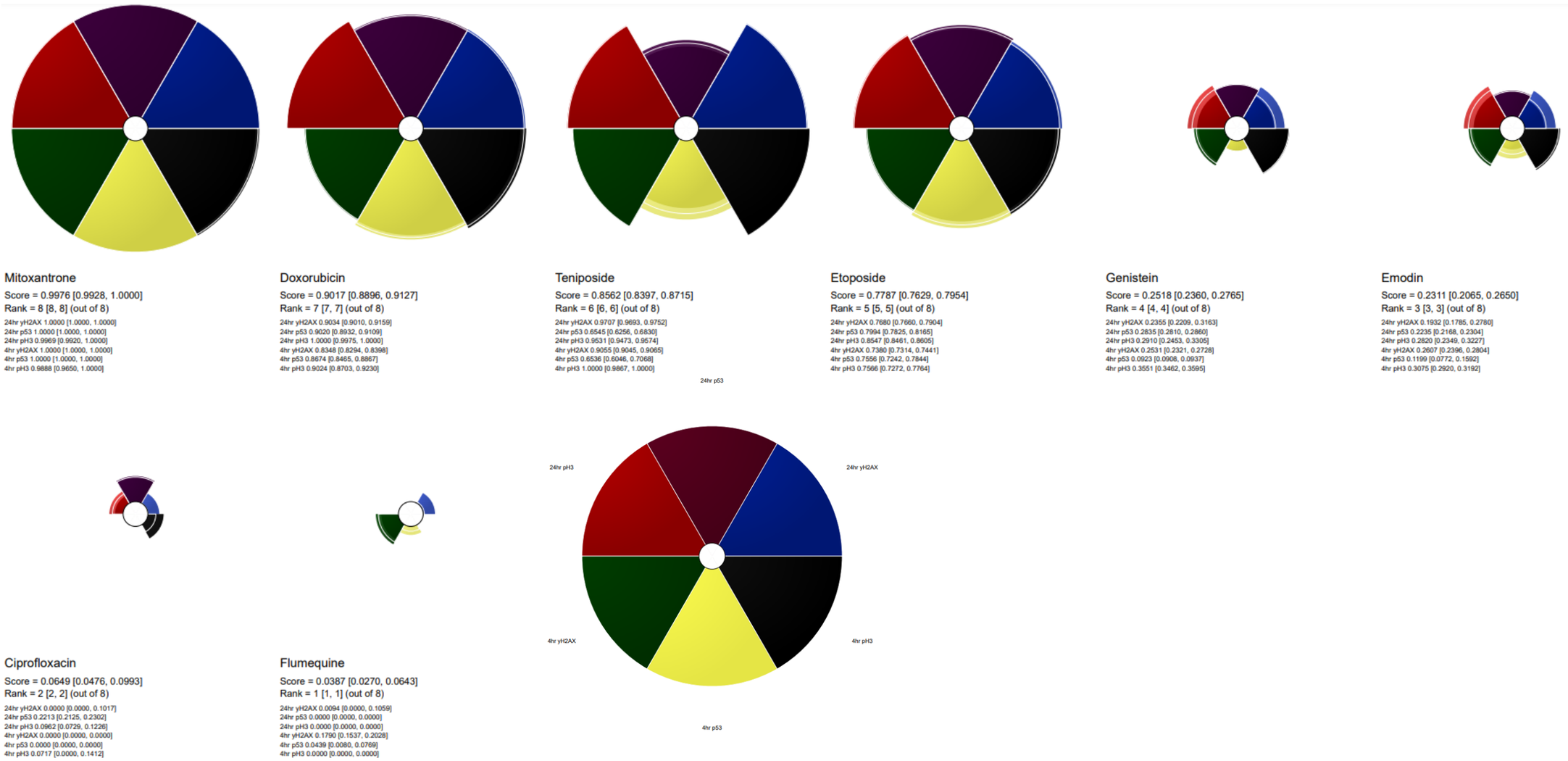


Figure 2.3.1.8 – ToxPi profiles for component/objective 1 generated from BMD data for each chemical, showing the biomarkers and their ToxPi scoring for each end point. Each slice represents a biomarker at a given time. Shading in the centre of the circle represents uncertainty for each slice, where darker centres show more uncertainty. The magnitude of each slice (to either the edge or white line at the centre of the CI) represents the relative potency of each biomarker, and widths represent weightings. The brighter regions presented in each is the bootstrap CI, demonstrating the degree of uncertainty surrounding the true BMD. Overall profile size varies according to the normalised ToxPi score; therefore, smaller profiles reflect lower relative compound potency within the dataset rather than a reduced signal at any given endpoint. These profiles highlight the individual reporters and their corresponding scores below each figure (4/24hr p-H3, 4/24hr yH2AX, and 4/24hr p53). Additionally, a ToxPi profile key illustrating the colour coordinated arrangement of MultiFlow DNA damage biomarkers at 4hr and 24hr timepoints as slices and the magnitude for each slice represents the fold change values of the median fluorescence intensity generated by propidium iodine flow cytometry, which are modelled into BMD figures that have their CIs -log10(x) transformed to construct the following ToxPi slice profiles accordingly.

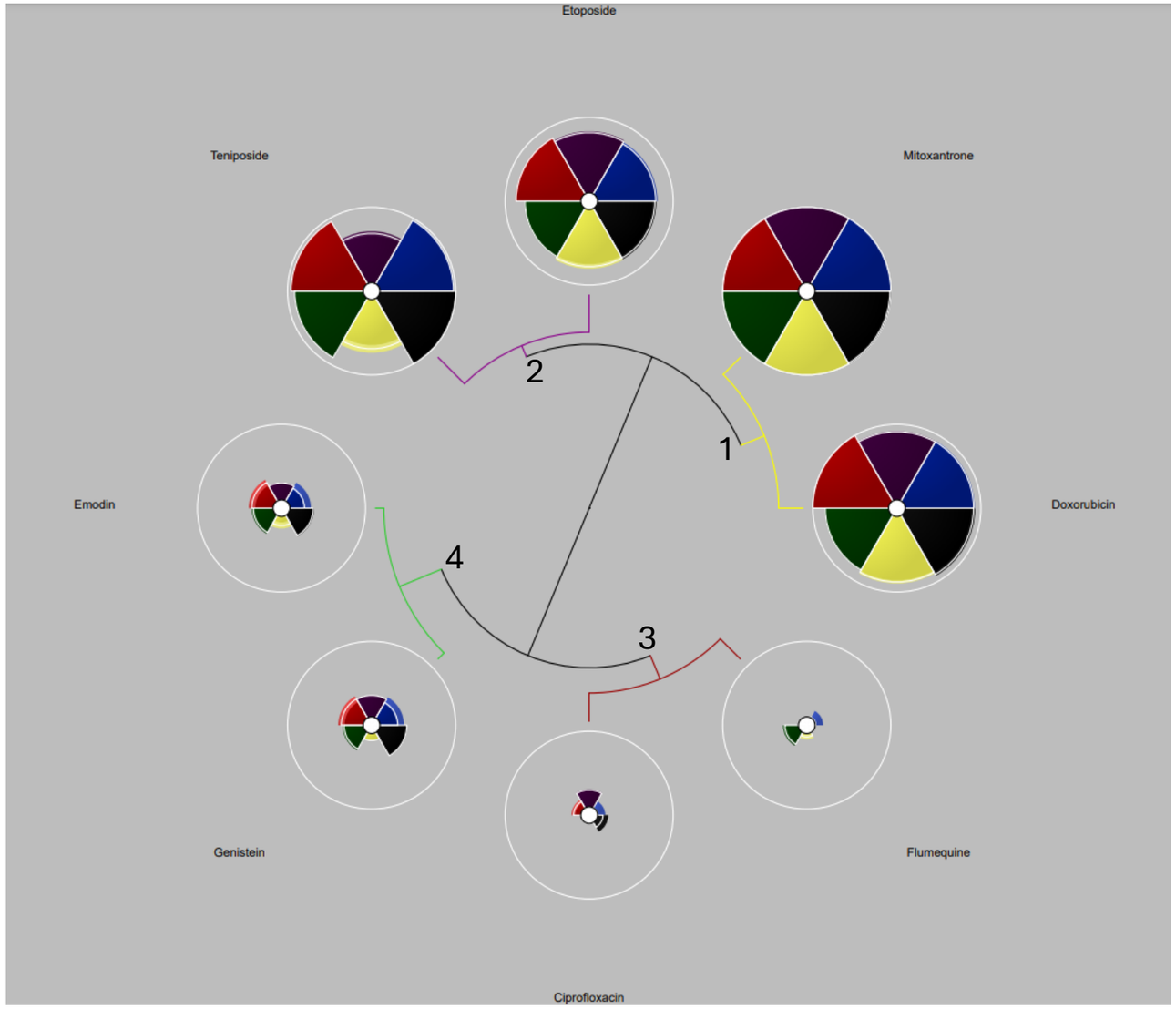


Figure 2.3.1.9 – Component 1 hierarchical AGNES clustering of each ToxPi profile generated through the complete clustering algorithm to show grouping of each topoisomerase II inhibitors based on similarity across four distinctive groupings based on their QSAR classifications described in Wheeldon et al., (2020) and Figure 2.1.6.1.3. Anthraquinones representing group 1 (yellow branch), podophyllotoxins for group 2 (purple branch), fluoroquinolones for group 3 (red branch), and isoflavones for group 4 (green branch) with emodin being the only compound that incorrectly clustered with isoflavone (genistein) instead of with the anthraquinones (mitoxantrone and doxorubicin).

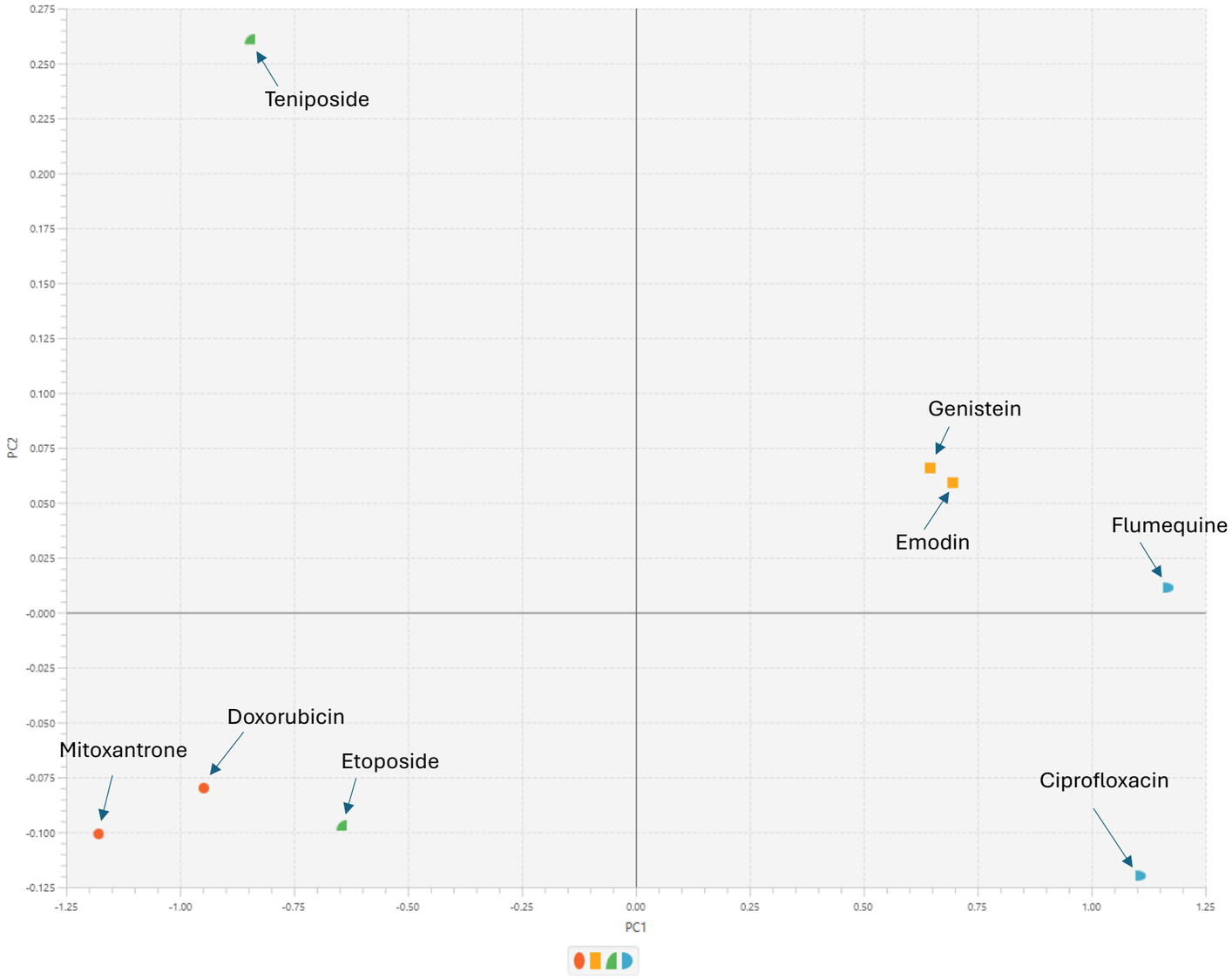


Figure 2.3.1.10 – The K -means clustering generated with a user-defined number of clusters set at 4, using divisive algorithms and PCA plot to graphically plot compounds based on their integrated genotoxic biomarker response patterns. These compounds clusters on a PCA basis where more potent compounds sit on the left (more negative PC1 axis) side, and compounds sat around lower values for the PC2 axis except for Teniposide (higher PC2). Each cluster correlates with the previous hierarchical clusters generated in figure 2.3.1.4, supporting the stability of the observed groupings across analytical methods. However, Teniposide appears to exhibit a larger Euclidian distance to Etoposide, likely due to reduced 4/24hr p53 response projections compared to other compounds and yH2AX and p-H3 reporters.

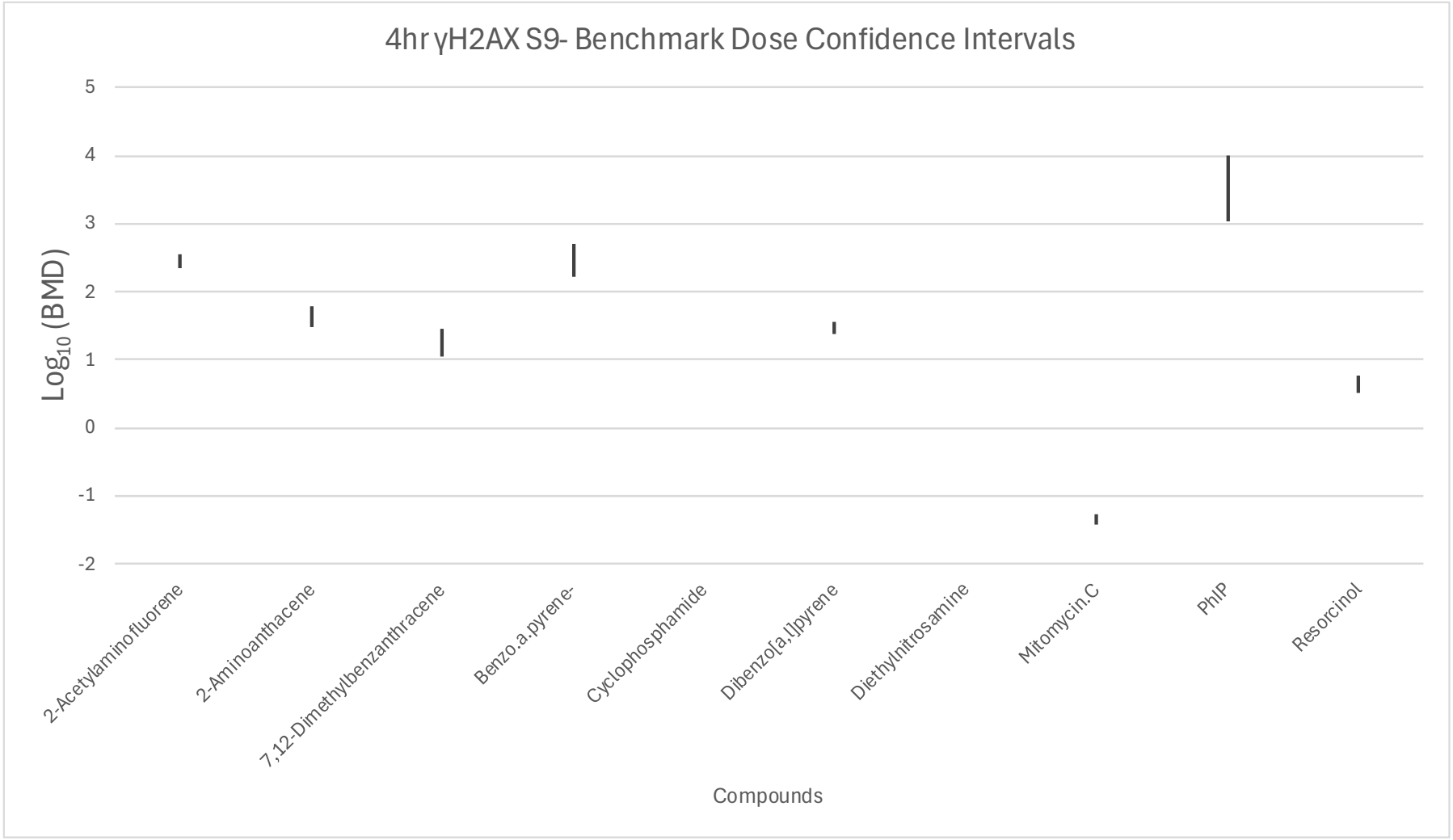


Figure 2.3.2.1 – BMDs (μM) with 90% CIs were generated using a CES of 0.3 in PROAST v65.5 using exponential models 3 and 5 to fit the dose-response curve for TK6 cells exposed to 10 clastogens. Here, γH2AX response at 4 hours in the absence of S9 metabolic activation was analysed.

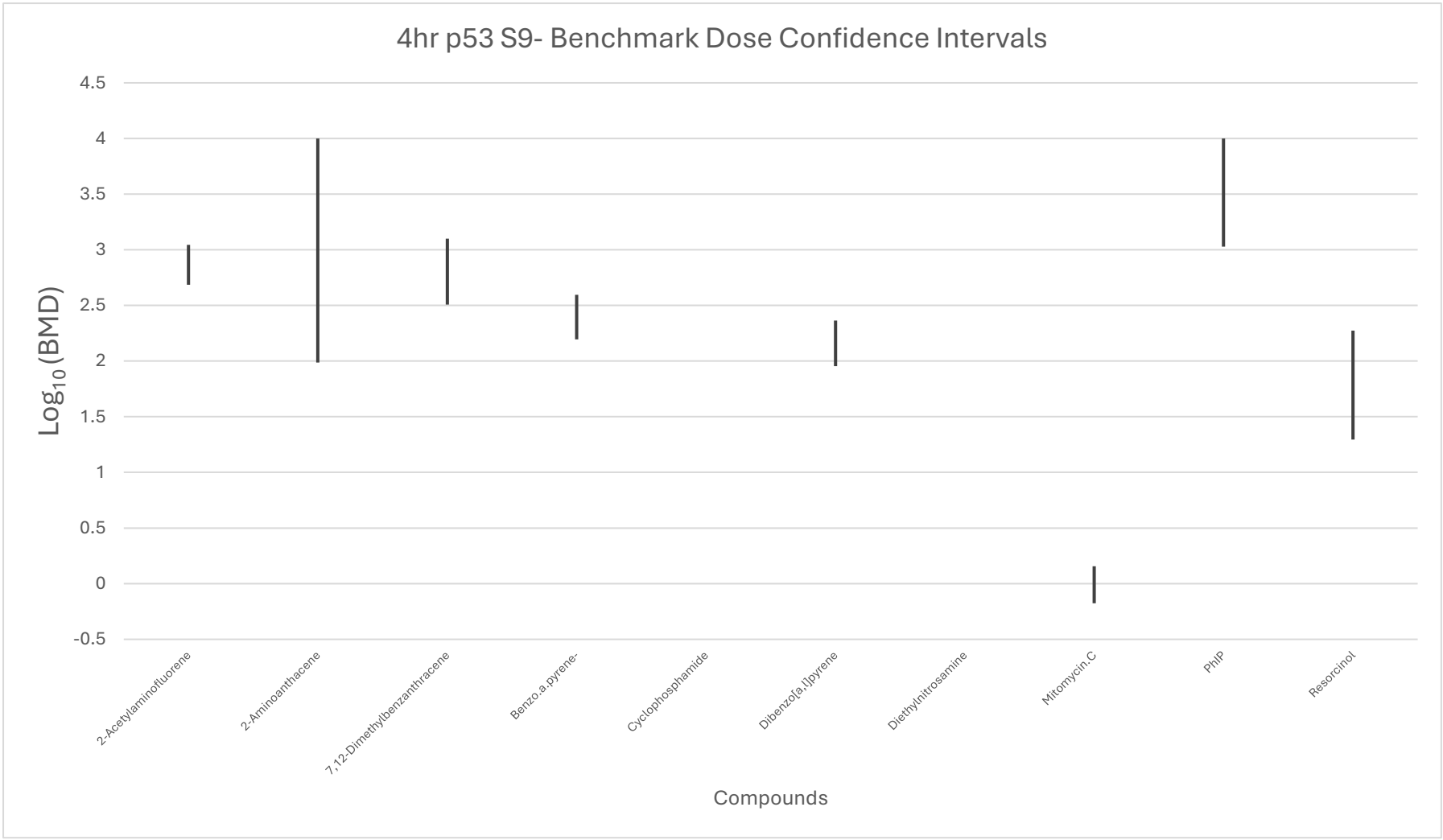


Figure 2.3.2.2 – BMDs (μM) with 90% CIs were generated using PROAST for the p53 response at 4 hours in TK6 cells without S9 activation. Higher log10(BMD) values indicate lower potency.

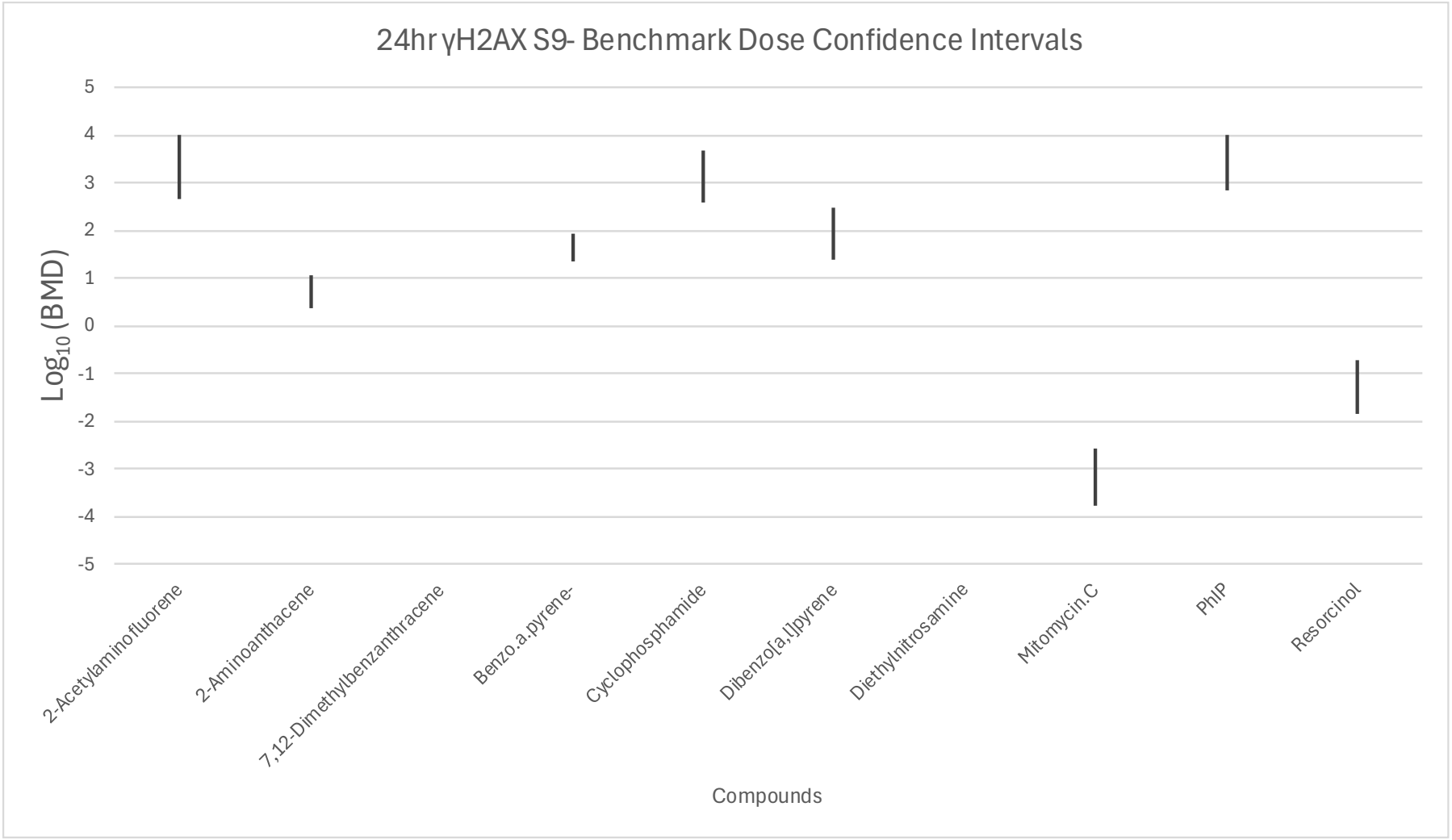


Figure 2.3.2.3 – BMDs (μM) with 90% CIs were generated using PROAST for the γH2AX response at 24 hours in TK6 cells without S9 metabolic activation.

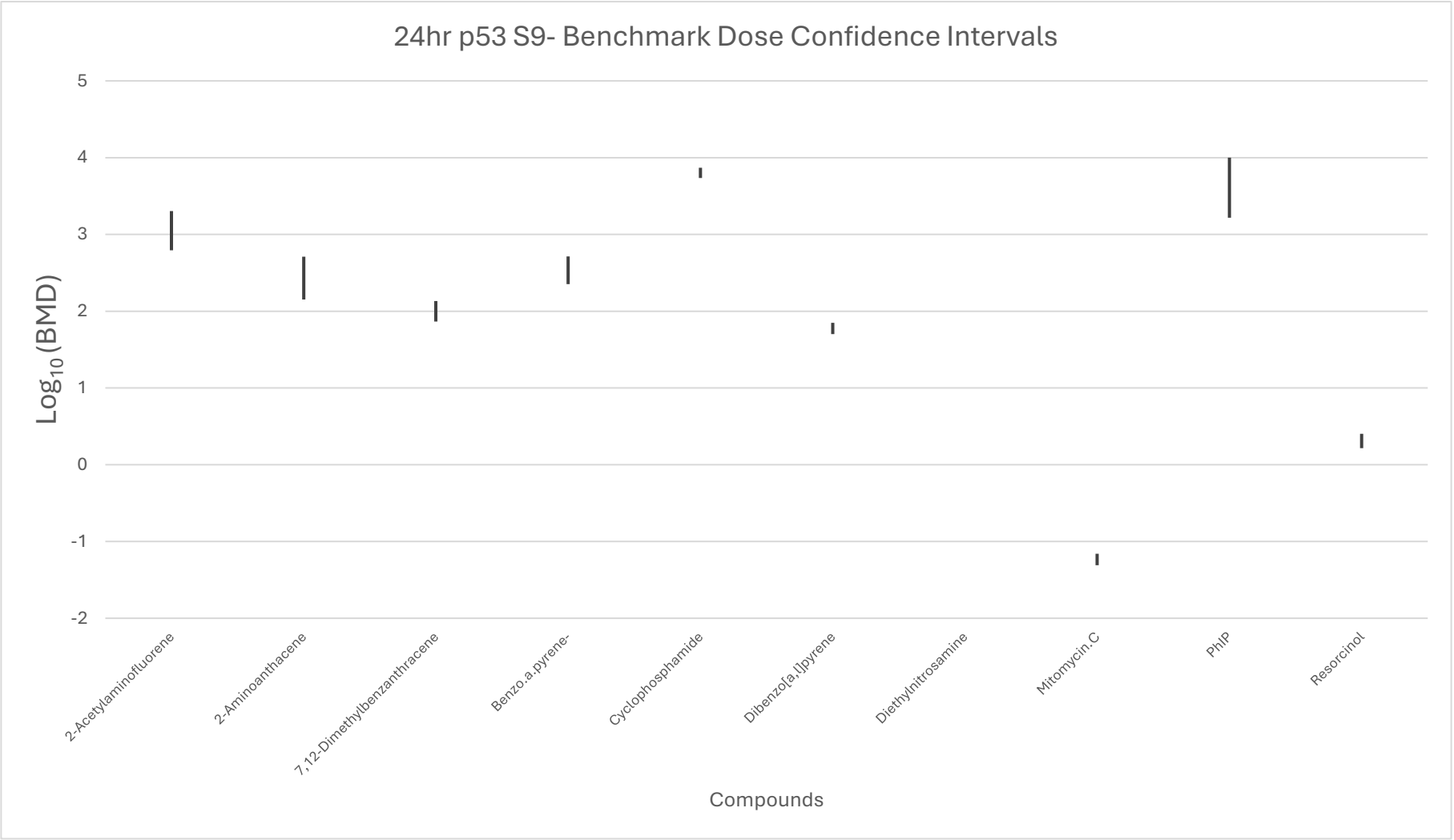


Figure 2.3.2.4 – BMDs (μM) with 90% CIs were generated using PROAST for the p53 response at 24 hours in TK6 cells without S9 activation, where cellular mitochondrial function remained intact whereas the S9 fraction itself contains phase I/II enzymes but lacks intact mitochondria.

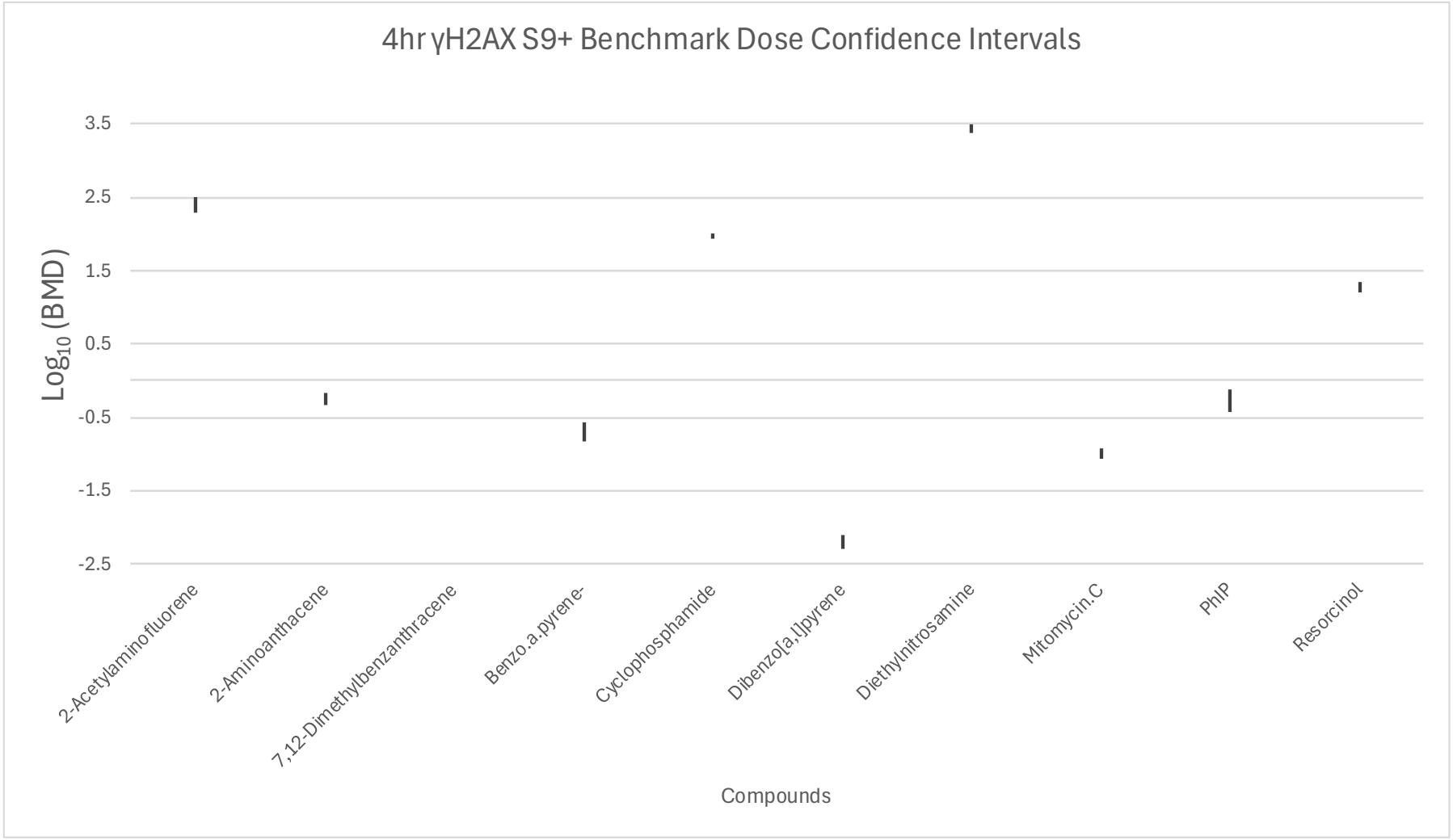


Figure 2.3.2.5 – BMDs (μM) with 90% CIs were generated using PROAST for the γH2AX response at 4 hours in TK6 cells with S9 metabolic activation.

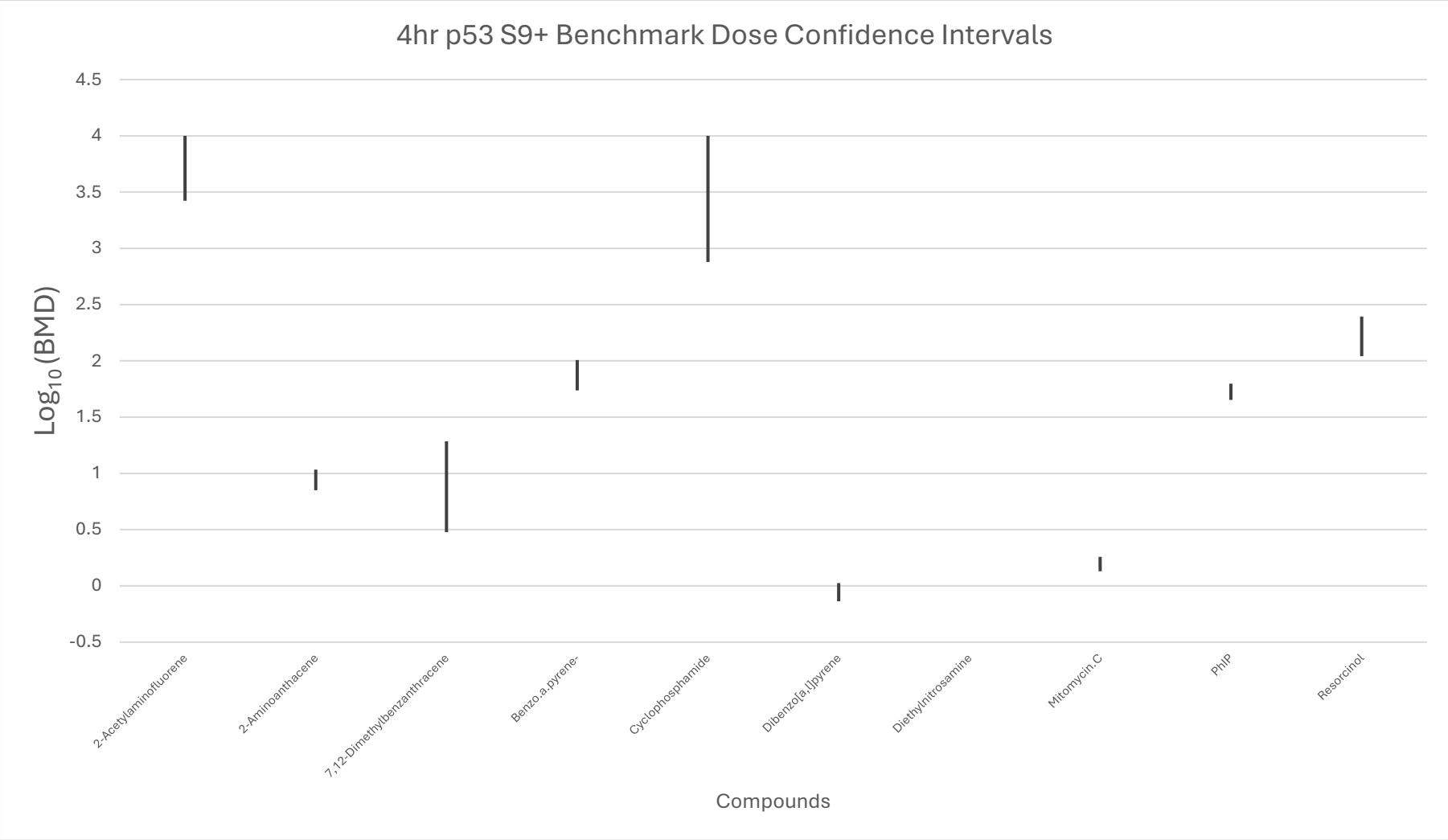


Figure 2.3.2.6 – BMDs (μM) with 90% CIs were generated using PROAST for the p53 response at 4 hours in TK6 cells with S9 metabolic activation. Higher log10(BMD) values indicate lower potency.

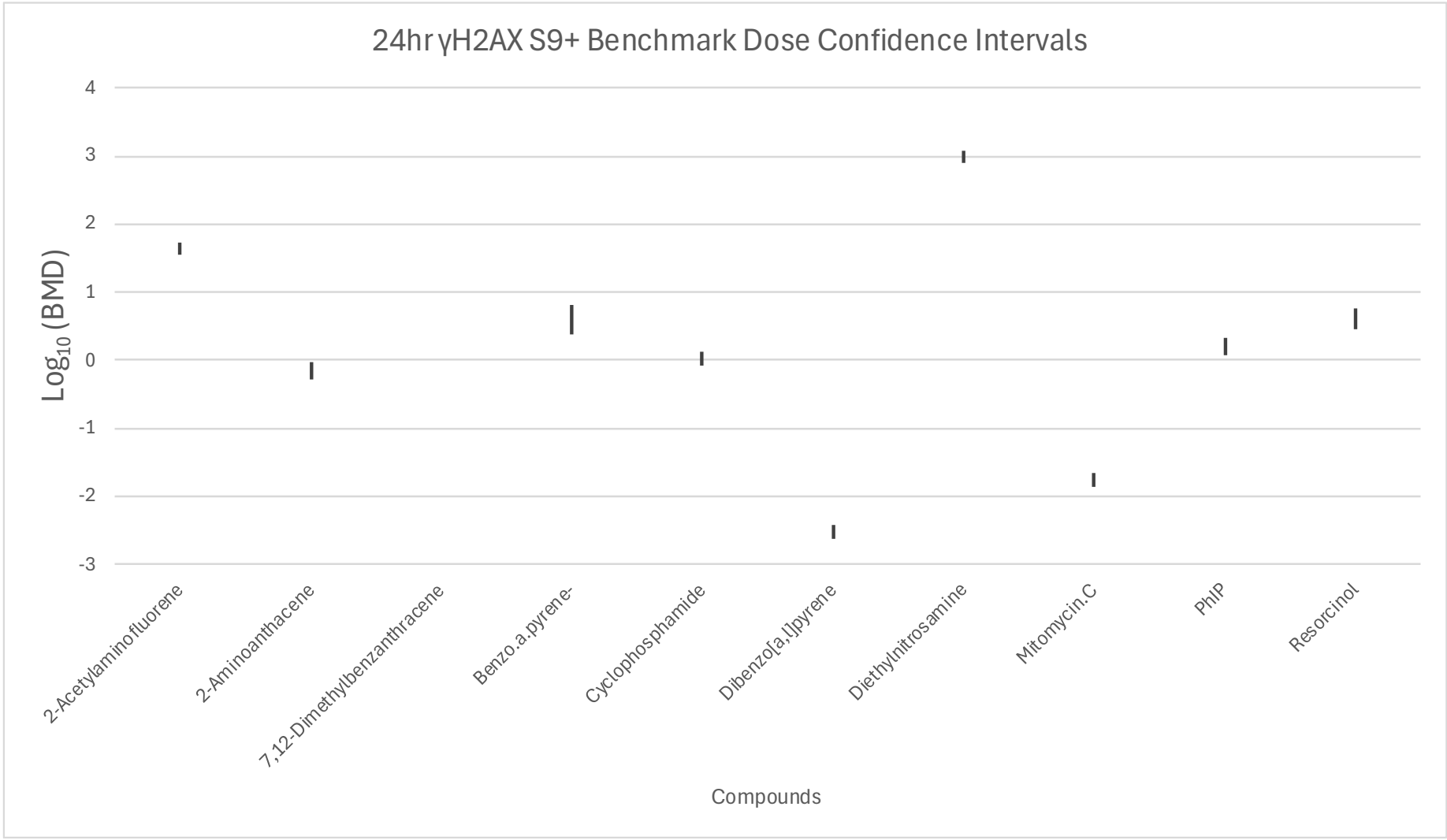


Figure 2.3.2.7 – BMDs (μM) with 90% CIs were generated using PROAST for the γH2AX response at 24 hours in TK6 cells with S9 metabolic activation.

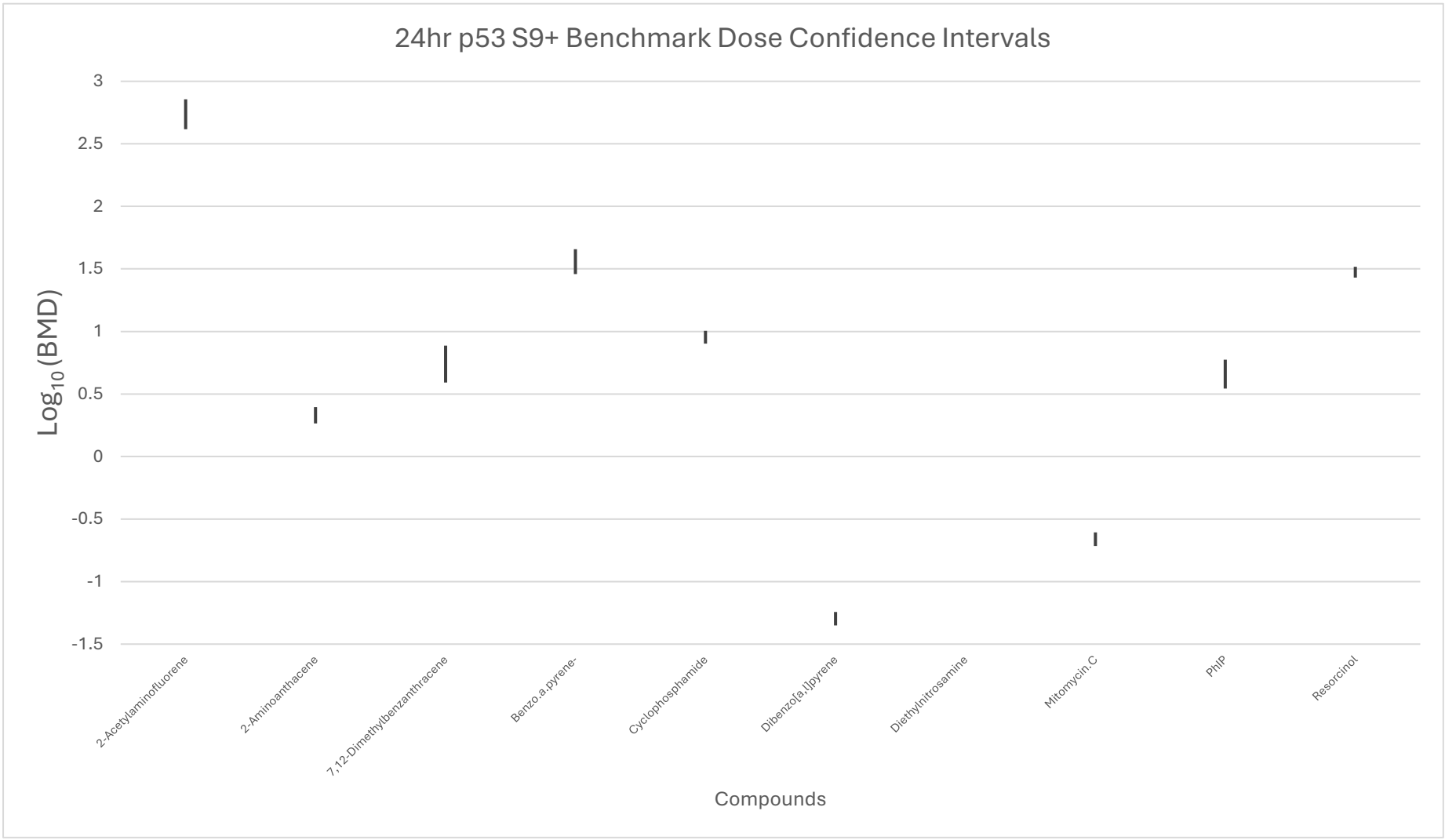


Figure 2.3.2.8 – BMDs (μM) with 90% CIs were generated using PROAST for the p53 response at 24 hours in TK6 cells with S9 metabolic activation.

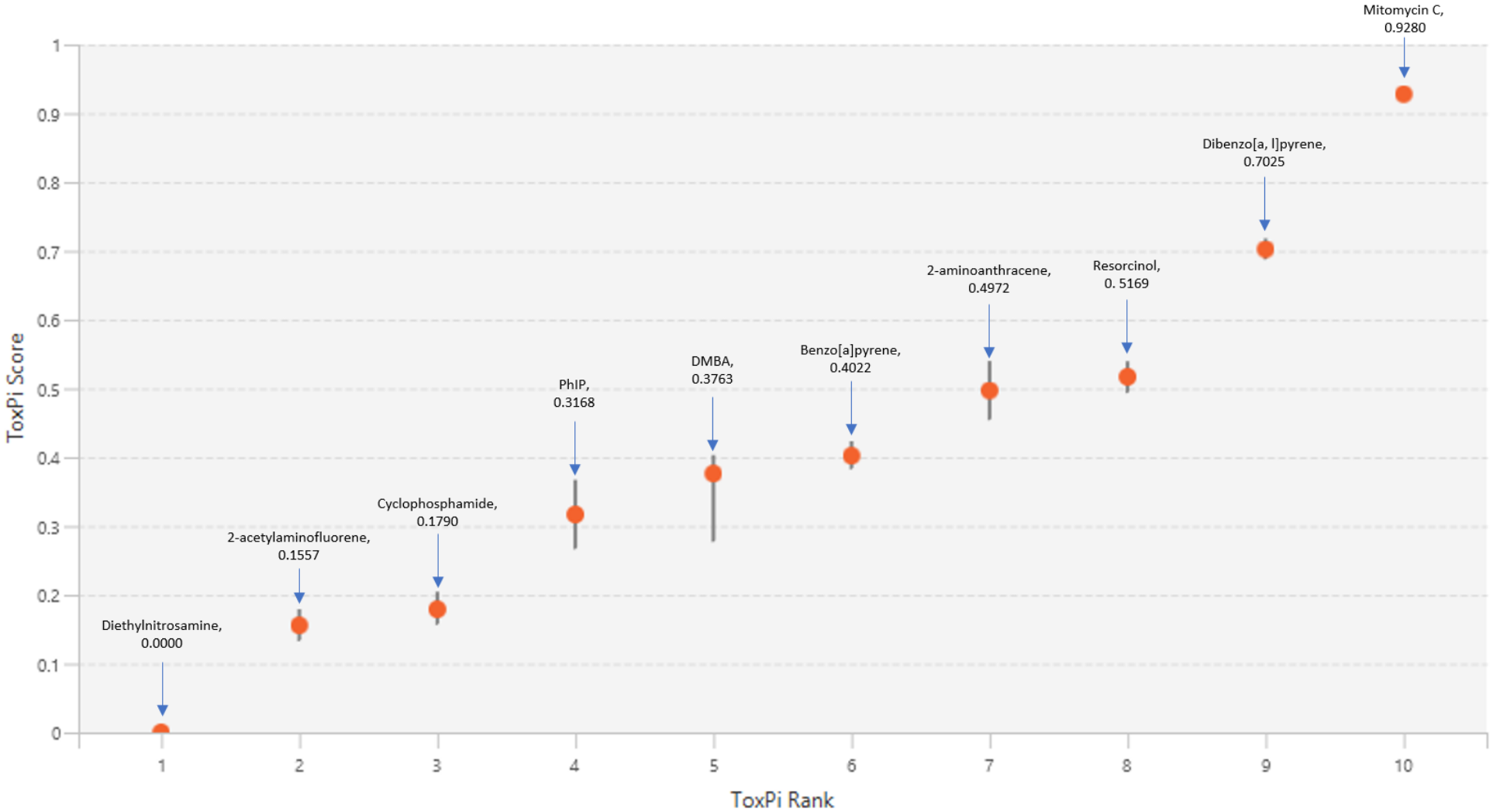


Figure 2.3.2.9 – ToxPi ranking for BMD data on 10 clastogens in the absence of S9 overall (1 = lowest potency and 10 = highest potency). ToxPi scores ranging between 0-1 where the higher score translates to higher genotoxic potency, with the calculation being described in section 2.2.7, and the error bars represent the CI of each compound.

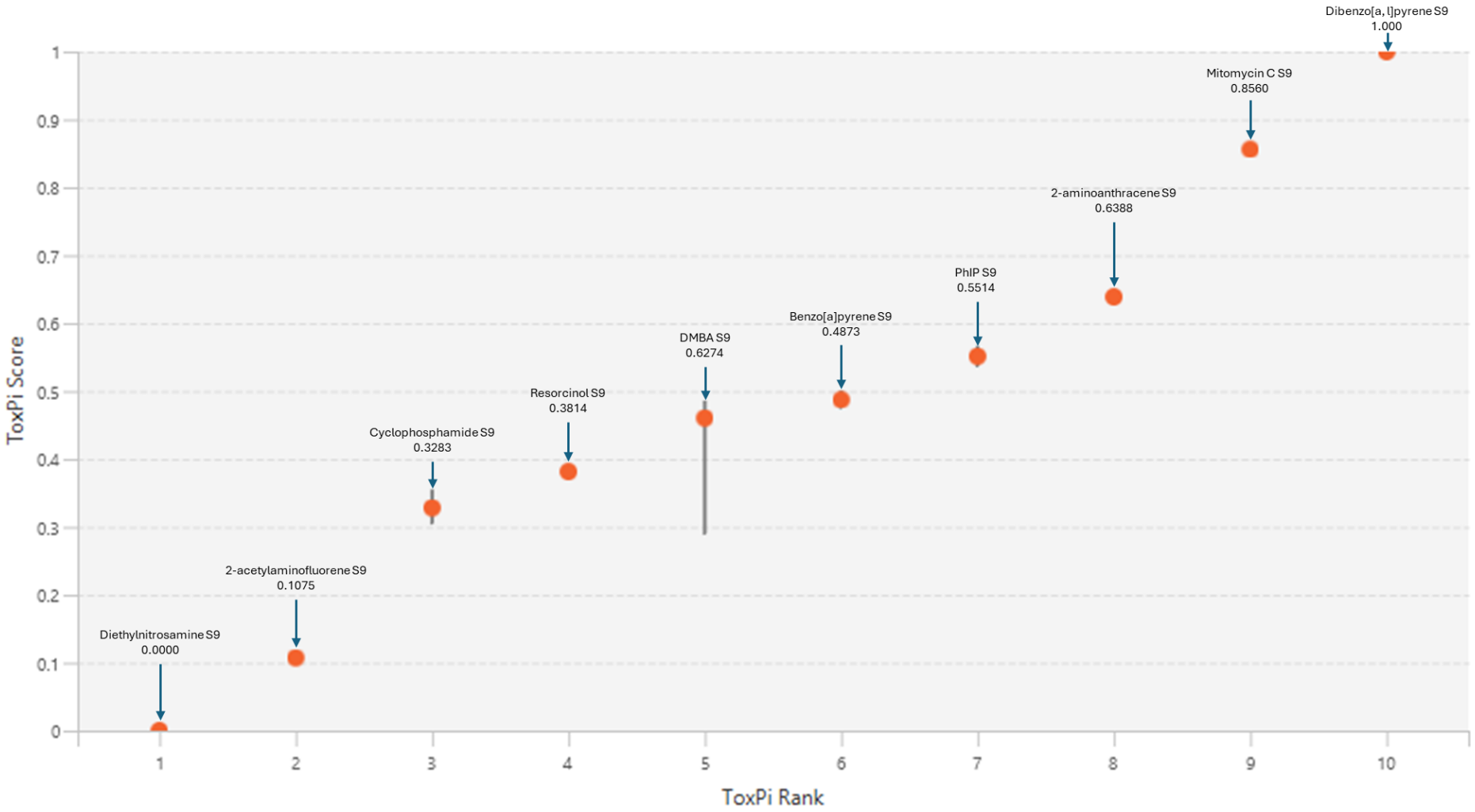


Figure 2.3.2.10 – ToxPi ranking for BMD data on 10 clastogens in the presence of S9 overall (1 = lowest potency and 10 = highest potency). ToxPi scores ranging between 0-1 where the higher score translates to higher genotoxic potency, with the calculation being described in section 2.2.7, and the error bars represent the CI of each compound.

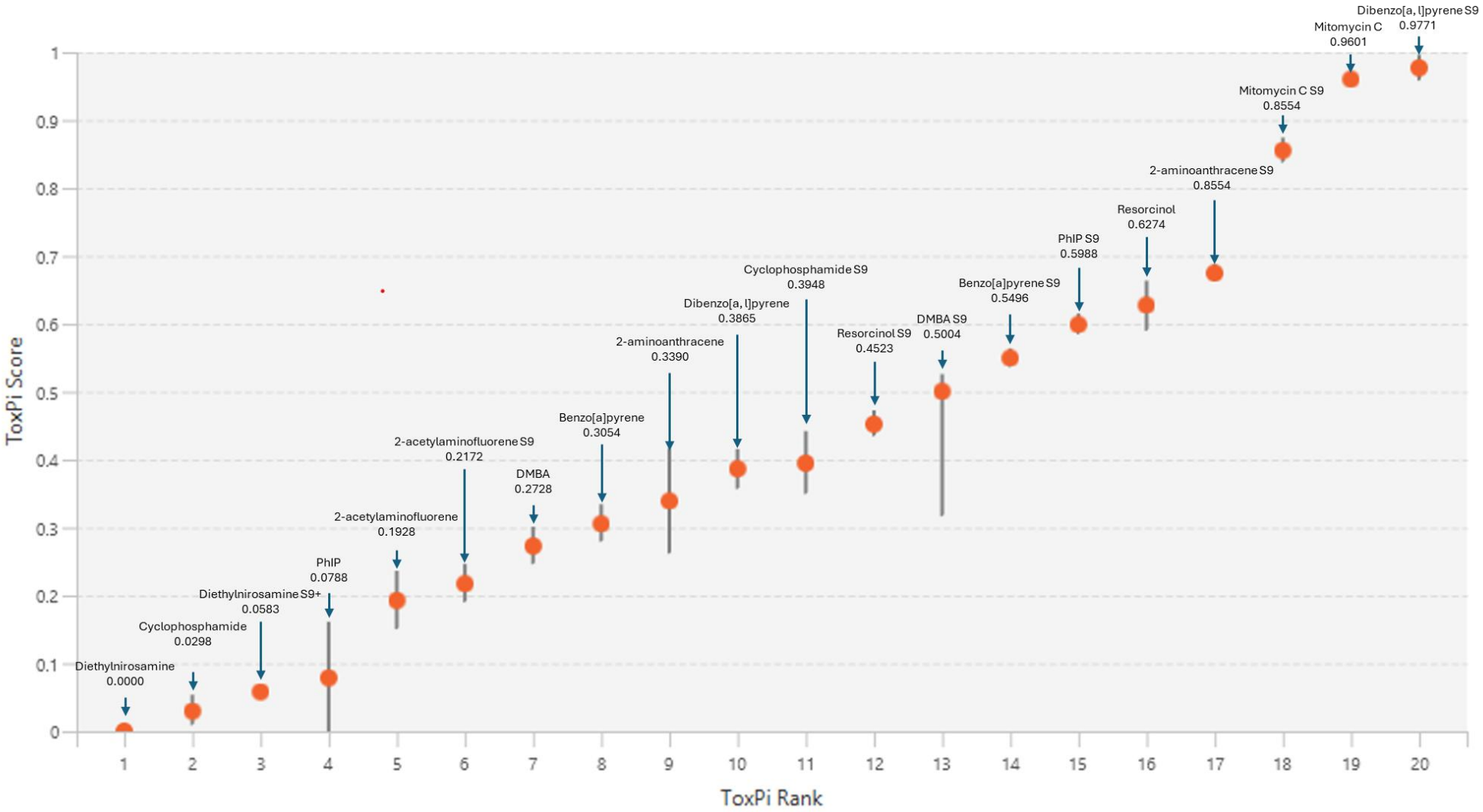


Figure 2.3.2.11 - ToxPi ranks for the 10 clastogens were separated based on the presence/absence of S9, with scores ranging from 0 to 1 produced through the methods given in section 2.2.7, where a higher score indicates greater genotoxic potencies. The error bars represent the confidence interval for each compound.

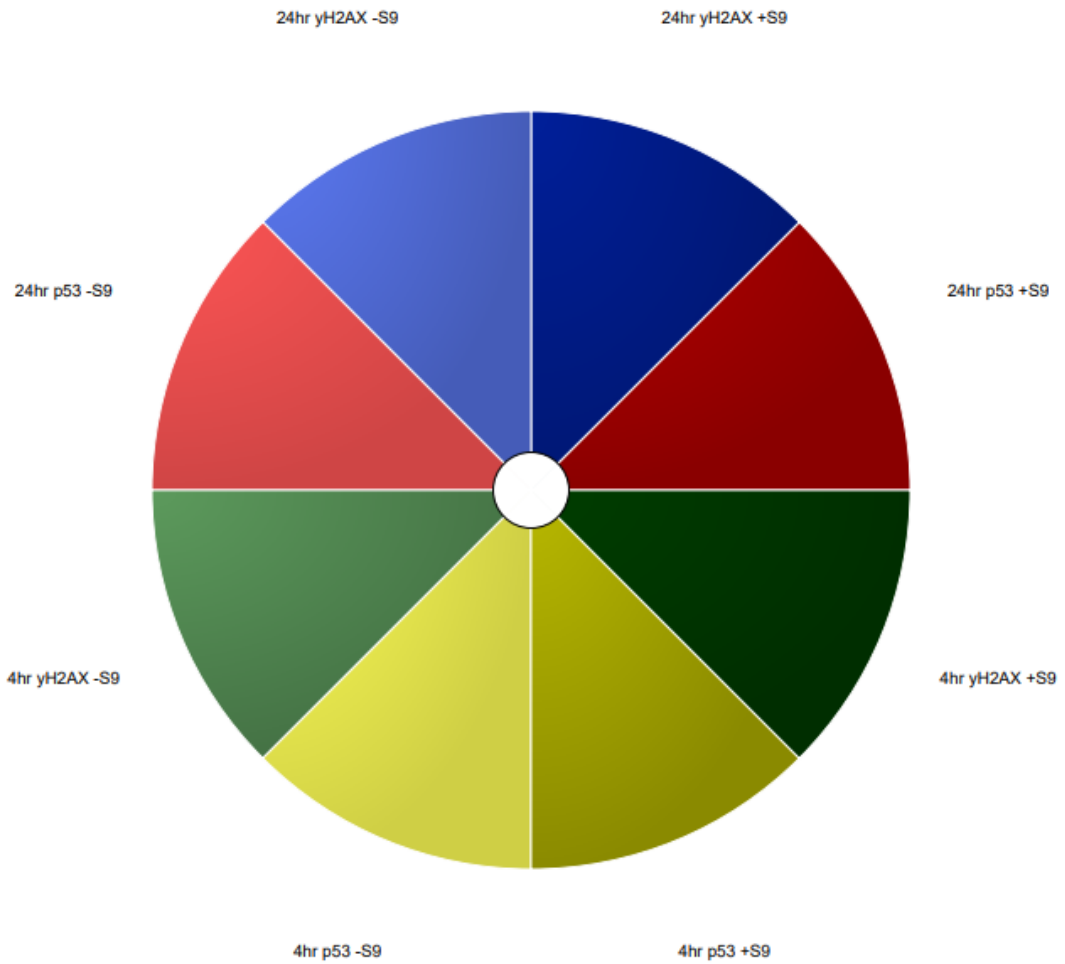
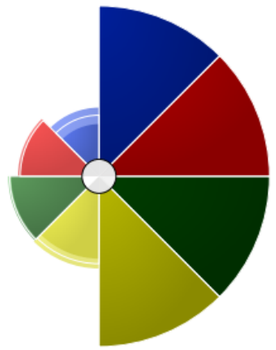


Figure 2.3.2.12 – ToxPi profile key illustrating each slice with their respective endpoints, including colour coordinated endpoints for with γH2AX and p53 endpoints, at 4/24-hour timepoints with S9 present (S9+) and absent (S9-).

Mitomycin.C

Score = 0.9280 [0.9264, 0.9295]
Rank = 10 [10, 10] (out of 10)

24hr p53 +S9 0.8423 [0.8396, 0.8448]
24hr γH2AX +S9 0.8615 [0.8610, 0.8619]
24hr γH2AX -S9 1.0000 [1.0000, 1.0000]
24hr p53 -S9 1.0000 [1.0000, 1.0000]
4hr γH2AX -S9 1.0000 [1.0000, 1.0000]
4hr p53 -S9 1.0000 [1.0000, 1.0000]
4hr p53 +S9 0.9336 [0.9248, 0.9415]
4hr γH2AX +S9 0.7865 [0.7856, 0.7874]



Dibenzo[a,l]pyrene

Score = 0.7025 [0.6899, 0.7156]
Rank = 9 [9, 9] (out of 10)

24hr p53 +S9 1.0000 [1.0000, 1.0000]
24hr γH2AX +S9 1.0000 [1.0000, 1.0000]
24hr γH2AX -S9 0.2873 [0.2289, 0.3366]
24hr p53 -S9 0.4024 [0.4024, 0.4172]
4hr γH2AX -S9 0.4716 [0.4620, 0.4810]
4hr p53 -S9 0.4591 [0.4256, 0.4899]
4hr p53 +S9 1.0000 [1.0000, 1.0000]
4hr γH2AX +S9 1.0000 [1.0000, 1.0000]

Resorcinol

Score = 0.5169 [0.4959, 0.5369]
Rank = 8 [7, 8] (out of 10)

24hr p53 +S9 0.3131 [0.2990, 0.3266]
24hr γH2AX +S9 0.4323 [0.4214, 0.4432]
24hr γH2AX -S9 0.7361 [0.7193, 0.7503]
24hr p53 -S9 0.6937 [0.6937, 0.6978]
4hr γH2AX -S9 0.6274 [0.6135, 0.6409]
4hr p53 -S9 0.5525 [0.4489, 0.6479]
4hr p53 +S9 0.3965 [0.3881, 0.4039]
4hr γH2AX +S9 0.3839 [0.3832, 0.3846]

2-Aminoanthacene

Score = 0.4972 [0.4564, 0.5375]
Rank = 7 [7, 8] (out of 10)

24hr p53 +S9 0.5967 [0.5927, 0.6005]
24hr γH2AX +S9 0.5687 [0.5632, 0.5741]
24hr γH2AX -S9 0.4571 [0.4471, 0.4657]
24hr p53 -S9 0.2720 [0.2500, 0.3135]
4hr γH2AX -S9 0.4424 [0.4212, 0.4629]
4hr p53 -S9 0.2509 [0.0000, 0.4819]
4hr p53 +S9 0.7352 [0.7227, 0.7464]
4hr γH2AX +S9 0.6546 [0.6542, 0.6550]

Benzo.a.pyrene-

Score = 0.4022 [0.3851, 0.4208]
Rank = 6 [6, 6] (out of 10)

24hr p53 +S9 0.2920 [0.2916, 0.2925]
24hr γH2AX +S9 0.4342 [0.4130, 0.4554]
24hr γH2AX -S9 0.3300 [0.3161, 0.3417]
24hr p53 -S9 0.2518 [0.2497, 0.2735]
4hr γH2AX -S9 0.2867 [0.2463, 0.3260]
4hr p53 -S9 0.3999 [0.3650, 0.4320]
4hr p53 +S9 0.4880 [0.4735, 0.5010]
4hr γH2AX +S9 0.7350 [0.7253, 0.7445]

7,12-Dimethylbenzanthracene

Score = 0.3763 [0.2794, 0.4007]
Rank = 5 [5, 5] (out of 10)

24hr p53 +S9 0.4952 [0.4803, 0.5105]
24hr γH2AX +S9 0.0000 [0.0000, 0.0000]
24hr γH2AX -S9 0.0000 [0.0000, 0.0000]
24hr p53 -S9 0.3577 [0.3577, 0.3704]
4hr γH2AX -S9 0.5127 [0.4802, 0.5443]
4hr p53 -S9 0.2984 [0.2340, 0.3576]
4hr p53 +S9 0.7510 [0.6829, 0.8270]
4hr γH2AX +S9 0.5958 [0.0000, 0.5958]

PhIP

Score = 0.3168 [0.2688, 0.3647]
Rank = 4 [4, 4] (out of 10)

24hr p53 +S9 0.5152 [0.5080, 0.5226]
24hr γH2AX +S9 0.5054 [0.5004, 0.5104]
24hr γH2AX -S9 0.0802 [0.0000, 0.1480]
24hr p53 -S9 0.0381 [0.0000, 0.1023]
4hr γH2AX -S9 0.0899 [0.0000, 0.1773]
4hr p53 -S9 0.1210 [0.0000, 0.2324]
4hr p53 +S9 0.5271 [0.4969, 0.5542]
4hr γH2AX +S9 0.6578 [0.6454, 0.6701]

Cyclophosphamide

Score = 0.1790 [0.1589, 0.2020]
Rank = 3 [3, 3] (out of 10)

24hr p53 +S9 0.4420 [0.4320, 0.4517]
24hr γH2AX +S9 0.5385 [0.5378, 0.5393]
24hr γH2AX -S9 0.1190 [0.0473, 0.1797]
24hr p53 -S9 0.0000 [0.0000, 0.0256]
4hr γH2AX -S9 0.0000 [0.0000, 0.0000]
4hr p53 -S9 0.0000 [0.0000, 0.0000]
4hr p53 +S9 0.0724 [0.0000, 0.1532]
4hr γH2AX +S9 0.2603 [0.2543, 0.2663]

2-Acetylaminofluorene

Score = 0.1557 [0.1352, 0.1768]
Rank = 2 [2, 2] (out of 10)

24hr p53 +S9 0.0000 [0.0000, 0.0000]
24hr γH2AX +S9 0.2441 [0.2439, 0.2444]
24hr γH2AX -S9 0.0933 [0.0000, 0.1722]
24hr p53 -S9 0.1495 [0.1351, 0.1865]
4hr γH2AX -S9 0.2901 [0.2754, 0.3043]
4hr p53 -S9 0.2828 [0.2483, 0.3145]
4hr p53 +S9 0.0000 [0.0000, 0.0000]
4hr γH2AX +S9 0.1858 [0.1791, 0.1924]

Diethylnitrosamine

Score = 0.0000 [0.0000, 0.0000]
Rank = 1 [1, 1] (out of 10)

24hr p53 +S9 0.0000 [0.0000, 0.0000]
24hr γH2AX +S9 0.0000 [0.0000, 0.0000]
24hr γH2AX -S9 0.0000 [0.0000, 0.0000]
24hr p53 -S9 0.0000 [0.0000, 0.0000]
4hr γH2AX -S9 0.0000 [0.0000, 0.0000]
4hr p53 -S9 0.0000 [0.0000, 0.0000]
4hr p53 +S9 0.0000 [0.0000, 0.0000]
4hr γH2AX +S9 0.0000 [0.0000, 0.0000]

Figure 2.3.2.13 – ToxPi profiles (with bootstrap CI) generated from BMD data for each chemical in component 2. The visualization displays biomarkers and their ToxPi scores for each endpoint. Each slice represents a biomarker at a specific time point, with and without S9. The shading in the centre of each circle indicates uncertainty, with darker shades signifying greater uncertainty. The radius of each slice, extending from the centre of the circle to either the edge or a white line within the slice, represents the potency of each biomarker, while the width reflects user-defined weightings. The brighter regions indicate the bootstrap confidence interval (CI), with the outer edge showing the upper CI and the inner edge showing the lower CI for each slice.

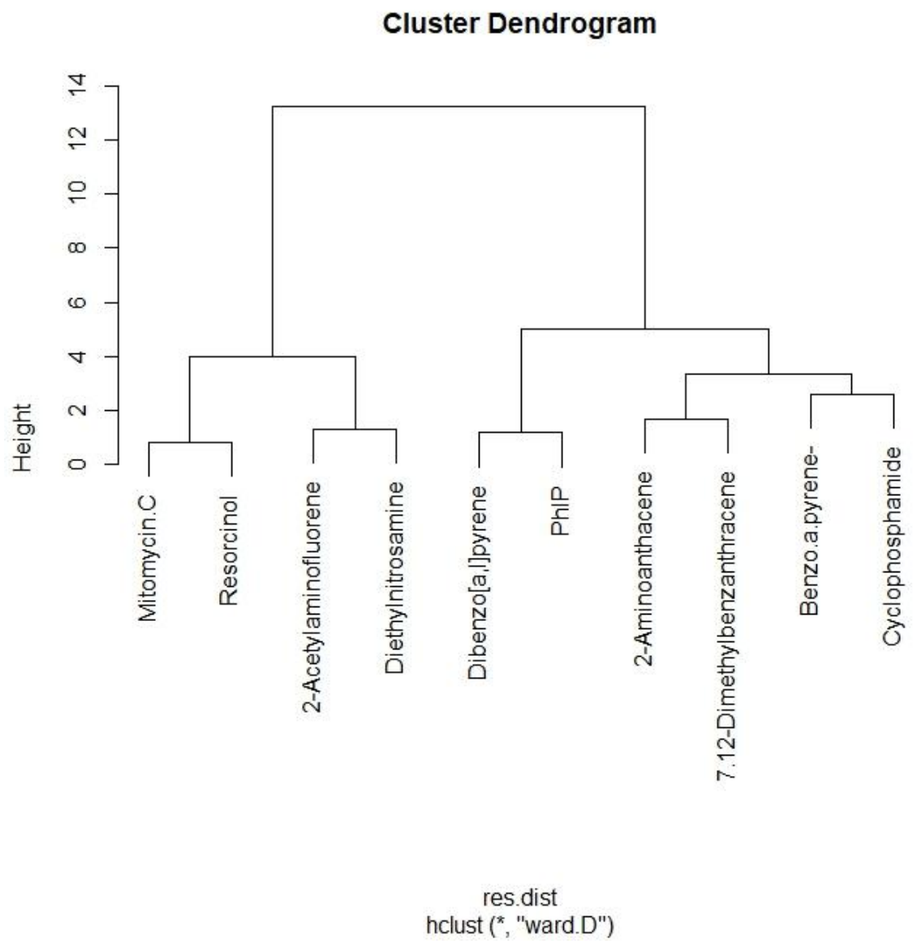


Figure 2.3.2.14 – Euclidean hierarchical clustering was performed in R using the *hclust* function with the ward.D algorithm to reveal distinct groupings of each compound based on their S9 potency ratio CI range values, as described in (Wheeldon et al., 2021; Tian et al., 2020). The clustering was based on the AGNES approach in the original study but differences between the S9- and S9+ conditions were performed using a subtractive approach for each endpoint tested: 4-hour γH2AX, 4-hour p53, 24-hour γH2AX, and 24-hour p53. The branch height on the y-axis represents the degree of compound relatedness and is not attributed to any direct measurement unit, other than to describe the degree to similarity between compounds, where smaller heights indicating increased similarity among compounds. In this case, branch height reflects compound similarity based on metabolic-dependent potency differences. The hclust analysis was included to enable direct comparison between clustering derived from raw BMD values and ToxPi clustering based on transformed BMD inputs, thereby assessing the consistency of compound grouping across alternative analytical frameworks.

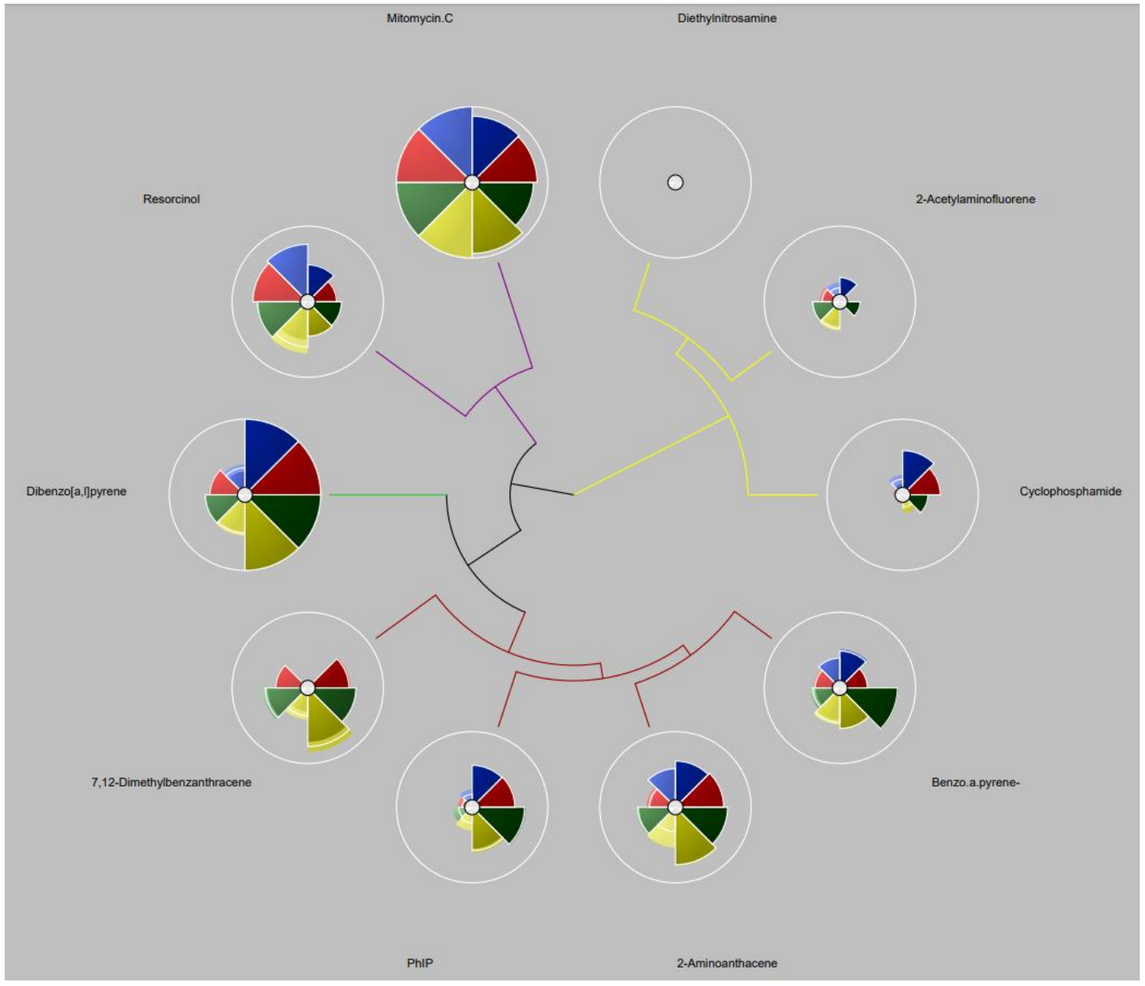


Figure 2.3.2.15 – Complete hierarchical clustering of ToxPi profiles was performed using complete AGNES clustering to identify chemical groupings, each color-coded based on their genotoxic potency in the presence of S9. Green represents compounds with a significant increase in genotoxic potency when S9 is present, red indicates compounds with a moderate increase, yellow denotes no change in potency with the introduction of S9, and purple signifies a decrease in genotoxic capacity due to S9. These groupings correspond to the S9 potency ratio confidence intervals reported by (Wheeldon et al., 2021).

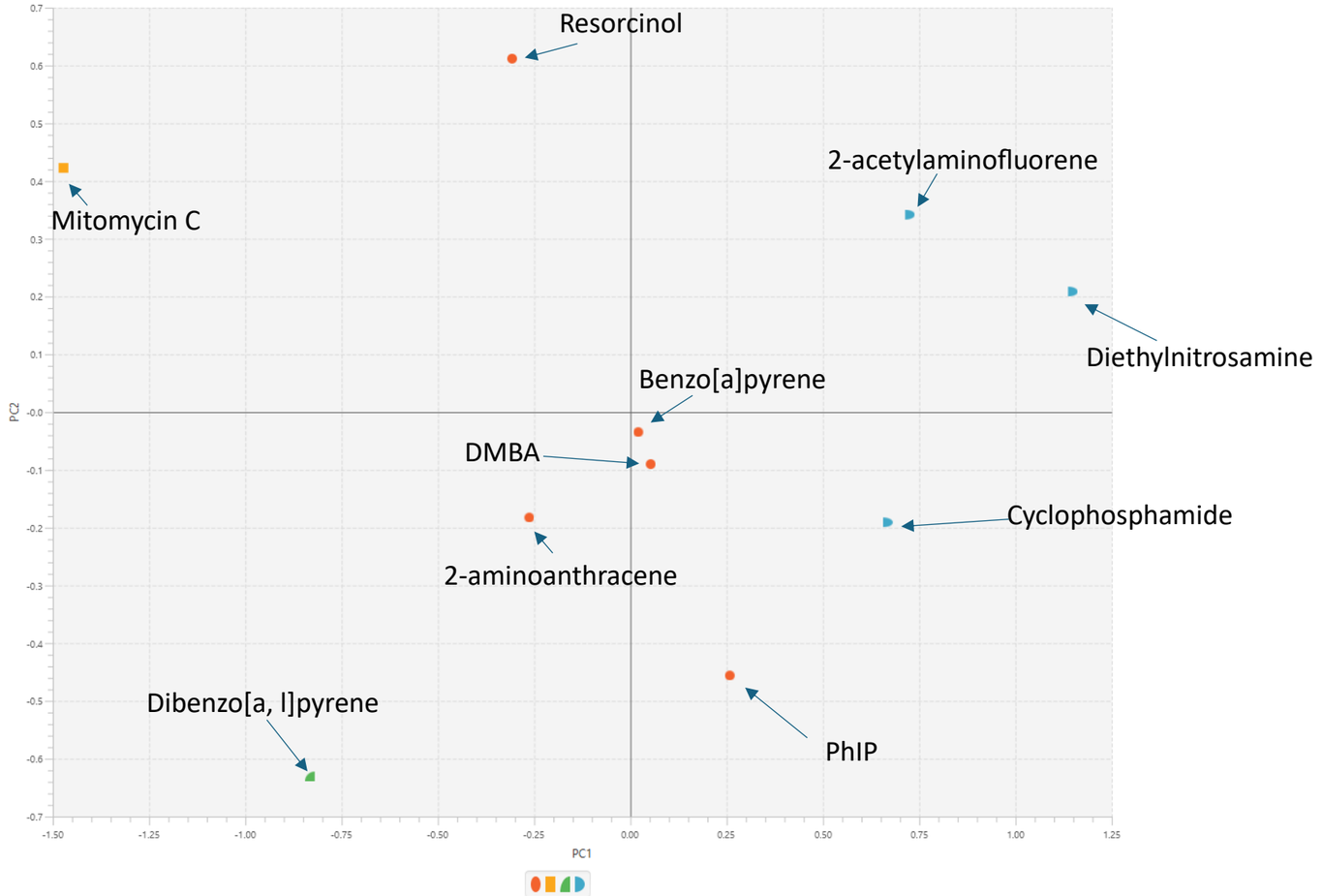


Figure 2.3.2.16 – ToxPi K-means divisive/PCA clustering where each compound is grouped based on variation across the ToxPi scores/genotoxicity and biomarker sensitivity. Each point represents a compound, and each colour/shape represents a cluster they have been placed into. The pattern located is that PC1 compounds with higher ToxPi scores are towards the left, and for PC2, compounds with increased potency with S9+ are at the bottom. Furthermore, the compounds cluster similarly in K-means colours/shapes according to their metabolic properties as described in the hierarchical cluster chart, with exception to mitomycin C and resorcinol: both compounds that have reduced potency with S9 presence. Cluster membership reflects similarity across the full multidimensional ToxPi space; therefore, compounds assigned to the same cluster may appear separated in the two-dimensional PCA projection due to variance captured in higher principal components and the forcing of the clusters based on the user selected K cluster number. While hierarchical provided clearer cluster groupings, the K-means clusters support the presence of broadly related compounds based on Euclidian proximity when evaluated across a 2D space.

The initial BMD graphs which were shown in the sections discussing the 8 Topoisomerase II inhibitors and 10 clastogens are not included in this section.

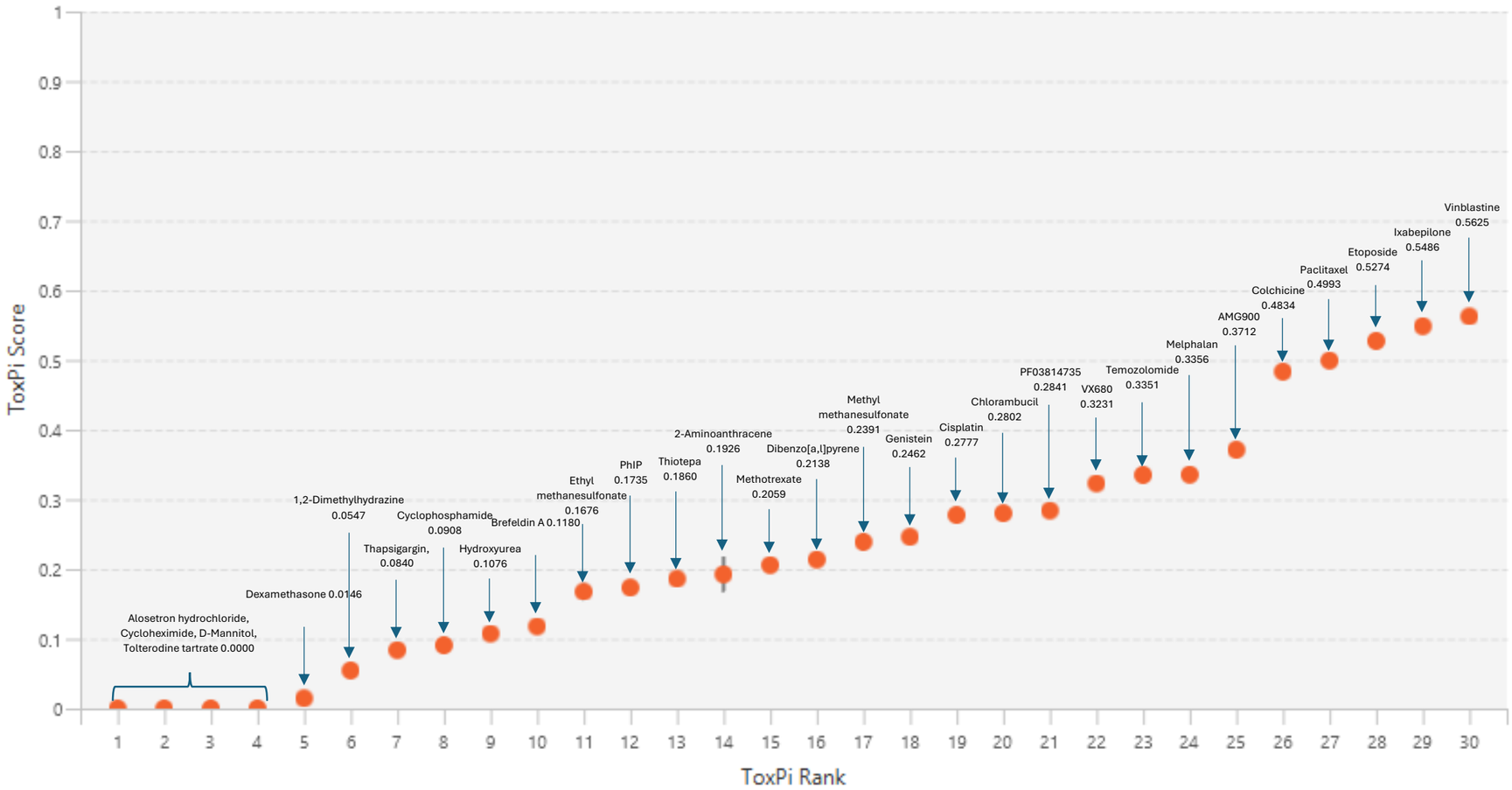


Figure 2.3.3.1 – ToxPi ranking for BMD data on 30 aneugens/clastogens (1 = lowest potency and 30 = highest potency). ToxPi scores, which were calculated using the normalisation formulae described in section 2.2.7 ranging between 0-1 where the higher score translates to higher genotoxic potency and the error bars represent the CI of each compound

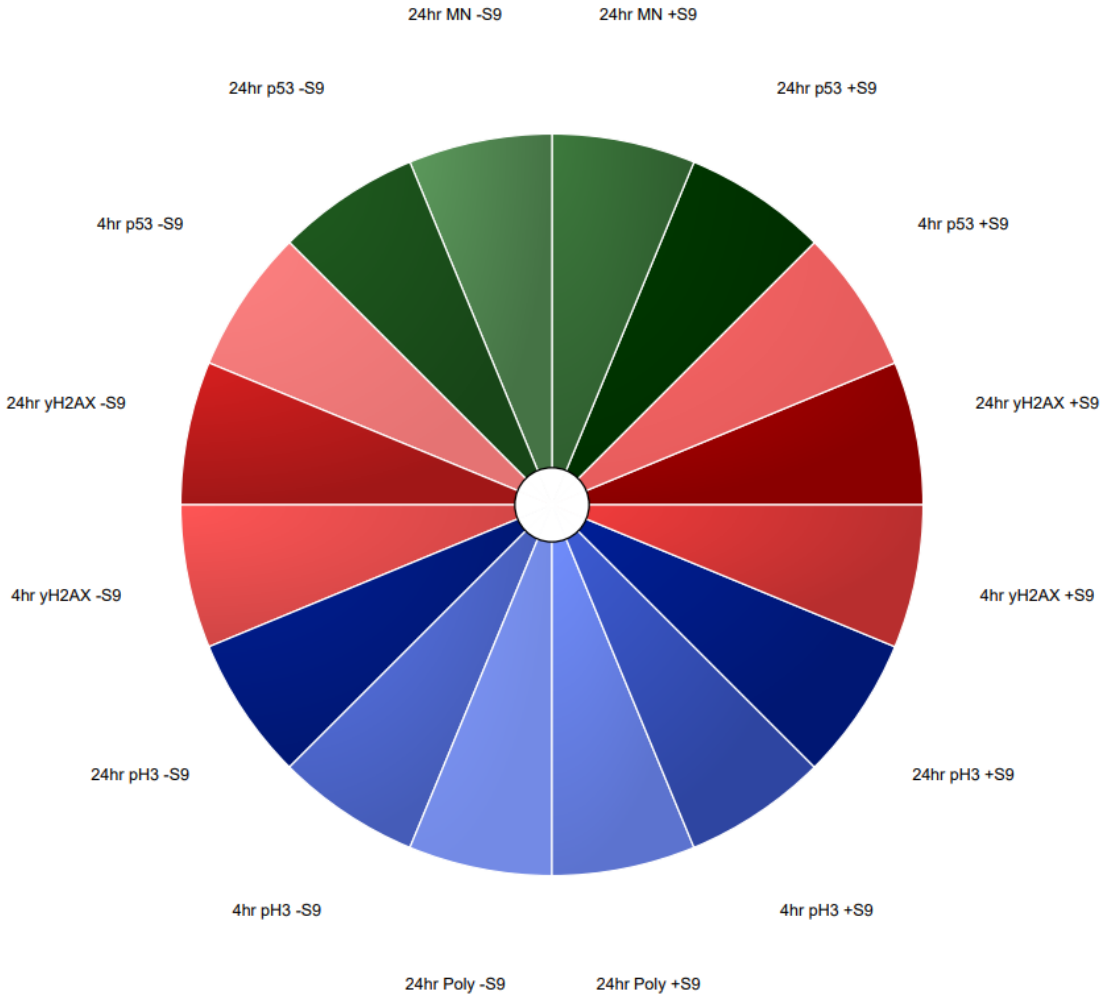
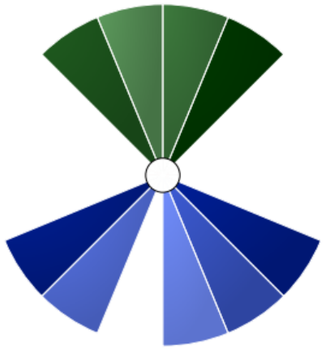


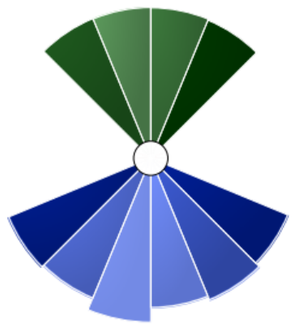
Figure 2.3.3.2 – ToxPi profile key illustrating each slice, and their colour coordinated across the MultiFlow/MicroFlow endpoints for both S9± conditions for each biomarker at 4-hour and 24-hour timepoints. Slice colour ordering were selected to reflect mechanistic groupings of genotoxicity biomarkers: green slices denoting downstream chromosomal damage (24hr MN and 24hr p53), red slices representing DNA lesion/DNA damage response signalling (4-hr p53 and 4-hr/24hr yH2AX) and blue slices indicating mitotic dysregulation associated with aneugenicity (4-hr/24hr p-H3 and 4-hr/24hr poly).



Vinblastine

Score = 0.5625 [0.5625, 0.5625]
Rank = 30 [30, 30] (out of 30)

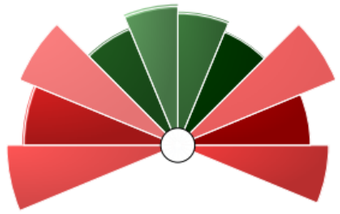
24hr γH2AX +S9 0.0000 [0.0000, 0.0000]
4hr p53 +S9 0.0000 [0.0000, 0.0000]
24hr p53 +S9 1.0000 [1.0000, 1.0000]
24hr MN +S9 1.0000 [1.0000, 1.0000]
24hr MN -S9 1.0000 [1.0000, 1.0000]
24hr p53 -S9 1.0000 [1.0000, 1.0000]
4hr p53 -S9 0.0000 [0.0000, 0.0000]
24hr γH2AX -S9 0.0000 [0.0000, 0.0000]
4hr γH2AX -S9 0.0000 [0.0000, 0.0000]
24hr pH3 -S9 1.0000 [1.0000, 1.0000]
4hr pH3 -S9 1.0000 [1.0000, 1.0000]
24hr Poly -S9 0.0000 [0.0000, 0.0000]
24hr Poly +S9 1.0000 [1.0000, 1.0000]
4hr pH3 +S9 1.0000 [1.0000, 1.0000]
24hr pH3 +S9 1.0000 [1.0000, 1.0000]
4hr γH2AX +S9 0.0000 [0.0000, 0.0000]



Ixabepilone

Score = 0.5486 [0.5455, 0.5517]
Rank = 29 [29, 29] (out of 30)

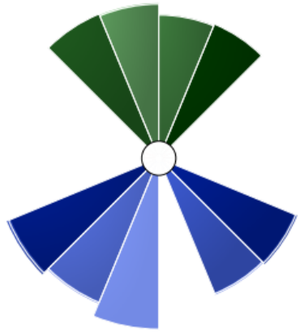
24hr γH2AX +S9 0.0000 [0.0000, 0.0000]
4hr p53 +S9 0.0000 [0.0000, 0.0000]
24hr p53 +S9 0.8597 [0.8586, 0.8608]
24hr MN +S9 0.8666 [0.8660, 0.8671]
24hr MN -S9 0.8687 [0.8645, 0.8730]
24hr p53 -S9 0.8712 [0.8710, 0.8714]
4hr p53 -S9 0.0000 [0.0000, 0.0000]
24hr γH2AX -S9 0.0000 [0.0000, 0.0000]
4hr γH2AX -S9 0.0000 [0.0000, 0.0000]
24hr pH3 -S9 0.8879 [0.8787, 0.8976]
4hr pH3 -S9 0.8729 [0.8651, 0.8808]
24hr Poly -S9 0.9545 [0.9525, 0.9564]
24hr Poly +S9 0.8592 [0.8517, 0.8666]
4hr pH3 +S9 0.8851 [0.8767, 0.8935]
24hr pH3 +S9 0.8520 [0.8437, 0.8606]
4hr γH2AX +S9 0.0000 [0.0000, 0.0000]



Etoposide

Score = 0.5274 [0.5241, 0.5306]
Rank = 28 [28, 28] (out of 30)

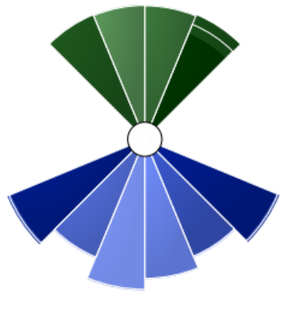
24hr γH2AX +S9 0.7504 [0.7503, 0.7505]
4hr p53 +S9 1.0000 [1.0000, 1.0000]
24hr p53 +S9 0.6841 [0.6784, 0.6897]
24hr MN +S9 0.7470 [0.7362, 0.7577]
24hr MN -S9 0.7922 [0.7784, 0.8056]
24hr p53 -S9 0.7053 [0.6962, 0.7145]
4hr p53 -S9 1.0000 [1.0000, 1.0000]
24hr γH2AX -S9 0.8899 [0.8799, 0.8999]
4hr γH2AX -S9 1.0000 [1.0000, 1.0000]
24hr pH3 -S9 0.0000 [0.0000, 0.0000]
4hr pH3 -S9 0.0000 [0.0000, 0.0000]
24hr Poly -S9 0.0000 [0.0000, 0.0000]
24hr Poly +S9 0.0000 [0.0000, 0.0000]
4hr pH3 +S9 0.0000 [0.0000, 0.0000]
24hr pH3 +S9 0.0000 [0.0000, 0.0000]
4hr γH2AX +S9 0.8692 [0.8670, 0.8714]



Paclitaxel

Score = 0.4993 [0.4963, 0.5023]
Rank = 27 [27, 27] (out of 30)

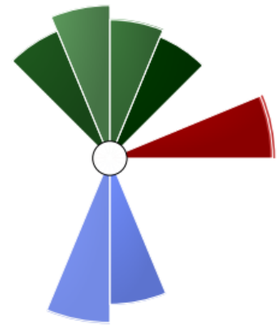
24hr γH2AX +S9 0.0000 [0.0000, 0.0000]
4hr p53 +S9 0.0000 [0.0000, 0.0000]
24hr p53 +S9 0.8322 [0.8310, 0.8334]
24hr MN +S9 0.8197 [0.8170, 0.8224]
24hr MN -S9 0.8913 [0.8851, 0.8975]
24hr p53 -S9 0.9033 [0.9030, 0.9036]
4hr p53 -S9 0.0000 [0.0000, 0.0000]
24hr γH2AX -S9 0.0000 [0.0000, 0.0000]
4hr γH2AX -S9 0.0000 [0.0000, 0.0000]
24hr pH3 -S9 0.9444 [0.9318, 0.9576]
4hr pH3 -S9 0.9041 [0.8968, 0.9113]
24hr Poly -S9 1.0000 [1.0000, 1.0000]
24hr Poly +S9 0.0000 [0.0000, 0.0000]
4hr pH3 +S9 0.8418 [0.8347, 0.8488]
24hr pH3 +S9 0.8516 [0.8416, 0.8620]
4hr γH2AX +S9 0.0000 [0.0000, 0.0000]



Colchicine

Score = 0.4834 [0.4761, 0.4909]
Rank = 26 [26, 26] (out of 30)

24hr γH2AX +S9 0.0000 [0.0000, 0.0000]
4hr p53 +S9 0.0000 [0.0000, 0.0000]
24hr p53 +S9 0.6969 [0.6325, 0.7619]
24hr MN +S9 0.7562 [0.7539, 0.7586]
24hr MN -S9 0.7591 [0.7552, 0.7631]
24hr p53 -S9 0.7648 [0.7644, 0.7651]
4hr p53 -S9 0.0000 [0.0000, 0.0000]
24hr γH2AX -S9 0.0000 [0.0000, 0.0000]
4hr γH2AX -S9 0.0000 [0.0000, 0.0000]
24hr pH3 -S9 0.8435 [0.8315, 0.8562]
4hr pH3 -S9 0.7190 [0.7132, 0.7248]
24hr Poly -S9 0.8652 [0.8587, 0.8718]
24hr Poly +S9 0.7937 [0.7893, 0.7982]
4hr pH3 +S9 0.7124 [0.7065, 0.7184]
24hr pH3 +S9 0.8235 [0.8117, 0.8359]
4hr γH2AX +S9 0.0000 [0.0000, 0.0000]



AMG900

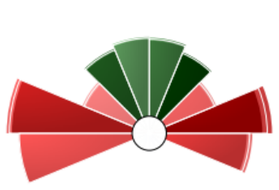
Score = 0.3712 [0.3687, 0.3737]
Rank = 25 [25, 25] (out of 30)

24hr γH2AX +S9 0.9466 [0.9314, 0.9627]
4hr p53 +S9 0.0000 [0.0000, 0.0000]
24hr p53 +S9 0.7419 [0.7413, 0.7426]
24hr MN +S9 0.7895 [0.7829, 0.7959]
24hr MN -S9 0.8824 [0.8794, 0.8854]
24hr p53 -S9 0.7831 [0.7802, 0.7859]
4hr p53 -S9 0.0000 [0.0000, 0.0000]
24hr γH2AX -S9 0.0000 [0.0000, 0.0000]
4hr γH2AX -S9 0.0000 [0.0000, 0.0000]
24hr pH3 -S9 0.0000 [0.0000, 0.0000]
4hr pH3 -S9 0.0000 [0.0000, 0.0000]
24hr Poly -S9 0.9570 [0.9507, 0.9635]
24hr Poly +S9 0.8384 [0.8340, 0.8429]
4hr pH3 +S9 0.0000 [0.0000, 0.0000]
24hr pH3 +S9 0.0000 [0.0000, 0.0000]
4hr γH2AX +S9 0.0000 [0.0000, 0.0000]

Melphalan

Score = 0.3356 [0.3313, 0.3399]
Rank = 24 [23, 24] (out of 30)

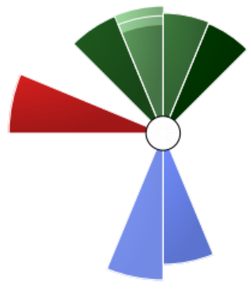
24hr γH2AX +S9 0.7087 [0.7038, 0.7140]
4hr p53 +S9 0.0000 [0.0000, 0.0000]
24hr p53 +S9 0.5231 [0.5178, 0.5283]
24hr MN +S9 0.6946 [0.6747, 0.7142]
24hr MN -S9 0.7040 [0.6866, 0.7209]
24hr p53 -S9 0.5272 [0.5209, 0.5335]
4hr p53 -S9 0.0000 [0.0000, 0.0000]
24hr γH2AX -S9 0.8218 [0.8107, 0.8327]
4hr γH2AX -S9 0.7569 [0.7564, 0.7574]
24hr pH3 -S9 0.0000 [0.0000, 0.0000]
4hr pH3 -S9 0.0000 [0.0000, 0.0000]
24hr Poly -S9 0.0000 [0.0000, 0.0000]
24hr Poly +S9 0.0000 [0.0000, 0.0000]
4hr pH3 +S9 0.0000 [0.0000, 0.0000]
24hr pH3 +S9 0.0000 [0.0000, 0.0000]
4hr γH2AX +S9 0.6328 [0.6291, 0.6368]



Temozolomide

Score = 0.3351 [0.3273, 0.3428]
Rank = 23 [23, 24] (out of 30)

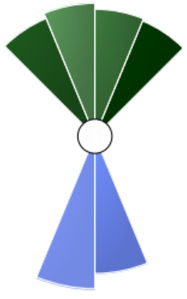
24hr γH2AX +S9 0.6562 [0.6237, 0.6869]
4hr p53 +S9 0.3492 [0.3386, 0.3594]
24hr p53 +S9 0.4515 [0.4439, 0.4591]
24hr MN +S9 0.5009 [0.4913, 0.5103]
24hr MN -S9 0.5076 [0.5000, 0.5151]
24hr p53 -S9 0.4748 [0.4660, 0.4835]
4hr p53 -S9 0.3474 [0.3361, 0.3583]
24hr γH2AX -S9 0.7988 [0.7816, 0.8157]
4hr γH2AX -S9 0.7345 [0.7321, 0.7370]
24hr pH3 -S9 0.0000 [0.0000, 0.0000]
4hr pH3 -S9 0.0000 [0.0000, 0.0000]
24hr Poly -S9 0.0000 [0.0000, 0.0000]
24hr Poly +S9 0.0000 [0.0000, 0.0000]
4hr pH3 +S9 0.0000 [0.0000, 0.0000]
24hr pH3 +S9 0.0000 [0.0000, 0.0000]
4hr γH2AX +S9 0.5414 [0.5233, 0.5588]



VX680

Score = 0.3231 [0.3184, 0.3278]
Rank = 22 [22, 22] (out of 30)

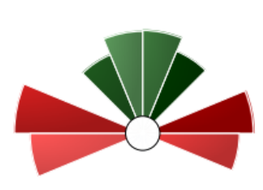
24hr γH2AX +S9 0.0000 [0.0000, 0.0000]
4hr p53 +S9 0.0000 [0.0000, 0.0000]
24hr p53 +S9 0.6610 [0.6608, 0.6611]
24hr MN +S9 0.6699 [0.6691, 0.6707]
24hr MN -S9 0.6525 [0.5915, 0.7122]
24hr p53 -S9 0.7114 [0.7106, 0.7121]
4hr p53 -S9 0.0000 [0.0000, 0.0000]
24hr γH2AX -S9 0.8912 [0.8867, 0.8958]
4hr γH2AX -S9 0.0000 [0.0000, 0.0000]
24hr pH3 -S9 0.0000 [0.0000, 0.0000]
4hr pH3 -S9 0.0000 [0.0000, 0.0000]
24hr Poly -S9 0.8427 [0.8384, 0.8470]
24hr Poly +S9 0.7414 [0.7373, 0.7456]
4hr pH3 +S9 0.0000 [0.0000, 0.0000]
24hr pH3 +S9 0.0000 [0.0000, 0.0000]
4hr γH2AX +S9 0.0000 [0.0000, 0.0000]



PF03814735

Score = 0.2841 [0.2828, 0.2855]
Rank = 21 [21, 21] (out of 30)

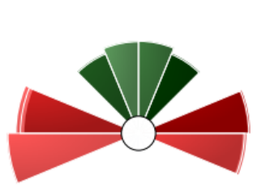
24hr γH2AX +S9 0.0000 [0.0000, 0.0000]
4hr p53 +S9 0.0000 [0.0000, 0.0000]
24hr p53 +S9 0.6996 [0.6993, 0.6999]
24hr MN +S9 0.7145 [0.7131, 0.7159]
24hr MN -S9 0.7531 [0.7488, 0.7575]
24hr p53 -S9 0.7171 [0.7150, 0.7192]
4hr p53 -S9 0.0000 [0.0000, 0.0000]
24hr γH2AX -S9 0.0000 [0.0000, 0.0000]
4hr γH2AX -S9 0.0000 [0.0000, 0.0000]
24hr pH3 -S9 0.0000 [0.0000, 0.0000]
4hr pH3 -S9 0.0000 [0.0000, 0.0000]
24hr Poly -S9 0.8804 [0.8722, 0.8888]
24hr Poly +S9 0.7811 [0.7762, 0.7861]
4hr pH3 +S9 0.0000 [0.0000, 0.0000]
24hr pH3 +S9 0.0000 [0.0000, 0.0000]
4hr γH2AX +S9 0.0000 [0.0000, 0.0000]



Chlorambucil

Score = 0.2802 [0.2775, 0.2829]
Rank = 20 [20, 20] (out of 30)

24hr γH2AX +S9 0.6087 [0.6002, 0.6177]
4hr p53 +S9 0.0000 [0.0000, 0.0000]
24hr p53 +S9 0.4653 [0.4611, 0.4694]
24hr MN +S9 0.5572 [0.5501, 0.5642]
24hr MN -S9 0.5518 [0.5459, 0.5576]
24hr p53 -S9 0.4402 [0.4335, 0.4470]
4hr p53 -S9 0.0000 [0.0000, 0.0000]
24hr γH2AX -S9 0.7230 [0.7189, 0.7271]
4hr γH2AX -S9 0.6199 [0.6163, 0.6237]
24hr pH3 -S9 0.0000 [0.0000, 0.0000]
4hr pH3 -S9 0.0000 [0.0000, 0.0000]
24hr Poly -S9 0.0000 [0.0000, 0.0000]
24hr Poly +S9 0.0000 [0.0000, 0.0000]
4hr pH3 +S9 0.0000 [0.0000, 0.0000]
24hr pH3 +S9 0.0000 [0.0000, 0.0000]
4hr γH2AX +S9 0.5167 [0.5135, 0.5201]



Cisplatin

Score = 0.2777 [0.2733, 0.2819]
Rank = 19 [19, 19] (out of 30)

24hr γH2AX +S9 0.6060 [0.5971, 0.6143]
4hr p53 +S9 0.0000 [0.0000, 0.0000]
24hr p53 +S9 0.4449 [0.4389, 0.4508]
24hr MN +S9 0.4871 [0.4841, 0.4902]
24hr MN -S9 0.4841 [0.4802, 0.4880]
24hr p53 -S9 0.4468 [0.4416, 0.4520]
4hr p53 -S9 0.0000 [0.0000, 0.0000]
24hr γH2AX -S9 0.6566 [0.6315, 0.6815]
4hr γH2AX -S9 0.7387 [0.7344, 0.7427]
24hr pH3 -S9 0.0000 [0.0000, 0.0000]
4hr pH3 -S9 0.0000 [0.0000, 0.0000]
24hr Poly -S9 0.0000 [0.0000, 0.0000]
24hr Poly +S9 0.0000 [0.0000, 0.0000]
4hr pH3 +S9 0.0000 [0.0000, 0.0000]
24hr pH3 +S9 0.0000 [0.0000, 0.0000]
4hr γH2AX +S9 0.5783 [0.5655, 0.5905]

Genistein

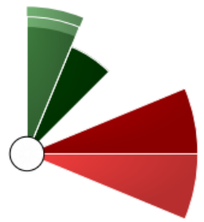
Score = 0.2462 [0.2412, 0.2511]
Rank = 18 [18, 18] (out of 30)

24hr yH2AX +S9 0.3542 [0.3496, 0.3587]
4hr p53 +S9 0.5205 [0.5101, 0.5306]
24hr p53 +S9 0.3062 [0.3013, 0.3111]
24hr MN +S9 0.3868 [0.3708, 0.4024]
24hr MN -S9 0.3685 [0.3572, 0.3796]
24hr p53 -S9 0.3163 [0.3102, 0.3224]
4hr p53 -S9 0.5238 [0.5149, 0.5325]
24hr yH2AX -S9 0.3866 [0.3731, 0.3999]
4hr yH2AX -S9 0.4479 [0.4446, 0.4513]
24hr pH3 -S9 0.0000 [0.0000, 0.0000]
4hr pH3 -S9 0.0000 [0.0000, 0.0000]
24hr Poly -S9 0.0000 [0.0000, 0.0000]
24hr Poly +S9 0.0000 [0.0000, 0.0000]
4hr pH3 +S9 0.0000 [0.0000, 0.0000]
24hr pH3 +S9 0.0000 [0.0000, 0.0000]
4hr yH2AX +S9 0.3287 [0.3280, 0.3293]

Methyl methanesulfonate

Score = 0.2391 [0.2290, 0.2490]
Rank = 17 [17, 17] (out of 30)

24hr yH2AX +S9 0.4778 [0.4693, 0.4857]
4hr p53 +S9 0.3466 [0.3345, 0.3583]
24hr p53 +S9 0.3947 [0.3131, 0.4755]
24hr MN +S9 0.3730 [0.3578, 0.3880]
24hr MN -S9 0.3476 [0.3336, 0.3613]
24hr p53 -S9 0.3321 [0.3252, 0.3390]
4hr p53 -S9 0.0000 [0.0000, 0.0000]
24hr yH2AX -S9 0.6134 [0.5966, 0.6301]
4hr yH2AX -S9 0.5201 [0.5154, 0.5246]
24hr pH3 -S9 0.0000 [0.0000, 0.0000]
4hr pH3 -S9 0.0000 [0.0000, 0.0000]
24hr Poly -S9 0.0000 [0.0000, 0.0000]
24hr Poly +S9 0.0000 [0.0000, 0.0000]
4hr pH3 +S9 0.0000 [0.0000, 0.0000]
24hr pH3 +S9 0.0000 [0.0000, 0.0000]
4hr yH2AX +S9 0.4198 [0.4184, 0.4210]



Dibenzo(a,l)pyrene

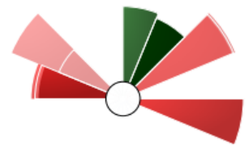
Score = 0.2138 [0.2095, 0.2181]
Rank = 16 [16, 16] (out of 30)

24hr yH2AX +S9 1.0000 [1.0000, 1.0000]
4hr p53 +S9 0.0000 [0.0000, 0.0000]
24hr p53 +S9 0.6420 [0.6389, 0.6451]
24hr MN +S9 0.7790 [0.7128, 0.8443]
24hr MN -S9 0.0000 [0.0000, 0.0000]
24hr p53 -S9 0.0000 [0.0000, 0.0000]
4hr p53 -S9 0.0000 [0.0000, 0.0000]
24hr yH2AX -S9 0.0000 [0.0000, 0.0000]
4hr yH2AX -S9 0.0000 [0.0000, 0.0000]
24hr pH3 -S9 0.0000 [0.0000, 0.0000]
4hr pH3 -S9 0.0000 [0.0000, 0.0000]
24hr Poly -S9 0.0000 [0.0000, 0.0000]
24hr Poly +S9 0.0000 [0.0000, 0.0000]
4hr pH3 +S9 0.0000 [0.0000, 0.0000]
24hr pH3 +S9 0.0000 [0.0000, 0.0000]
4hr yH2AX +S9 1.0000 [1.0000, 1.0000]

Methotrexate

Score = 0.2059 [0.2048, 0.2070]
Rank = 15 [14, 15] (out of 30)

24hr yH2AX +S9 0.8838 [0.8687, 0.8997]
4hr p53 +S9 0.0000 [0.0000, 0.0000]
24hr p53 +S9 0.6785 [0.6775, 0.6796]
24hr MN +S9 0.0000 [0.0000, 0.0000]
24hr MN -S9 0.0000 [0.0000, 0.0000]
24hr p53 -S9 0.7318 [0.7308, 0.7328]
4hr p53 -S9 0.0000 [0.0000, 0.0000]
24hr yH2AX -S9 1.0000 [1.0000, 1.0000]
4hr yH2AX -S9 0.0000 [0.0000, 0.0000]
24hr pH3 -S9 0.0000 [0.0000, 0.0000]
4hr pH3 -S9 0.0000 [0.0000, 0.0000]
24hr Poly -S9 0.0000 [0.0000, 0.0000]
24hr Poly +S9 0.0000 [0.0000, 0.0000]
4hr pH3 +S9 0.0000 [0.0000, 0.0000]
24hr pH3 +S9 0.0000 [0.0000, 0.0000]
4hr yH2AX +S9 0.0000 [0.0000, 0.0000]



2-Aminoanthracene

Score = 0.1926 [0.1690, 0.2156]
Rank = 14 [13, 15] (out of 30)

24hr yH2AX +S9 0.0000 [0.0000, 0.0000]
4hr p53 +S9 0.6651 [0.6539, 0.6765]
24hr p53 +S9 0.4555 [0.4552, 0.4559]
24hr MN +S9 0.4891 [0.4870, 0.4911]
24hr MN -S9 0.0000 [0.0000, 0.0000]
24hr p53 -S9 0.0000 [0.0000, 0.0000]
4hr p53 -S9 0.3395 [0.0000, 0.6700]
24hr yH2AX -S9 0.4607 [0.4375, 0.4836]
4hr yH2AX -S9 0.0000 [0.0000, 0.0000]
24hr pH3 -S9 0.0000 [0.0000, 0.0000]
4hr pH3 -S9 0.0000 [0.0000, 0.0000]
24hr Poly -S9 0.0000 [0.0000, 0.0000]
24hr Poly +S9 0.0000 [0.0000, 0.0000]
4hr pH3 +S9 0.0000 [0.0000, 0.0000]
24hr pH3 +S9 0.0000 [0.0000, 0.0000]
4hr yH2AX +S9 0.6717 [0.6707, 0.6728]

Thiotepa

Score = 0.1860 [0.1835, 0.1885]
Rank = 13 [13, 14] (out of 30)

24hr yH2AX +S9 0.5604 [0.5555, 0.5655]
4hr p53 +S9 0.0000 [0.0000, 0.0000]
24hr p53 +S9 0.4210 [0.4134, 0.4285]
24hr MN +S9 0.5381 [0.5326, 0.5436]
24hr MN -S9 0.4808 [0.4760, 0.4856]
24hr p53 -S9 0.3661 [0.3574, 0.3747]
4hr p53 -S9 0.0000 [0.0000, 0.0000]
24hr yH2AX -S9 0.6102 [0.6019, 0.6185]
4hr yH2AX -S9 0.0000 [0.0000, 0.0000]
24hr pH3 -S9 0.0000 [0.0000, 0.0000]
4hr pH3 -S9 0.0000 [0.0000, 0.0000]
24hr Poly -S9 0.0000 [0.0000, 0.0000]
24hr Poly +S9 0.0000 [0.0000, 0.0000]
4hr pH3 +S9 0.0000 [0.0000, 0.0000]
24hr pH3 +S9 0.0000 [0.0000, 0.0000]
4hr yH2AX +S9 0.0000 [0.0000, 0.0000]

PhIP

Score = 0.1735 [0.1670, 0.1797]
Rank = 12 [12, 12] (out of 30)

24hr yH2AX +S9 0.9455 [0.8930, 0.9951]
4hr p53 +S9 0.0000 [0.0000, 0.0000]
24hr p53 +S9 0.5615 [0.5489, 0.5740]
24hr MN +S9 0.5059 [0.4938, 0.5179]
24hr MN -S9 0.0000 [0.0000, 0.0000]
24hr p53 -S9 0.0000 [0.0000, 0.0000]
4hr p53 -S9 0.0000 [0.0000, 0.0000]
24hr yH2AX -S9 0.0000 [0.0000, 0.0000]
4hr yH2AX -S9 0.0000 [0.0000, 0.0000]
24hr pH3 -S9 0.0000 [0.0000, 0.0000]
4hr pH3 -S9 0.0000 [0.0000, 0.0000]
24hr Poly -S9 0.0000 [0.0000, 0.0000]
24hr Poly +S9 0.0000 [0.0000, 0.0000]
4hr pH3 +S9 0.0000 [0.0000, 0.0000]
24hr pH3 +S9 0.0000 [0.0000, 0.0000]
4hr yH2AX +S9 0.7627 [0.7357, 0.7884]

Ethyl methanesulfonate

Score = 0.1676 [0.1577, 0.1772]
Rank = 11 [11, 11] (out of 30)

24hr yH2AX +S9 0.4162 [0.3967, 0.4345]
4hr p53 +S9 0.0000 [0.0000, 0.0000]
24hr p53 +S9 0.2293 [0.2121, 0.2464]
24hr MN +S9 0.3417 [0.3063, 0.3766]
24hr MN -S9 0.3107 [0.2821, 0.3387]
24hr p53 -S9 0.2191 [0.2037, 0.2345]
4hr p53 -S9 0.0000 [0.0000, 0.0000]
24hr yH2AX -S9 0.5089 [0.4815, 0.5361]
4hr yH2AX -S9 0.3697 [0.3634, 0.3757]
24hr pH3 -S9 0.0000 [0.0000, 0.0000]
4hr pH3 -S9 0.0000 [0.0000, 0.0000]
24hr Poly -S9 0.0000 [0.0000, 0.0000]
24hr Poly +S9 0.0000 [0.0000, 0.0000]
4hr pH3 +S9 0.0000 [0.0000, 0.0000]
24hr pH3 +S9 0.0000 [0.0000, 0.0000]
4hr yH2AX +S9 0.2853 [0.2775, 0.2928]

Brefeldin A

Score = 0.1180 [0.1173, 0.1186]
Rank = 10 [10, 10] (out of 30)

24hr yH2AX +S9 0.0000 [0.0000, 0.0000]
4hr p53 +S9 0.0000 [0.0000, 0.0000]
24hr p53 +S9 0.0000 [0.0000, 0.0000]
24hr MN +S9 0.5724 [0.5687, 0.5761]
24hr MN -S9 0.6336 [0.6335, 0.6337]
24hr p53 -S9 0.0000 [0.0000, 0.0000]
4hr p53 -S9 0.0000 [0.0000, 0.0000]
24hr yH2AX -S9 0.0000 [0.0000, 0.0000]
4hr yH2AX -S9 0.0000 [0.0000, 0.0000]
24hr pH3 -S9 0.0000 [0.0000, 0.0000]
4hr pH3 -S9 0.0000 [0.0000, 0.0000]
24hr Poly -S9 0.6819 [0.6752, 0.6885]
24hr Poly +S9 0.0000 [0.0000, 0.0000]
4hr pH3 +S9 0.0000 [0.0000, 0.0000]
24hr pH3 +S9 0.0000 [0.0000, 0.0000]
4hr yH2AX +S9 0.0000 [0.0000, 0.0000]

Hydroxyurea

Score = 0.1076 [0.1046, 0.1106]
Rank = 9 [9, 9] (out of 30)

24hr yH2AX +S9 0.3653 [0.3634, 0.3671]
4hr p53 +S9 0.0000 [0.0000, 0.0000]
24hr p53 +S9 0.2391 [0.2326, 0.2455]
24hr MN +S9 0.2397 [0.2291, 0.2503]
24hr MN -S9 0.2798 [0.2700, 0.2893]
24hr p53 -S9 0.2270 [0.2195, 0.2345]
4hr p53 -S9 0.0000 [0.0000, 0.0000]
24hr yH2AX -S9 0.3712 [0.3593, 0.3829]
4hr yH2AX -S9 0.0000 [0.0000, 0.0000]
24hr pH3 -S9 0.0000 [0.0000, 0.0000]
4hr pH3 -S9 0.0000 [0.0000, 0.0000]
24hr Poly -S9 0.0000 [0.0000, 0.0000]
24hr Poly +S9 0.0000 [0.0000, 0.0000]
4hr pH3 +S9 0.0000 [0.0000, 0.0000]
24hr pH3 +S9 0.0000 [0.0000, 0.0000]
4hr yH2AX +S9 0.0000 [0.0000, 0.0000]

Cyclophosphamide

Score = 0.0908 [0.0877, 0.0940]
Rank = 8 [8, 8] (out of 30)

24hr yH2AX +S9 0.5587 [0.5552, 0.5619]
4hr p53 +S9 0.0000 [0.0000, 0.0000]
24hr p53 +S9 0.3839 [0.3785, 0.3894]
24hr MN +S9 0.4221 [0.4153, 0.4289]
24hr MN -S9 0.0175 [0.0000, 0.0345]
24hr p53 -S9 0.0000 [0.0000, 0.0000]
4hr p53 -S9 0.0000 [0.0000, 0.0000]
24hr yH2AX -S9 0.0713 [0.0538, 0.0886]
4hr yH2AX -S9 0.0000 [0.0000, 0.0000]
24hr pH3 -S9 0.0000 [0.0000, 0.0000]
4hr pH3 -S9 0.0000 [0.0000, 0.0000]
24hr Poly -S9 0.0000 [0.0000, 0.0000]
24hr Poly +S9 0.0000 [0.0000, 0.0000]
4hr pH3 +S9 0.0000 [0.0000, 0.0000]
24hr pH3 +S9 0.0000 [0.0000, 0.0000]
4hr yH2AX +S9 0.0000 [0.0000, 0.0000]

Thapsigargin

Score = 0.0840 [0.0829, 0.0851]
Rank = 7 [7, 7] (out of 30)

24hr yH2AX +S9 0.0000 [0.0000, 0.0000]
4hr p53 +S9 0.0000 [0.0000, 0.0000]
24hr p53 +S9 0.0000 [0.0000, 0.0000]
24hr MN +S9 0.6288 [0.6175, 0.6399]
24hr MN -S9 0.7152 [0.7082, 0.7221]
24hr p53 -S9 0.0000 [0.0000, 0.0000]
4hr p53 -S9 0.0000 [0.0000, 0.0000]
24hr yH2AX -S9 0.0000 [0.0000, 0.0000]
4hr yH2AX -S9 0.0000 [0.0000, 0.0000]
24hr pH3 -S9 0.0000 [0.0000, 0.0000]
4hr pH3 -S9 0.0000 [0.0000, 0.0000]
24hr Poly -S9 0.0000 [0.0000, 0.0000]
24hr Poly +S9 0.0000 [0.0000, 0.0000]
4hr pH3 +S9 0.0000 [0.0000, 0.0000]
24hr pH3 +S9 0.0000 [0.0000, 0.0000]
4hr yH2AX +S9 0.0000 [0.0000, 0.0000]

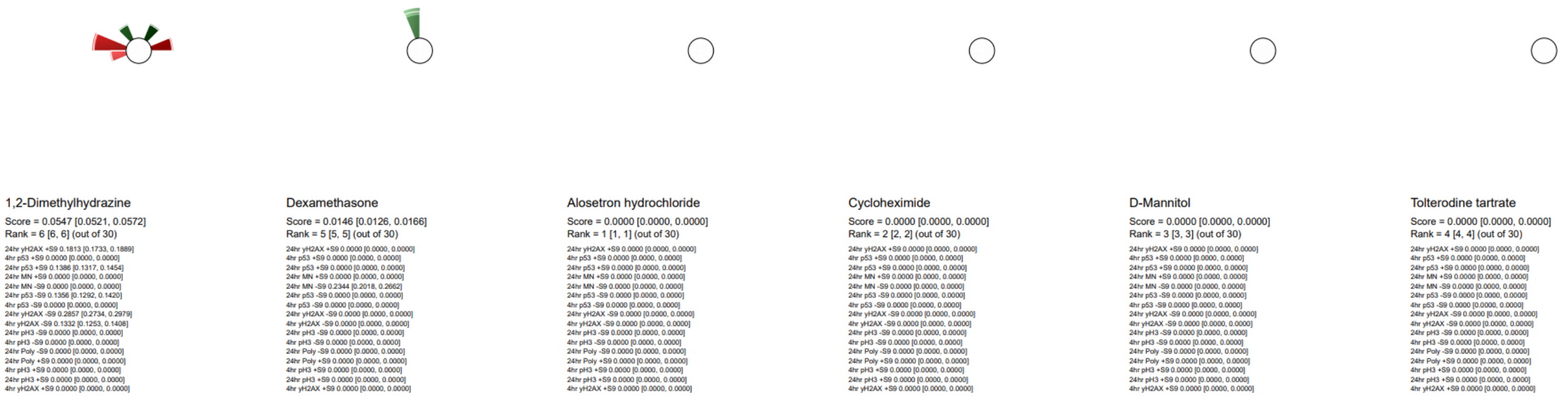


Figure 2.3.3.3 – ToxPi profiles (with bootstrap CI) generated from BMD data for all 30 aneugenic/clastogenic compounds assessed in this component, analysing potency across MultiFlow/MicroFlow endpoints at 4/24-hour timepoints in the presence/absence of S9. Each slice represents a biomarker at a specific time point, both with and without S9. The shading in the centre of each circle denotes uncertainty, with darker shades indicating higher levels of uncertainty. The radius of each slice, from the centre to either the edge or a white line within the slice, illustrates the potency of each biomarker, while the width corresponds to user-defined weightings. The brighter regions represent the bootstrap confidence interval (CI), with the outer edge marking the upper CI and the inner edge indicating the lower CI for each slice.

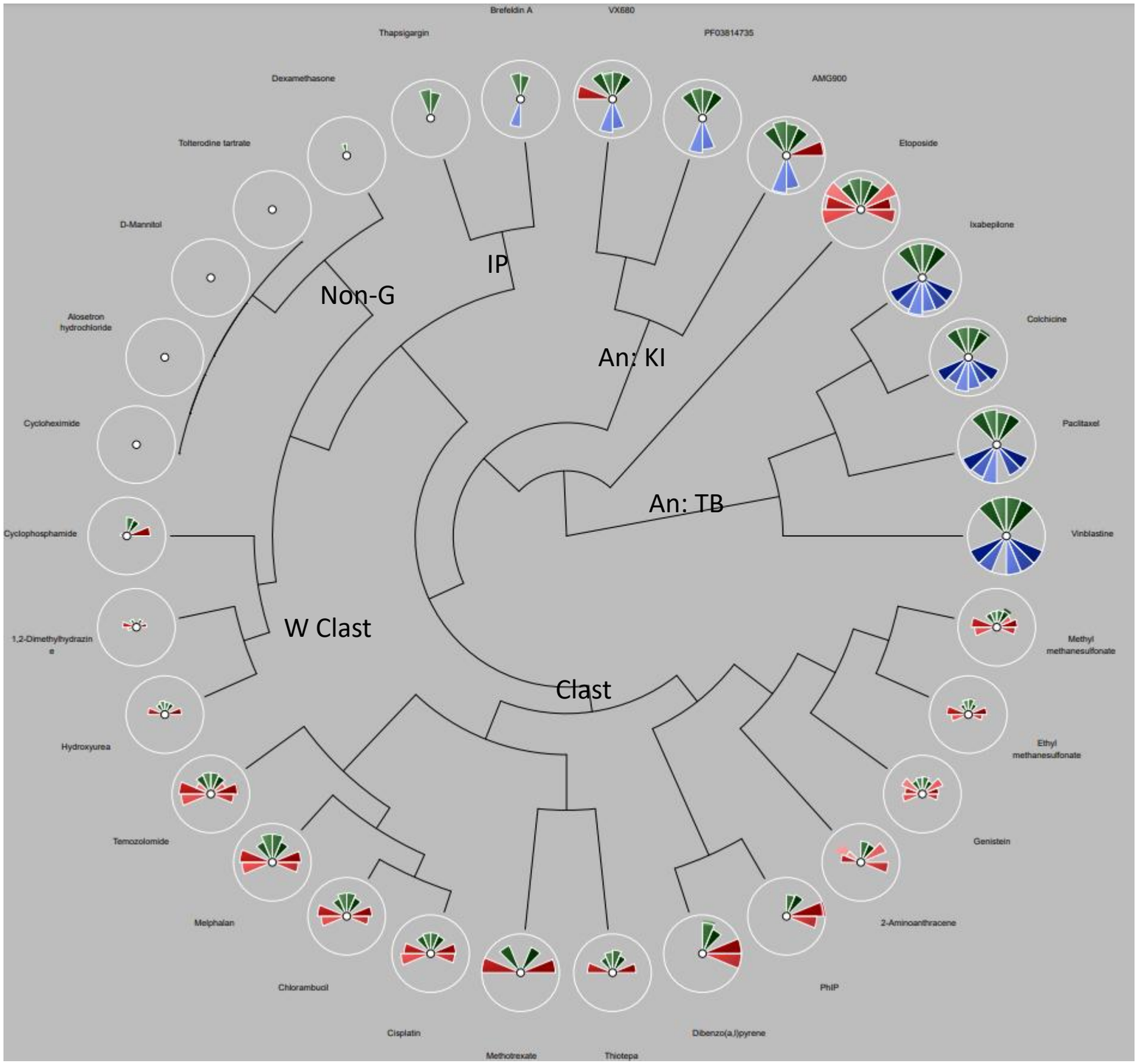


Figure 2.3.3.4 – Complete hierarchical clustering of ToxPi profiles was performed using complete AGNES clustering to identify chemical groupings, each color-coded based on their biomarker activation and MoA. ToxPi separated the compounds into their known subgroupings: clastogens (Clast), weak clastogens (W Clast), aneugenic tubulin binders (An: TB), aneugenic aurora kinase inhibitors (An: KI), irrelevant positives (IP), and non-genotoxicants (Non-G), with some compounds being identified as irrelevant positives for having unexpected projections as they were known non-genotoxic compounds, which included Brefeldin A and Thapsigargin.

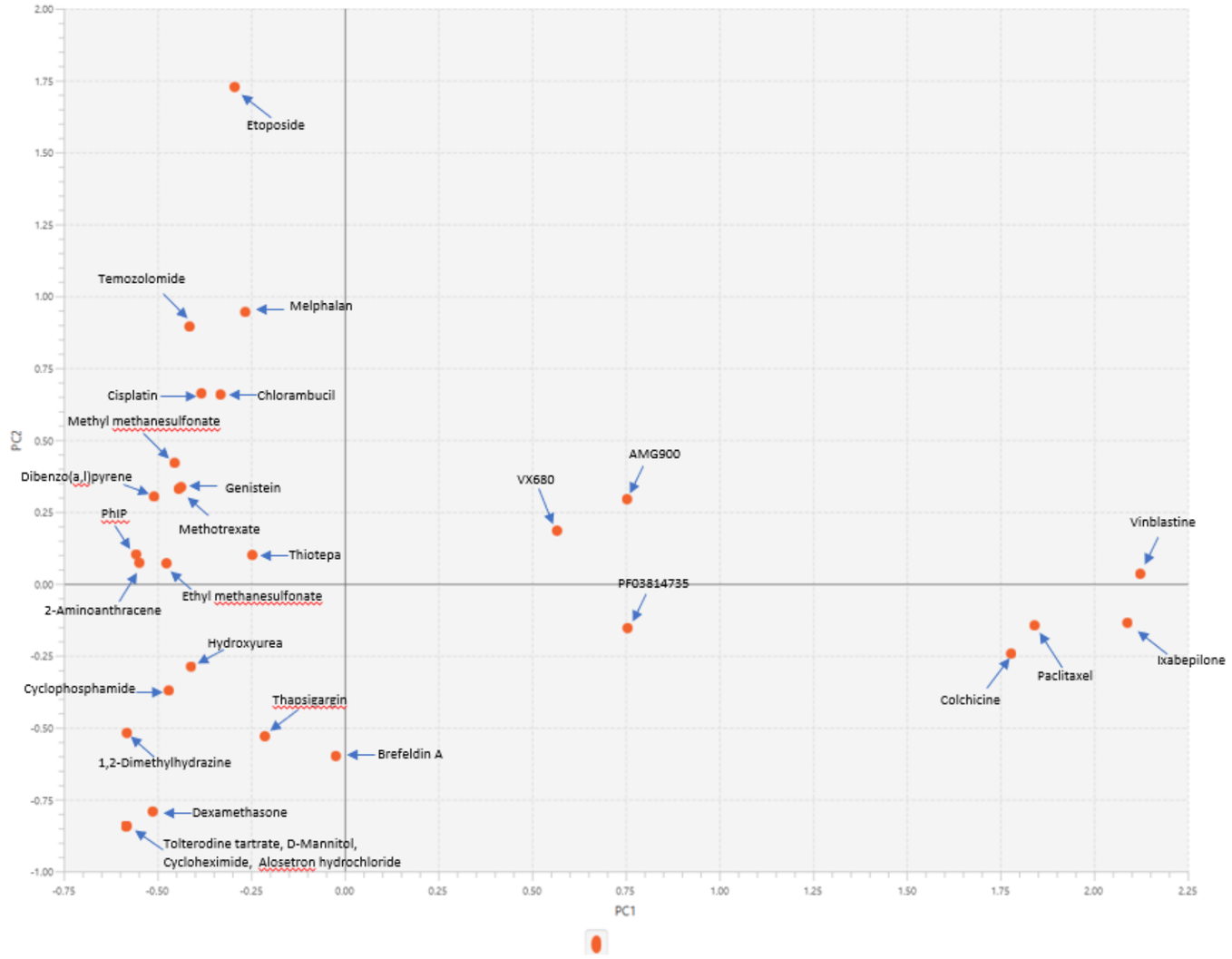


Figure 2.3.3.5 – ToxPi K-means divisive/PCA clustering where each compound is grouped based on variation across the ToxPi scores/genotoxicity and biomarker sensitivity. Unlike previous chapters, as a fixed cluster number was not set, the compounds were instead manually observed, and compounds were grouped based on PC1 and PC2 analyses. PC1 compounds with higher ToxPi scores are towards the right of the graph, however, no such pattern was found for PC2 as more varying endpoints were assessed. Instead, compounds with similar modes of action grouped within closer Euclidian distances from each other, while distant compounds with greatly different biomarker activation ToxPi profiles had greater Euclidian distances between them. Importantly, as DNA damage signalling rather is primarily captured instead of direct mutagenicity, the clustering is better interpreted to reflect similar/differing degrees of genotoxic stress rather than discrete mechanistic classes.

2.4 – Discussion

This chapter conducted a comprehensive analysis across three key components, as previously outlined in the results section 2.3. This discussion aims to critically evaluate the applicability and effectiveness of the methodologies employed, particularly focusing on the rationale behind the outcomes derived from the multiplexed imaging techniques used in this study. The techniques were leveraged to generate genotoxic potency rankings, develop biomarker response profiles, and apply unsupervised machine learning algorithms. By utilising the Benchmark Dose (BMD) approach to calculate CIs and to cluster assay data, this study seeks to create a framework for optimizing information extraction and utilization. This approach holds the potential to enhance the retention of pharmaceutical compounds that may be viable candidates for treating various illnesses and diseases.

If the BMD could not be modelled due to the response not reaching the CES or the true BMD estimate exceeding the tested concentration range, no value was recorded. When PROAST successfully modelled the BMDL and BMDU, the values were recorded for each chemical, biomarker, timepoint, and S9 activation property. These BMD CIs provide insight into how metabolic activation affects clastogenic genotoxic potency, however, as stated in the previous section, fails to combine them into a singular figure that actively compares all endpoints simultaneously; hence ToxPi was implemented to generate further figures.

2.4.1 – BMD CIs High-Low Charts

The experimental dose – response data from the MultiFlow/MicroFlow assays was inserted into PROAST as described in section 2.2.5 to map and calculate the 90% BMD CIs, which were subsequently reproduced in the high-low charts in the previous results sections. However, if the BMD could not be modelled due to the response not reaching the CES or the true BMD estimate exceeding the tested concentration range, no value was recorded. When PROAST successfully modelled the BMDL and BMDU, the values were recorded for each chemical, biomarker, timepoint, and S9 activation property. These BMD CIs provide insight into how metabolic activation affects clastogenic genotoxic potency.

Several options and methods are available to model and display BMD CI for comparative purposes, as highlighted by the high-low charts produced in each subsection. These BMD CI graphs provide a means to illustrate the differences and overlaps in compound potency across

various singular endpoints and it generates 90% CIs that function as error bars, indicating a range where the true BMD value might lie, reflecting the uncertainty inherent in the modelling process and offering a more robust representation of the data, despite how BMD CIs as a potency assessment tool may be considered problematic because it cannot provide a single, definitive BMD value. However, more importantly, they fail to integrate all endpoints, time points, and metabolic states into a unified analytical metric that would allow for a direct comparison of overall compound potency and individual endpoints. This means that high-low figures show some potency ranking ability in comparing compounds across endpoints quite well; however, it can only display this information for one endpoint at a time which is substantially limits the comparative capacity of this analysis methodology as it there are more efficient methods to represent the responses across all biomarkers more effectively using automated tools such as ToxPi. This limitation would complicate the comparative analysis across genotoxicity assays, particularly when assessing hazards/drug safety. Additionally, the current approach does not account for the varying significance of certain endpoints across regulatory frameworks or apply weighted significance to biomarkers. By relying solely on raw BMD values to construct high-low plots, the analysis can rank potencies and compare compounds across different endpoints but remains restricted to a single endpoint at a time. This limitation reduces the analysis's comparative effectiveness, especially when more advanced methods, such as ToxPi, can offer a more comprehensive evaluation across multiple biomarkers simultaneously.

2.4.2 – Topoisomerase II inhibitors outputs

This component attempted was to observe how PROAST/ToxPi would display compounds with differing genotoxic MoA with an identical drug target and observing their potency rankings, biomarker scoring and collective clustering. Key findings include that the potency ranking remained mostly consistent with figure 2.3.1.7 and the BMD CIs in the original Wheeldon et al., (2020) study except for doxorubicin, which was ranked higher compared to teniposide in the ToxPi scoring figure 2.3.1.7 with a ToxPi scores of 0.9017 and 0.8562 respectively. However, the separated BMD CIs between both compounds overlapped at the 4/24hr p53 endpoints suggesting no significant difference in potency at those biomarkers, however, the ToxPi scoring suggests that overall, they are significant differences in potency overall due to the lack of overlaps in figure 2.3.1.7. This is likely due to the remaining endpoints not overlapping, which would introduce differences when all biomarkers are combined into a singular score, and furthermore, visual differences between figure 2.3.1.7

and the original BMD CIs figure Wheeldon et al., (2020) are most likely caused by differences in rounding when the BMDs are logarithmically transformed.

Hierarchical clustering figure 2.3.1.9 shows compounds clustering into their distinctive groupings aligned to their structural QSAR relationships, apart from emodin, which unexpectedly grouped with genistein. This disparity is likely caused by the lower potency activation across all its biomarkers, and with reduced potency, the data skewed, leading emodin to share biomarker potency activation properties with genistein across all endpoints. This resulted in emodin clustering with isoflavones rather than with anthraquinones implying that segregation based purely on structural information was not followed and potency ranking/profiles skewed data away from expected groupings since we would expect compounds with similar biomarker activation properties and structural information to classify with each other. The K-means clustering applied shows the Euclidian relationship between compounds using divisive clustering and all compounds graphically clustered identically to the ToxPi hierarchical clustering system. However, this is subject to the reporters selected, so any modification to the biomarker panel is expected to alter compound positioning and cluster groupings.

Despite the strength of this dataset, BMD required to be set to ensure that datasets did not skew heavily, and initial limited information was able to be extracted as only Multiflow data was used and further assay data incorporation from polyploidy, RNC and MN could provide further details into compounds, further bringing question in potential changes to their potency ranking and clustering. CI data alone does provide further insight into endpoint specific comparisons, however, ToxPi facilitates a more integrated interpretation by enabling simultaneous comparison across all endpoints.

2.4.3 – 10 Clastogens outputs

In the original study design performed by Tian et al., (2020), to generate the BMD CI, the CES chosen for both p53 and γH2aX endpoints predominantly assessed in this component was 0.3 to a slightly lower suggestion for the medium effect size, while MN and polyploidy CES values remain at +1.0 and +5.0 respectively, however, neither of these latter endpoints are included in this analysis. Additionally, the absence of RNC CIs in this analysis further contributed to the deviations observed between these mentioned studies and this investigation. The covariates are set as the S9 conditions, and the same exponential models

used in the previous section are applied to fit the dose response curves to extract the BMDs. Reported BMDs were split into biomarkers at different time points in the presence and absence of S9 and where PROAST was unable to fit the appropriate dose-response models/the BMD value exceeded the top concentration tested; no BMD values were recorded in those conditions as stated in section 2.2.5.2. Furthermore, for the original hierarchical clustering using the ward. D algorithm, an S9 potentiation ratio was determined, however, unlike the original Tian et al., (2020) and Wheeldon et al., (2021) studies, a subtractive approach was utilised as opposed to the divisive approach to obtain metabolic activation differences in both S9- and S9+ conditions for each compound. Furthermore, multiple ToxPi analyses and extra clustering figures needed to be produced to more effectively perform comparative analysis between compounds, especially due to limitations in the tools that ToxPi's provides, such as increasing specific clustering options/functionality**.** Reduced user control over biomarker inclusion made it challenging to differentiate endpoints aligned with clastogenic mechanisms from those reflecting alternative genotoxic pathways. For example, p-H3, aneugen-associated reporter, was excluded to minimise mechanistic noise, dataset skewing and the improvement in the interpretability of DNA damaging strand break-related biomarkers. This approach reflects a targeted analytical strategy for clastogenic compounds, rather than selective data handling while recognising that some compounds may exhibit mixed MoAs. Regarding ToxPi profiles (figure 2.3.2.13), biomarker activation appears consistent with the Wheeldon et al., (2021) study, however, to draw these comparisons, more interpretation is required with the original paper. There are differences in the profiling due to the methodologies applied to analysing individual endpoints as BMD CIs and combined ToxPi analyses. The approaches inherently generate variation in outputs relative to the original study design meaning any differences are considered methodological rather than indicative of insufficient data quality. Hence, this explains why compound potency, metabolism and clustering were preserved across analyses. ToxPi provided a stronger comprehensive and condensed breakdown of biomarker activations making it easier for comparative purposes and maximising information extraction with the ability to apply modifications to the study to extract new information. For example, having combined metrics for the ToxPi figures and separating them out based on metabolic activation across figures 2.3.2.9, 2.3.2.10 and 2.3.2.11 respectively. Differences between the metabolically separated ToxPi rankings and overall, 10 clastogens final ranks are due to the increased number of combined metrics in the overall analysis, leaning the data away from its expected results. For example, dibenzo[a,l]pyrene S9+ ranks a lot higher compared to dibenzo[a,l]pyrene S9-

which is consistent with the ToxPi profiles, *hclust*/ToxPi clustering and original Wheeldon et al., (2021) study heat map, however, when more metrics are combined in the overall ToxPi ranking, the ToxPi score and ranking for dibenzo[a,l]pyrene appears to be less potent than mitomycin C, despite how well the combined figures visualises the ToxPi profiles. This leads back into a need for multiple ToxPi runs being required to succinctly compare compounds biomarker responses and potency figures. The potency ranking, profiling and clustering generated for both *hclust* and ToxPi are both consistent with the literature and any unexpected findings mentioned are due to the reasons stated in this section.

2.4.3.1 – *hclust* functionality and metabolically separated ToxPi outputs

The Euclidian hierarchical clustering dendrogram was produced from the *hclust* function in R Gui produced the clustering that is observed in figure 2.3.2.14, where the ward.D algorithm was applied. Although this clustering algorithm worked as proposed in the previous section, additional ToxPi analysis was required to clearly distinguish between potencies in the presence and absence of S9, where metabolic states between compounds were separated in figure 2.3.2.11, so any differences between potencies are clearly illustrated. The metabolically separated ToxPi rankings clearly shows how much S9 changes compound and by how much relative to each dataset, therefore it is supplementing the Euclidian hierarchical clustering allows for strong classifications of parent compound and their metabolites. Resorcinol and mitomycin C show strong clustering in the figure 2.3.2.14, dendrogram and both metabolites ranked significantly lower in potency compared to their parent compounds in the ToxPi ranks (figure 2.3.2.11) with no overlaps between each set of CIs. Meanwhile, compounds like DMBA and 2-aminoanthracene, displayed similar increases in genotoxic potency in the presence of S9, also forming a distinct cluster. Interestingly, both benzo[a]pyrene and cyclophosphamide, which require metabolic activation to their potent forms, displayed similar modest increases in potency in the presence of S9. This observation aligns with the ToxPi rankings and Euclidean hierarchical cluster figures, further supporting the role of metabolic activation in their biological activity. Even though ToxPi hierarchical clustering system separated these compounds, when observing their ToxPi profiles, they appear to be more related than what the original results suggest, which likely highlights the differences between the AGNES approaches being implemented where ToxPi uses mathematically transformed figures (0 – 1 ToxPi scores across all reporters) when clustering as opposed to the original *hclust* function which uses the original BMD figures. Similarly to this, PhIP and dibenzo[a, l]pyrene, both which exhibit a significant increase in potency when

metabolically activated, which both the *hclust* function and ToxPi ranking/profiles exhibiting this, despite what figure 2.3.2.14 suggests, alluding to differences in clustering approaches used between programs. This further suggests that potentially selecting a different number of clusters might bring these compounds closer to similarity than what is initially presented. Finally, 2-acetylaminofluorene and diethylnitrosamine, two compounds that did not show a significant difference in genotoxic potency regardless of the presence or absence of S9, were clustered together in the analysis, an outcome aligns with expectations, as their behaviour was consistent across conditions. These findings further allude to how no atypical clustering or potency ranking was observed in this part of the investigation and all findings remain consistent. However, diethylnitrosamine is a known clastogens and a ToxPi scores of 0.000 initially suggests that it has virtually no potency, which is untrue. ToxPi operates relatively to each dataset, so it is really suggesting that compared to the other 9 compounds, diethylnitrosanime potency is significantly lower across the MultiFlow endpoints tested within the bounds of the investigation. When the calculation of the ToxPi score (as described in 2.2.7) is considered and the original BMDs are transformed to the biological testing range of 0.0001 – 10000 μM is taken into account, this result is likely due to the original data across most endpoints having either no recorded BMDs or a high proportion of BMDs amounting to 10000 μM, meaning that mathematically, a ToxPi score of 0.0000 would be produced.

Furthermore, diethylnitrosamine is a known potent carcinogen, yet its ToxPi score of 0.000 may misleadingly suggest a lack of potency. However, this interpretation overlooks the fact that ToxPi scores are relative to other compounds within the same dataset. In this context, the low score reflects diethylnitrosamine's comparatively lower potency across various endpoints within this dataset and does not imply an absence of its carcinogenicity. This result is influenced by the lack of BMD values for certain endpoints or the presence of higher BMDs for diethylnitrosamine compared to other compounds, with some values adjusted to 10,000 μM, as outlined in the methods. These variations could be attributed to the nature of O6-alkylguanine DNA lesions , which likely require specialized detection methods—such as specific probes, antibodies, or microbes—making their detection in the MultiFlow assay more challenging as the γ-H2AX endpoint does is activated when dsDNA breaks are detected, likely reflecting downstream replication stress from the pro-mutagenic base modifications rather than the initial lesions themselves. However, it does not specifically target O6-alkylguanine lesions associated with diethylnitrosamine (Bartsch & Montesano, 1984). This may have led to higher BMD values in the original MultiFlow experiments,

lowering its ToxPi score. Consequently, the score of 0.000 reflects diethylnitrosamine's relatively reduced potency in this particular study, despite its recognized mutagenic potential. A potential solution to this issue could involve scaling the results against a set of positive and negative controls or introduce a mutation-based reporter into the biomarker panel. This approach would help mitigate the effects of potency disparities between stronger and weaker genotoxicants (with very low and high BMDs respectively), thereby reducing the likelihood of potentially misleading interpretations.

2.4.4 – Comprehensive 30 aneugen/clastogens outputs

Displaying BMD CIs from Avlasevich et al., (2020) as the CI charts would be disadvantageous as it does provide insight into very specific detailed BMD data, however, fails to visualise it in the most efficient and economical way possible, hence only ToxPi analyses were ran for this assessment, as the shortcomings in components once and two regarding limited compounds and biomarkers were no longer issues in this department. Instead, a comprehensive demonstration of the effectiveness of ToxPi's visualisation capabilities were displayed for MultiFlow and MicroFlow compound assessment as both potency ranking and ToxPi profiling remained consistent with the literature presented and allow for highly detailed assessment of compounds. However, the increase in compounds/biomarkers had resulted in irrelevant positive results being detected, even with the 0.0001 – 10000 μM filtering having been applied. Furthermore, the density of the data had resulted in differences in metabolically activated (S9) compounds vs parent compounds to be less intuitively shown and represented. For the hierarchical clustering, the expanded dataset produced groupings that naturally segregated according to their known MoAs, with clusters reflecting relative similarities in MoA and endpoint responses, apart from some irrelevant positives (thapsigargin and brefeldin A). Across the assessment, 6 distinct groupings were found (weak clastogens, irrelevant positives, aneugens kinase inhibitors, aneugens tubulin binding, and non-genotoxicants) as shown in figure 2.3.3.4. Specific biomarkers option selections for distinct compounds were unable to be performed in ToxPi due to limitations in the tools provided when it comes to the clustering component, which proved problematic when analysing clastogens with irrelevant biomarker phosphor-histone 3, which may skew overall ranking for clastogens down and contribute towards alterations in clustering. For Figure 2.3.3.5, PC2 could not be clearly defined. While the K-means clustering PC1 aligns with compound potency, PC2 remains ambiguous due to the influence

of too many overlapping factors, however, any MoA interpretation is made after the clustering, not required beforehand. Additionally, the inability to assess higher principal components (e.g., PC3, PC4, etc.) is a limitation of ToxPi. To further elucidate the patterns in this plot, additional analysis using third-party software, such as JMP, would be necessary. Despite this, aurora kinase inhibiting and tubulin binding aneugens that activated p-H3/polyploidy biomarkers were separated from the remaining clastogens, irrelevant positives and non-genotoxic compounds, showing a layer of comparative assessment is still able to be performed.

2.4.5 - Improvements

The BMD is analysed only after the original dose-response data has been accurately modelled. This approach ensures that the derived BMD values are based on well-fitted dose-response curves, minimising uncertainty and enhancing the reliability of the findings. However, accurately modelling the dose-response relationship depends on selecting appropriate endpoints that truly reflect the biological effects of interest. In some cases, existing endpoints may have limitations that restrict the depth of interpretation, necessitating the extraction of additional mechanistic insights. Expanding the analysis to incorporate multiple MoAs could provide a more comprehensive understanding of the compound's effects, improving confidence in the derived BMD values.

A critical consideration in this context is the integration of multiple assays, as relying solely on multiflow/microflow assay data may not provide a complete representation of a compound's potency, its MoAs, or the physical manifestations associated with micronuclei formation. While genotoxicity assays, such as MultiFlow and Microflow offer valuable information, they may not capture the full spectrum of cellular responses. Therefore, incorporating complementary assays could bridge this gap, providing a more holistic view of compound toxicity and enabling more robust dose-response analysis. This would, in turn, enhance decision-making processes related to compound potency assessment.

Additionally, the applicability of the BMD/ToxPi approach needs to be evaluated across a diverse range of assays to determine its versatility and reliability in different experimental settings. Assessing multiple endpoints across various assays would provide insights into the consistency of the approach and its potential for broader implementation in toxicological studies. Understanding how the BMD/ToxPi methodology performs across different

biological contexts is crucial for establishing it as a standardised tool for dose-response analysis.

Another important aspect is ensuring the consistency of data transformations used in the analysis. It is essential to determine whether these transformations are unique to this particular workflow or if they should be standardised for broader applications across different assay types. Consistency in data processing ensures comparability between studies, reduces variability, and strengthens the overall reliability of the results.

Finally, further refinement is needed in terms of biomarker specificity and the weighting strategies used in the analysis. The selection and weighting of biomarkers play a critical role in accurately characterising dose-response relationships and predicting toxicological outcomes. Careful consideration must be given to how biomarkers are chosen and prioritised within the BMD/ToxPi framework. Additionally, the implementation of advanced analytical tools such as BMDExpress could further enhance the robustness of the methodology, allowing for more sophisticated data interpretation and improved decision-making in toxicological assessments.

2.5 – Chapter Conclusions

The methodologies discussed in this chapter illustrate the capacity to visualize dose-response data using the BMD approach effectively. When combined with PROAST and ToxPi analyses, the BMD method offers a compelling alternative to the manual generation of CIs, serving as a valuable tool for comparative compound assessment. PROAST and ToxPi enhance data visualization by using dimensionality reduction techniques, which condense complex multiplexed chemical dose-response datasets. These tools simplify the interpretation of data by transforming compound biomarker response profiles, rankings of genotoxic compounds, and unsupervised clustering algorithms into accessible outputs like graphs, charts, and images. While Toxpi cannot directly classify MoAs, it enables the relative potency contribution of each endpoint to be effectively visualised, making it useful for identifying subtle effects in compounds, such as weak aneugens.

Furthermore, our findings suggest that a broader evaluation of compounds is necessary, especially given the limitations associated with drugs that have unique MoAs, which complicate grouping strategies. ToxPi has demonstrated its effectiveness in ranking, profiling, and clustering multiple compounds simultaneously, making it particularly useful for applications such as MultiFlow and MicroFlow assays. However, it is critical to implement

data filtering processes to prevent skewed results from BMD values that are exceedingly high or low, such as those above 10000 µM or below 0.0001 µM. Consequently, we established biologically relevant limits to maintain transparency and consistency in our analyses and prevent cellular damage through excessive cytotoxicity.

It is crucial to recognize that ToxPi's results depend heavily on the numerical data input, as the program utilizes machine learning algorithms to process these values. Although setting BMD limits helps to mitigate skewing, it can also result in ToxPi scores of 0.0000 if all endpoints reach 10000 µM, potentially leading to misleading outcomes. This risk is particularly significant if compounds with some degree of genotoxicity are assigned an overall potency score of 0.0000. Additionally, ToxPi operates on a relative basis within the dataset, which can pose challenges when assessing unknown compounds with uncertain potency or clustering attributes. This also prevents the comparison of ToxPi/slice scores across multiple datasets that are fed through ToxPi as it operates relatively to the data inserted as the ToxPi figures operate on a relative basis which can fluctuate and change based on the introduction of additional reporters and compounds. This limits the comparative capacity the figures have across multiple projects.

While ToxPi cannot fully replace existing tools for assessing potency, it provides a substantial enhancement to dose-response analyses in MultiFlow and MicroFlow assays by serving as an additional comparative figure. Moreover, ToxPi is particularly adept at identifying unexpected deviations from expected values within complex multiplexed datasets, underscoring its value as a supplementary tool in toxicological assessments. As ToxPi has been applied to the MultiFlow and MicroFlow assays, any further questioning on the procedures used here are continuing to be further enhanced and more questions surrounding the subject matter will continue to be addressed in the future, hence, the subsequent third chapter continues the exploration of these methodologies across the pharmacology/toxicology landscape by delving into the application and refinement of these tools within the ToxTracker assay. Additionally, it may be beneficial to run an analysis on compounds with similar MoAs and subtle biochemical differences, (i.e. a series of antivirals), to reduce overrepresentation of highly potent compounds and instead evaluate the quantitative sensitivity of ToxPi. In this context, ToxPi adds value by integrating multiple biomarker responses into a comparative framework to enable potency ranking while supporting mechanistic interpretation.

Chapter 3 – The combinational assessment of *in vitro* Toxtracker assay data with the TGx-DDi biomarker for genotoxicity hazard assessment on eighteen chemicals with positive/negative *in vitro* and *in vivo* cytotoxic calls.

3.1 Introduction

3.1.1 – The demand for combined approaches to toxicological for the innovation of new modern hazard assessment pipelines for mechanistic/potency information extraction

The development of new genotoxicity testing pipelines is crucial for advancing safety and hazard assessment tools. Identifying potential genetic hazards, such as those posed by existing or emerging chemicals, pharmaceuticals, and environmental pollutants, is vital, as these agents can potentially increase the risk of mutations, DNA damage, and chromosomal alterations, ultimately threatening human health through genetic disorders, diseases, and cancers. Addressing these concerns will enhance the development of potency assessment tools, improving human safety and survival.

Traditionally, as no single test can detect all forms of genotoxicity, testing strategies have revolved on the combination of both *in vitro* and *in vivo* assays in a battery format to establish a strong set of protocols, which allows for a wide range of genetic endpoints to be detected (Kirkland et al., 2007; Kirkland, Kasper, Muller, Corvi, & Speit, 2008). However, this produced limitations, such as the requirement of multiple highly sensitive *in vitro* assays needing to be performed to increase the number of genetic endpoints detected, which had the effect of reducing specificity across the assays, leading to the increased production of misleading positives for the *in vitro* findings (Kirkland et al., 2007; ICH S2, 2012). The detection of misleading positives is alleviated with the *in vivo* battery system, however, the phasing out of *in vivo* models due to ethical/logistical concerns by animal right advocacy groups and the limited human relevance/mechanistic resolution of this approach means that an alternative to this system that provides both increased mechanistic insight and potency information is highly desirable (ICH S2, 2012).

Recent advancements have led to the development of novel *in vitro* genotoxicity assays that offers a highly sensitive detection system which integrates an increased degree mechanistic understanding of both genetic and toxicological assessments for chemicals. Hence, combinational assay pipelines have developed where multiple *in vitro* assays are employed to

address the shortcomings of individual tests/traditional battery testing systems. These strategies aim to adhere to the 3Rs principle, where current testing approaches are eventually replaced, reducing our reliance on whole-animal models, and prioritising alternative novel genotoxicity testing methods to help develop and distribute new compounds in pharmaceuticals. This will also aid in identifying existing physical, biological, and environmental hazards, ensuring the continued protection of human health and safety (Russel and Burch, 1959; Sewell et al., 2016).

This chapter aims to display the approach involving the combination of the reporter assay ToxTracker assay which identifies cellular responses across mES cell lines and the TGx-DDI biomarker test for a transcriptomic approach using TK6 cell lines which uses an array of gene profiling to classify compounds as DDI/non-DDI (Hendriks et al., 2011; 2016; Li et al., 2019). By applying both assays as an alternative pipeline, both cellular biomarker responses and gene expression changes are monitored to provide a holistic and robust assessment for a compound's induced biological perturbations, supporting weight-of-evidence evaluations of genotoxic liability and MoA assessment**.** This would allow us to adequately and confidently distinguish between genotoxic/non-genotoxic compounds, enhancing our ability to accurately predict adverse outcomes earlier in the testing process, facilitating safer chemical development and the identification of potentially sources of genotoxicity. Furthermore, computational modelling using PROAST/ToxPi would enable more precision for data interpretation and predictive toxicological assessment of chemicals. This chapter aimed to delve into this innovative combination using the study performed by Allemang et al., (2020) as a dataset-basis to highlight the advantages, challenges of this testing strategy system.

3.1.2 – The ToxTracker Assay

As described in Chapter 1, the *in vitro* Toxtracker assay is a high-throughput genotoxic screening assay was developed by Toxys B.V. which draws parallels to the MulfiFlow/MicroFlow assays by monitoring cellular responses associated with DNA damage using genetically engineered mES cell lines fluorescently stained with GFP to analyse 6 genotoxic/non-genotoxic endpoints (Hendriks et al., 2011). Similarly to the MultiFlow assay, these biomarkers are measured for their activation using a MACSQuant® Analyser 10 flow cytometer to quantitatively measure DNA damage events, assisting in recognition of the MoA attributable to each compound tested, bringing a heightened understanding behind chemicals and their abilities to cause damage to genetic material (Hendriks et al., 2016).

Monitoring the activation of these endpoints, it brings a comprehensive overview of compound potency and their ability to induce cellular stress and mechanistic information on chemical biological damage pathways to enable a portrayal of a compound's MoA. The Toxtracker assay shows its value in the ability to collectively assess existing compounds, early drug development and chemical risk assessment purposes by providing mechanistic insight into cellular stress responses, particularly oxidative stress and, to a lesser extent, aneugenicity. The biomarkers utilised include Bscl2 and Rtkn, which reflects the downstream activation of ATR and ATM induced DNA damage pathways respectively associated with mutagenic DNA lesions and double-strand DNA breaks, as well as Btg2 for p53 activation (DNA damage, cytotoxicity, and apoptosis), and ddit3 for protein damage related to the unfolded/misfolded protein response, while Srxn1 and Blvrb are markers of ROS production, reflecting oxidative stress (Hendriks et al., 2016). Given its clear strengths, this chapter aims to utilise data generated by Procter and Gamble in Allemang et al., (2020) to combine both the ToxTracker assay and TGx-DDI biomarker, to analyse 18 test chemicals with positive/negative *in vitro* and *in vivo* cytotoxic calls and aim to evaluate how integrating these assays impacts the potency ranking and clustering generated by ToxPi (Allemang et al., 2020; Li et al., 2019).

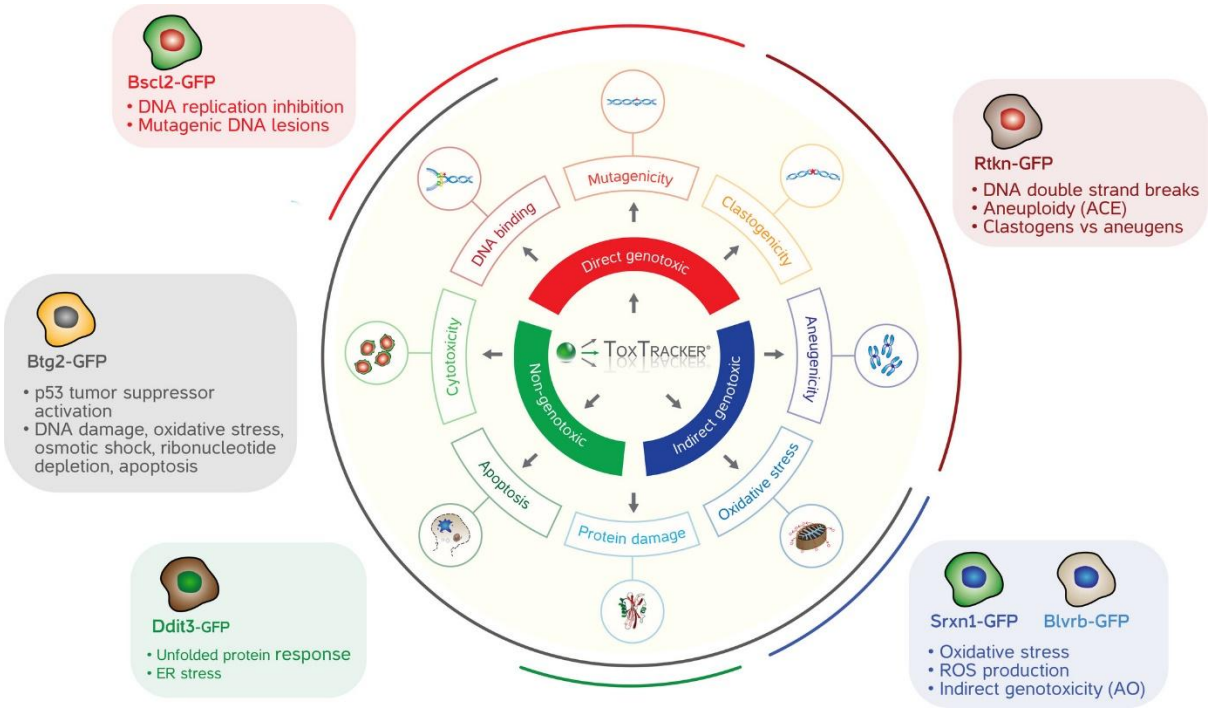


Figure 3.1.2.1 – A representation of the genotoxic/non-genotoxic endpoints covered in the Toxys ToxTracker Assay indicating each of the six GFP reporters with the associated MoA (Hendriks et al., 2024).

3.1.3 – TGx-DDI Transcriptomic Biomarker Test

Measuring carcinogenicity remains a complex subject matter as chemicals act through both genotoxic or non-genotoxic mechanisms that may ultimately genomic damage to humans, however, non-genotoxic carcinogens typically do so indirectly through processes such as oxidative stress, chronic inflammation, mitogenic stimulation, or disruption of chromosomal segregation rather than through direct DNA interaction. A high throughput testing system that can elicit specific gene biomarker information would be highly beneficial in extracting AOP data to more readily identify a highly detailed understanding behind compound MoA. As described in chapter 1, the TGx-DDI biomarker test is a transcriptomic test that provides an in-depth genetic assessment of compounds to analyse their ability to induce DNA damage (labelled DDI) using DNA damage biomarkers and signalling pathways involved in cell cycle control, DNA damage repair, and cellular apoptosis to identity this. The development and validity of this transcriptomic test performed by the HESI GTTC group shows that TGx-DDI fulfils the eSTAR standards and the three-pronged method following the nearest shrunken centroid, PCA and 2D clustering located 64 DDI endpoints that consistently identified DDI/non-DDI compounds with 90% probability. To predict whether compounds classification, a variety of gene expression technologies in TK6 cells are implemented (i.e. DNA Microarrays, qRT-PCR. High throughput nanostring nCounter platform and RNA-seq) to obtain fold change figures and comparative analysis is performed where the 64-gene profiles of compounds with known DDI/non-DDI compounds are compared to the chemical tested to help with classification (Buick et al., 2021). As described earlier, DDI identification is a crucial aspect in chemical assessment for human health protection, environmental safeguarding and cancer prevention from new/existing compounds and safety measures surrounding these compounds are easier drawn and ensuring regulatory compliance during drug development and hazard identification. Transcriptional profiles capture bidirectional gene expression changes, whereas reporter assays quantify inducible pathway activation, providing a constrained but more interpretable measure of cell perturbation. However, transcriptional changes suffer from interpretative limitations as gene expression reflects activation of cell stress and DNA damage response pathways instead of the DNA lesions themselves.

The original study by Allemang et al., (2020) analysed 18 compounds with varying MoA, focusing on both *in vitro* and *in vivo* DDI calls. Although aneugens were undetectable, due to their non-DNA reactive mechanisms, vinblastine and colchicine were still included in the

analysis. The combination of the ToxTracker assay and TGx-DDI test enables the detection of complementary cellular stress responses and transcriptional perturbations associated with DNA damage, oxidative stress, protein damage, cell stress, and drug-induced liver injury (DDI). Simultaneous assessment of these biomarkers aims to provide additional insights into the compounds' MoA, which can influence potency ranking and clustering using PROAST and ToxPi. To accurately characterise and quantify the different endpoints while maintaining standardisation, unique methodologies were developed. These approaches are described in the next section to outline the analysis of both ToxTracker and TGx-DDI endpoints. The heatmap in figure 3.1.3.1 shows the TGx-DDI 64 Affymetrix gene panel, illustrating the transcriptional patterns described earlier to classify compounds as DDI or non-DDI.

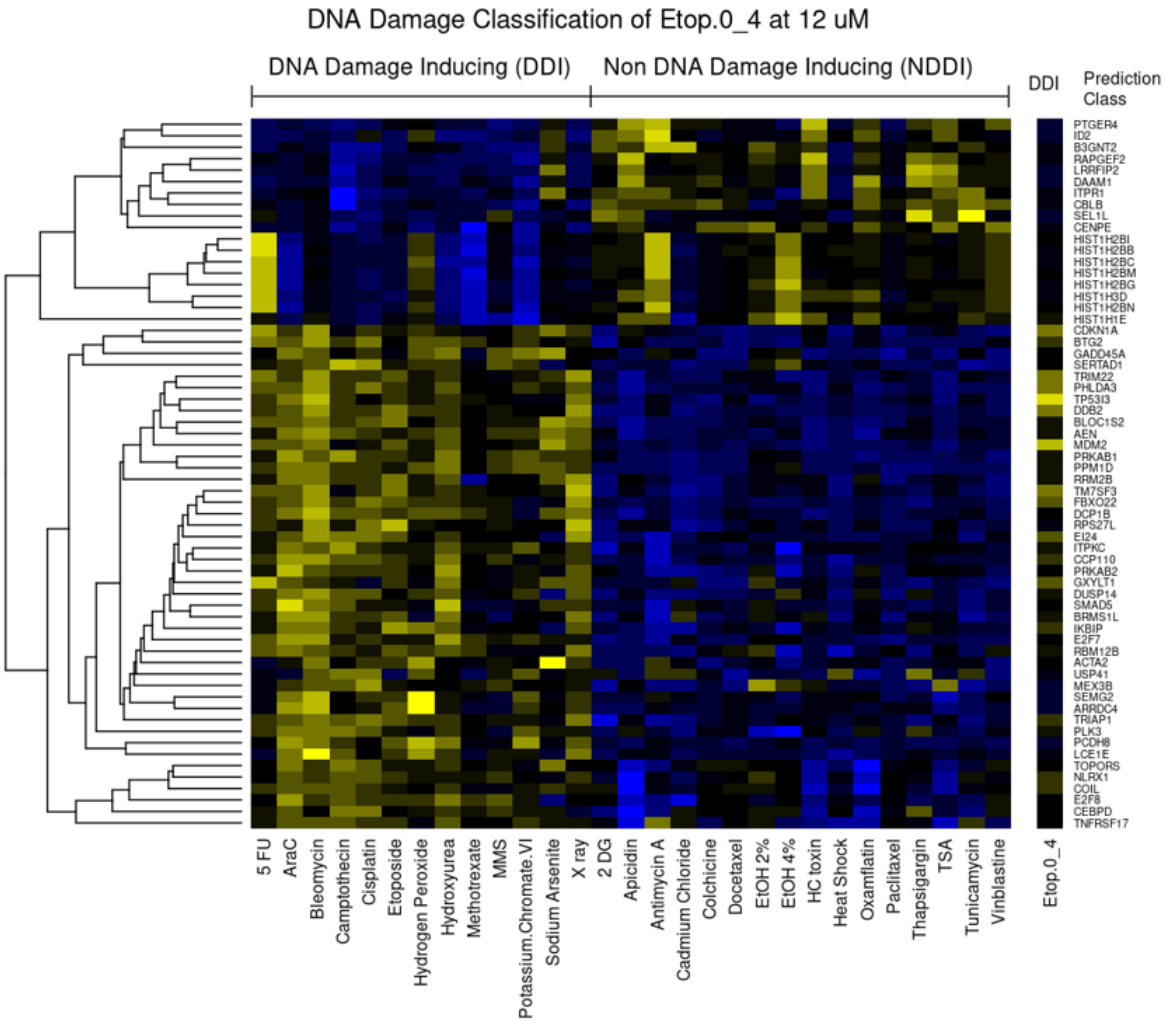


Figure 3.1.3.1 – Heatmap of the TGx-DDI biomarker across 64 genes after TK6 cells were exposed to 12 µM etoposide after 4 hours and analysed using Affymetrix whole genome analysis (Allemang et al., 2020). Blue indicates a fold change decrease relative to the control and yellow indicates an increase in fold change. The Etop.0_4 band (representing etoposide) on the right of the heatmap was compared to the fold change profiles 13 known DDI compounds and 16 non-DDI reference compounds to predict whether etoposide at a concentration of 12uM is classified as DDI or non-DDI, where etoposide was classified as DDI.

3.1.4 – Gene expression technologies in the TGx-DDI assay

Gene expression technologies (including DNA microarrays, qRT-PCR. high throughput NanoString nCounter platform and RNA-seq) quantify mRNA transcript abundance for thousands of genes in parallel and these are applied in the TGx-DDI assay to obtain fold change signatures that enables the identification of DDI compounds (Markowitz et al., 2003; Li et al., 2019). Quantitative gene expression information from TK6 cells is essential and opens opportunities for hazard identification, chemical screening, MoA development, weight-of-evidence analysis and potency assessment to validate the TGx-DDI biomarker results (Li et al., 2017). Although transcriptional processes occur downstream of the molecular event exhibited temporal lag, they provided an integrated measure of cell stress pathway activation. TK6 cells were treated with chemicals of known MoAs using concentration setting experiments (dose-response studies), where microarrays measured transcriptomic disturbances mediated by compounds alongside bioinformatic analysis to identify the subset of genes that responded to each genotoxic/non-genotoxic agent, highlighting the genes involved in the DNA damage response (Li et al., 2017). Once compared, the biomarkers with significantly perturbated findings to genotoxic stress and no changes to non-genotoxic agents were identified and collated into the TGx-DDI biomarker as signature genes used to predict DDI/non-DDI compounds and once these 64 genes were established, they were validated through further microarray analyses (Li et al., 2015; 2017; 2019; Yauk et al., 2016). High resolution and modern transcriptome profiling is performed with RNA-seq, a high throughput sequencing system that provides detailed insight on transcriptomic changes caused by drug exposure, to explore the performance of TGx-DDI in TK6/HepaRG cells (Buick et al., 2020). RNA-seq brings refinement to the TGx-DDI biomarker by identifying genotoxicity related genes/genetic events that could contribute towards genotoxicity using RNA extraction (Buick et al., 2020). Once the gene signatures were identified, the cell viability assays are performed and the RNA is isolated, the qRT-PCR phenotypic protocol is applied to precisely quantify RNA transcripts to measure individual gene expression, validate the key genes, and further confirmed, whether any of the selected genes were involved in cell response to genotoxic agents using gene validation steps. ATF3 (stress agents in TK6), *GADD45A* and *CDKN1A* (p53-regulated response) were used as housekeeping genes with *GAPDH* and *18sRNA* being the standard internal control genes (Li et al., 2015). While quantitative transcriptional responses can vary across cell systems due to differences in baseline gene regulation/stress pathway sensitivity, TK6 cells provides a well characterised and regulatory accepted model for genotoxicity assessment. The high throughput NanoString nCounter offers an

amplification-free direct platform for the TGx-DDI biomarker to measure gene expression once the gene-set is determined, without the need for sequencing/amplification (Geiss et al., 2008; Li et al., 2017). Together, these allowed for the collective 64 gene expression signatures to be identified, validated and applied to classify whether drugs are identified as DDI/non-DDI.

3.1.5 – Use of different cell lines for *in vitro* genotoxicity studies

Different cell line offers unique characteristics and provides additional information which each provide advantages based on their tissue origin, metabolic behaviour, reactivity, genetic stability and DNA repair mechanisms. For *in vitro* genotoxicity assays, TK6 cells has lymphoblastoid origin in humans (derived from WI-L2 cells cultured from a male patient with hereditary spherocytosis) and are often used as the standard cell line for regulatory hazard/safety assessment. Their p53-proficiency, high DNA damage sensitivity, stable karyotypes, well-understood genetic background regarding mutations, and a simplified culturing procedure with versatile applicability makes it a cell line with reliable detection of genotoxic events with low background mutation frequency (Li et al., 2022). Additionally, their acceptance within international frameworks (i.e. OECD, 2016 guidelines) reinforces their mechanistic suitability for genotoxicity testing, while their ease of culturing enables wide applicability across multiple assay formats. Despite this, TK6 cells have limited intrinsic metabolic competence, often requiring S9 to enable metabolite versions for compounds to be assessed, as TK6 has lower levels of phase I enzymes and incompletely characterised phase II enzyme profile. The incorporation of metabolic activation systems is important when assessing compounds whose genotoxicity is metabolite-dependent (Li et al., 2022). One such cell line, used in the ToxTracker assay is mES cells. These are derived from mouse embryonic stem cells and are highly appropriate for the assay due to their stable genome, low background mutation frequency, highly conserved cell stress response pathways and the inherent nature of cancer cells theorised to originate from stem cells. Their rapid proliferative rate and sensitivity to DNA damage pathways further supports reliable reporter responses (Hendriks et al., 2011). However, like many *in vitro* systems, they lack the intrinsic metabolic capacity to assess metabolites, so S9 incorporation is required to adequately evaluate compounds in their metabolically activated forms. Furthermore, the cell lines used in the TGx-DDI component are TK6/HepaRG cell line with wild-type p53; the latter deriving from hepatocellular carcinoma from a human female donor. This would provide very strong primary hepatocyte-like properties whilst expressing a relevant number of phase I and phase

II metabolic enzymes, transporters and nuclear receptors and differentiate into hepatocyte-like cells (Buick et al., 2021). HepaRG cells therefore have increased sensitivity to the detection of genotoxic compounds, cell culture stability, and metabolic competence for a multitude of genotoxic assays, making it an incredibly versatile cell line option (Buick et al., 2021). However, due to their ability to bioactivate pro-mutagens and their high sensitivity, even weak clastogens can elicit strong responses in HepaRG cells, which may overestimate potency and complicate clastogenicity detection. Despite this, TK6 cells continue to be widely used and serve as the default cell line for this assay in the context of this investigation. Using a combination of cell lines can prove to be advantageous to maximise their capability of analysing a variety of mechanisms, reactivities, and genetic events attributed to genotoxic compounds, hence displaying flexibility with cell line selection and using rational justifications that adhere to regulatory constraints within cell lines would maximise information extraction across all assays/tests covered in this chapter.

3.1.6 –The combinational ToxTracker and TGx-DDI testing assay

This chapter attempts to highlight the effects and benefits of combining metrics produced from the ToxTracker and TGx-DDI assays to produce a minimal set of *in vitro* analytical figures that compares compound potency ranking, endpoint activation, hierarchical AGNES clustering, and K-means divisive/PCA clustering by PROAST/ToxPi, producing a system that minimises the requirements of *in vivo* follow-up testing, making it an effective tool in potentially reducing unnecessary animal testing. Comparative data on 18 chemicals tested in Allemang et al., (2020) were analysed, presenting both positive/negative *in vivo* and *in vitro* results linked to their cytotoxicity. A combinational MN test incorporating the ToxTracker, TGx-DDI assay, and whole-genome clustering was employed to investigate the DNA damage-response and chromosomal-level processes underlying each compound's effects. The analysis focused on differentiating between direct and indirect MoAs, examining how the chemicals clustered based on shared properties, and identifying signatures indicative of aneugenic or clastogenic mechanisms, with mutagenic potential inferred from DNA damage-related responses rather than directly measured. This integrated approach sought to uncover associations with cancer-related stress responses through a synthesis of multiple analytical metrics.

ToxTracker and TGx-DDI testing assays both offer substantial benefits as described in earlier sections and as an individual test, easy fulfil their desired purposes with robust and refined findings. To produce the most information-dense and effective figures for highlighting

specific details behind a compound's genotoxicity, an analytical approach would require the employment of both tests in tandem. This would generate efficient figures that allow for a holistic evaluation of a compound's overall genotoxicity, including potency, endpoint activation, and chemical grouping or clustering. These figures will provide a comprehensive breakdown of compound mechanistic properties, allowing the complexities surrounding a collective compound MoA and modelling to be uncovered and the improvement of the accuracy of prediction models while minimizing false positive and negative results during data handling/processing. This integrated approach holds potential promise for enhancing the early identification of harmful compounds, reducing reliance on animal models by enabling improved compound prioritisation before *in vivo* testing, and advancing the field of genetic toxicology testing systems.

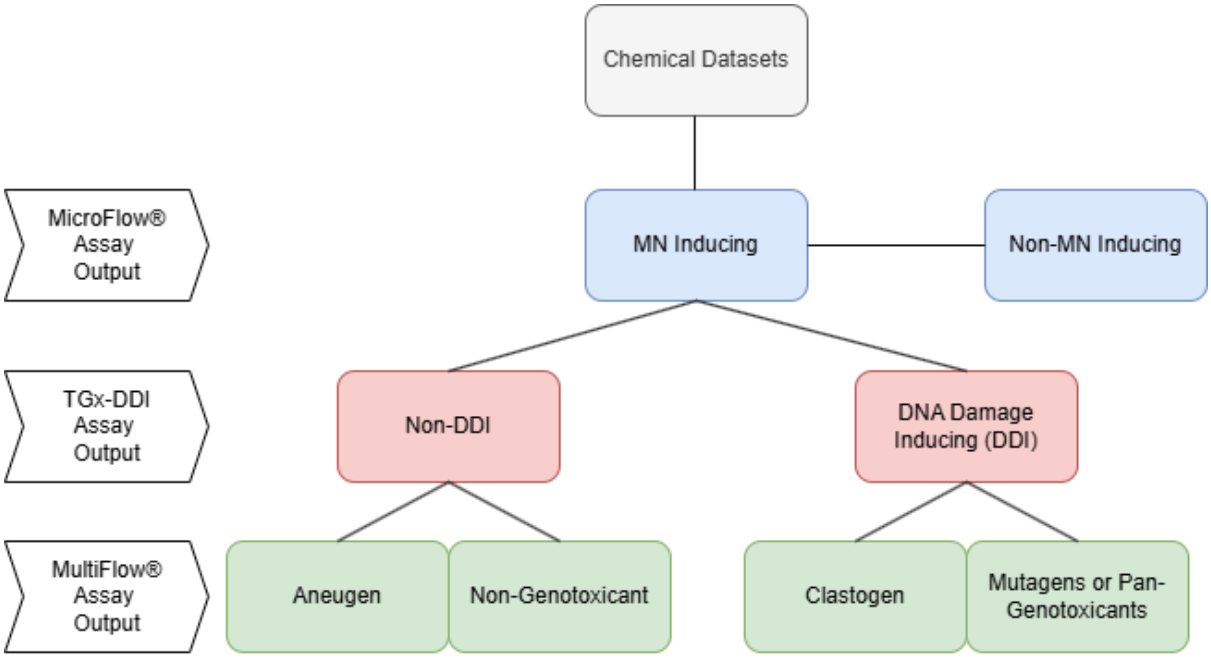


Figure 3.1.6.1 – NAM-based testing strategy framework flowchart depicting the combination of MicroFlow, TGx-DDI and MultiFlow tests to group compounds into their respective MoA classifications –aneugens, clastogens, mutagens and non-genotoxicants (adapted from (Buick et al., 2021)).

3.2 Materials and Methods

This chapter outlines the technologies, cell lines and methodologies applied within the bounds of the both the ToxTracker assay and TGx-DDI biomarker test.

3.2.1 – ToxTracker Dataset Curation

Toxys B.V. (Leiden, Netherlands) performed the ToxTracker assay on mES cell lines outlined in (Hendriks et al., 2011; 2016). The wild-type mES cells (C57/B4418) were cultured in mES knockout medium ((Gibco) containing 10% fetal calf serum, 2 mM glutamax, 1 mM sodium pyruvate, 100 μM β-mercaptopurine and leukaemia inhibitory binding factor) and seeded for 24-hours in gelatine-coated plates in buffalo rat liver (BRL) prior to exposure (Hendriks et al., 2016). The cells are then exposed to each test chemical for 24-hours and for metabolically activated versions of each test chemical, cells were exposed to 1% rat liver S9 extract (in 3 mM potassium chloride, 0.8 mM magnesium chloride, 0.5mM glucose-6-phosphate, and 0.4 mM NADP) for 3-hours prior to seeding in BRL (Hendriks et al., 201; 2016). Before seeding, the solution is washed in phosphate buffered saline (PBS) and cultured in the BRL medium for 24-hours (Hendriks et al., 2011; 2016).

Bacterial artificial chromosome (BAC) recombineering generated GFP reporters using mouse BAC finder to select the bacterial strains with the appropriate biomarker genes. Furthermore, the putative biomarker genes are modified with a C-terminal GFP marker using the BAC modification kit and PCR fragments. Using the KAPA HiFi PCR system, PCR fragments were created with a GFP-IRES-neomycin/kanamycin reporter and 50 nucleotide-long homology sequences to the biomarker gene. BAC strains were transformed with the GFP-IRES-Neo PCR fragment by electroporation and then plated on kanamycin plates. The modified BACs were purified with a Nucleobond kit, and the modification was confirmed by sequencing prior to BRL seeding and 24-hour transfection. GFP induction figures are obtained for GFP-tagged reporter lines Bscl2, Rtkn (DNA damage), Srxn1, Blvrb (Oxidative stress), Btg2 (p53 activation), and Ddit3 (protein damage) were obtained using high throughput flow cytometry (Guava easyCyte 6HT, EMD Millipore) where cells exposed to genotoxic agents in 96 gelatine-coated wells (dissolved in DMSO/ PBS) and diluted in BRL. After 24-hours, cells are washed with PBS, trypsinized and resuspended before proceeding with flow cytometry. Reporter activity was measured using the mean fluorescence intensity of 5000 intact cells where activation of a reporter is considered when a greater or equal to 1.5-fold change (5x Standard deviation) is observed of GFP expression, where noted results,

and further data processing is initiated. Biomarker performance was confirmed through the comparison of each compound to the control compounds set by Toxys for each MoA (Cisplatin for DNA damage, Alfatoxin B1 for p53 activation to metabolic stress, diethyl maleate for oxidative stress, and tunicamycin for protein misfolding) (Allemang et al., 2020).

The top concentrations used was 1 mM or the solubility limit with exception of o-toluidine which had a limit set to 10 mM. Once fold-change values are obtained across all compounds, concentrations where cell cytotoxicity exceeded 75% were not considered for further data processing/analysis for reasons explained in chapter 2.2.5.1 and the raw concentrations-GFP induction figures were placed into PROAST for BMD generation. Below is the list of all compounds assessed, including their MoA and *in vitro/in vivo* positive/negative calls:

Table 3.2.1.1 – Table outlining the compounds assessed in both the ToxTracker and TGx-DDI assay.

Compound	CASRN	MoA	Classification
2-Deoxy-D-glucose	157-17-6	Non-genotoxic	*In vitro* and *in vivo* negative
Amitrol	61-82-5	Non-genotoxic	*In vitro* and *in vivo* negative
Cyclohexanone	108-94-1	Non-genotoxic	*In vitro* and *in vivo* negative
D-Mannitol	69-65-8	Non-genotoxic	*In vitro* and *in vivo* negative
Sodium diclofenac	15307-79-6	Non-genotoxic (oxidative stress at higher doses)	*In vitro* and *in vivo* positive
5-Fluorouracil	51-21-8	Nucleoside analogue	*In vitro* and *in vivo* positive
Camptothecin	7689-03-4	Topoisomerase I inhibitor	*In vitro* and *in vivo* positive
Colchicine	64-86-8	Tubulin Binder	*In vitro* and *in vivo* positive
N-ethyl-N-nitrosourea (ENU)	759-73-9	Alkylating agent	*In vitro* and *in vivo* positive
Etoposide	33419-42-0	Topoisomerase II inhibitor	*In vitro* and *in vivo* positive
Methyl methanesulfonate (MMS)	67-27-3	Alkylating agent/oxidative stress	*In vitro* and *in vivo* positive

o-toluidine	95-53-4	DNA adduct formation	*In vitro* and *in vivo* positive
Vinblastine	143-67-9	Tubulin Binder	*In vitro* and *in vivo* positive
Ethionamide	536-33-4	Unknown MoA	*In vitro* positive, *in vivo* negative
Eugenol	97-53-0	Antioxidant/mixed MoA	*In vitro* positive, *in vivo* negative
Isobutyraldehyde	48-84-2	Oxidative Stress	*In vitro* positive, *in vivo* negative
Quercetin	117-39-5	Topoisomerase II inhibitor/Oxidative stress	*In vitro* positive, *in vivo* negative
Saccharin sodium	128-44-9	Indirect mutagen (irrelevant to humans)	*In vitro* positive, *in vivo* negative

3.2.2 – TGx-DDI dataset curation

The TGx-DDI framework employs a triad of methodologies—gene expression profiling, pathway interrogation, and machine learning techniques—to discern drug-induced injury signals. Each component independently scrutinises the data, and their integrated results offer a comprehensive evaluation of potential DDI hazards. The three distinct interpretations of the data include the nearest shrunken centroid probability method (chemicals classed as DDI when the probability was greater than 90%), PCA analysis using machine learning on the original training set where chemicals are classed as clean/unclean DDI, and 2-dimensional clustering to measure patterns and categorise gene expression data based on response similarities.

Health Canada analysed the Affymetrix data using the TGx-DDI biomarker approach following the procedures outlined in (Li et al., 2015; 2017; 2019). These studies highlight that TK6 cells were cultured in RPMI and 10% foetal bovine serum (FBS) and grown cells from 1-10x10^5 cells/ml (maintenance level) to a density of 4-5x10^5 cells/ml for treatment. A stock solution of each chemical was used to treat cells simultaneously and solvents (H_2O, ethanol, methanol, DMSO, and NaCl) used as controls were selected based on MoA and dose range (determined through dose response studies) as highlighted in Table 1 in the Li et al., (2015) publication and cells were exposed to γ-radiation irradiator at 0.87 Gy/min to a dose of 4 Gy.

RNA extraction/isolation was performed using the Trizol protocol (by Invitrogen - Thermo Fisher Scientific). Once purified, the RNA sample was subjected to spectrometry and bioanalyser analysis to determine sample quality and quantity. For analysed samples, qRT-PCR is performed using an icycler (Bio-Rad) to measure individual gene expression (RNA isolation → extraction → reverse transcription → qPCR) where housekeeping genes (controls that maintain normal basal cell functions) include *GADD45A*, *CDKN1A*, and *ATF3* with the following primer sequences:

- *GADD45A*: Forward 5′-TGCGAGAACGACATCAACAT-3′, Reverse 5′-CGGTCAGCAGGAATGACAGT-3′
- *CDKN1A*: Forward 5′-CAGACCAGCATGACAGATTTCT-3′, Reverse 5′-GGATTAGGGCTTCCTCTTGG-3′.
- *ATF3*: Forward 5′-CAGAGCAAGTGTCAGCATCC-3′, Reverse 5′-GCTGGTCAACGCAGCATTA-3′.

The assay was performed in triplicates and *18sRNA* was used to normalise gene expression levels while *GAPDH* was used for antimitotic agents, histone decyclases. Relative mRNA induction fold changes more than 1.7 (p-value < 0.05) were classified as significant, reflecting a data-driven threshold established during the biomarker validation to distinguish biologically meaningful transcriptional responses from baseline variability (Li et al., 2015). Additionally, iScript one-step RT-PCR kits for probes were used with the probe mixes used deriving from TaqMan gene expression assay (ABI, Foster city, CA).

Gene expression microarray data was labelled with SuperScript indirect cDNA labelling system (Invitrogen) (without an amplification step) and hybridised on Agilent human whole genome 44k oligo microarrays. Overall, these steps allow for the TGx-DDI biomarker to curate a set of genes that are responsive to DNA damage, allowing researchers to evaluate the genotoxicity of substances based on gene expression patterns.

Upon completion of data processing, compound with positive DDI calls had BMD data gathered to prevent false positives where non-DDI compounds can sometimes produce BMDs despite their classification. Furthermore, vinblastine and colchicine measurements were included in the analyses, despite being aneugenic and their MoA remaining undetectable in the TGx-DDI assay. This weight-of-evidence approach in theory, would provide a comprehensive production of BMD figures and models that provide further insight on compound genotoxicity across a diverse range of genes, encapsulated as a singular DDI/non-DDI endpoint.

3.2.3 – Working in PROAST

3.2.3.1 – ToxTracker

For the purposes of the ToxTracker assay, prior to BMD analysis using PROAST, all compounds had responses across all 6 ToxTracker endpoints assessed, and individual dose-response values were obtained using compound concentration and GFP induction figures, using the concentration of 0 μM with a GFP response of 1.000 set as the control values for the investigation. For subsequent concentrations, any concentration where GFP values exceeded 1.000 implied and increase in response from that endpoint at that concentration. Once all compound arrays are collected, any doses/responses where the cytotoxicity value >60% with a relative cell survival rate <40% had their results removed from the dataset to retain relevancy from excessive cell cytotoxicity figures induced by unrelated secondary

toxicity effects and excessive cytotoxicity findings from interfering with the datasets. This lower threshold reflects a preliminary analysis conducted prior to the adoption of the standard >80% cut-off in later validation studies. Additionally, any concentrations response figures of 0 were subsequently removed to prevent issues with detection limits/PROAST errors, further preventing any experimental outliers/errors from influencing the true BMD figure as a response of 0 constitutes the presence of some experimental noise within the detection process. Finally, a minimum of 5 concentrations required use in PROAST for each compound at each of the six endpoints before BMD analysis commences; this was no concern as this requirement was achieved.

In PROAST, the curated ToxTracker data had "exponential model 3 or 5" applied where concentration and GFP induction were selected as independent variable and dependent/ response values respectively. Covariates selected was the S9± conditions and a CES of +0.5 over control was set, producing 90% BMD CIs from 200 bootstraps (referenced in section 2.2.5.2). Log10-transformed BMD plots were created for dose and response post-curve evaluation, a single dose-response curve for ToxTracker for each compound at each individual endpoint; these responses were used for further ToxPi analysis.

3.2.3.2 – TGx-DDI

BMD analysis was performed using PROAST once data produced from the TGx-DDI biomarker was curated, and the responses across all 64 individual genes were gathered and aligned with their appropriate concentrations from the original dataset. Firstly, BMD values were only run for DDI compounds as to reduce any false positives where non-DDI compounds could numerically produce a dose-response curve leading to the calculation of misleading BMD CIs. Each chemical had 64 gene fold change figures per singular concentration and required normalisation as the data contained both positive and negative numerical values, which wouldn't be appropriate for further BMD analysis for PROAST/ToxPi. Therefore, normalisation involved adding a fixed numerical value to all fold change figures by adding 1 to all values, thereby eliminating negative values for the dataset and retaining numerical integrity across all values, eliminating skewing. Furthermore, a control row was added for all DDI compounds fold change is set at 1.0 at a concentration of 0 once the previously stated transformations were applied, meaning compounds with GFP values above 1.0 implies increase in fold change at that gene, correlating with an increased endpoint response, as only elevated gene expression was considered here. The datasets also

required no minimum number of concentrations as the results themselves are very data-intensive, allowing for PROAST to produce stronger mapping of the dose response curves for each compound when the BMD is determined, especially during the bootstrapping process.

In PROAST, an "exponential model 3 or 5 from various model families" was selected, with concentration/dose set as the dose and fold-change values set as the response. Covariates directly influencing the dependent variable were chosen based on their availability, which in this instance is the genes within the TGx-DDI assay. The CES was set at +0.5 above the control, and 200 bootstraps were performed to generate 90% BMD confidence intervals, as described in section 2.2.5.2. This process produced BMD graphs with log10(dose) and log10(response) axes. After assessing the curve, a single TGx-DDI dose-response curve was generated for each compound, summarizing data from all 64 genes. This TGx-DDI biomarker, representing a combined endpoint, was then used in the ToxPi evaluation.

3.2.4 – ToxPi settings and calibrations

Before processing data through ToxPi, the original BMD CI data is firstly gathered as a CSV file and 0.0001 – 10000 μM limits to each BMD are imposed as described in chapter 2.2.7 for skewing purposes. Once both ToxTracker and TGx-DDI BMD CI data is curated, the endpoints selected are as follows: Bscl2, Rtkn, Btg2, Ddit3, Srxn1, Blvrb, and TGx-DDI. The $-\log_{10}X$ transformation is applied and weighting remained identically set at 7.1% across all endpoints highlighted with exception to the TGx-DDI biomarker which had 14.3% (2-fold). This was applied on the basis that it is an entire assay summarised as a single endpoint, therefore holds higher significance compared to individual endpoints in the ToxTracker assay. As no officially prescribed weighting system is established, the numerical selection of weightings remains flexible. The numerical value applied here is based on a significance ranking. Therefore, the TGx-DDI endpoint has been weighted twice as much as the other ToxTracker endpoints. This approach is like the one used in the study by (Fortin et al., 2023). Once set, the endpoints are colour coordinated according to the ToxPi profile key in Figure 3.3.2.2, the -log10(x) function is applied, and the ToxPi results had their respective bootstrap CIs produced for the hierarchical clustering component, the complete clustering algorithm is selected.

3.3 Results

This chapter will outline the figures produced in this chapter from the gathering of BMD figures from the combination of both ToxTracker and TGx-DDI assays to build visuals and profiles that extract and provides greater depth driven information on compound toxicity, endpoint activation and grouping.

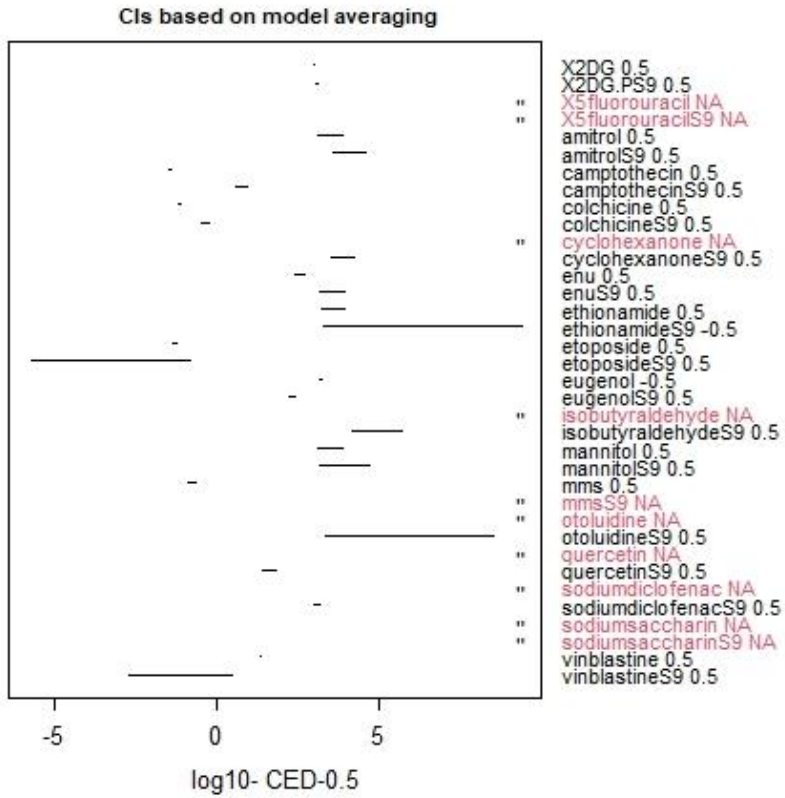


Figure 3.3.1.1 – 90% BMD CIs produced and plotted by PROAST for the Blvrb endpoint, a biomarker for oxidative stress (ROS), across all 18 compounds assessed in the ToxTracker assay, where exponential models 3 and 5 were selected, the concentration was set as the independent variable and the GFP induction was set as the response with the covariate being the S9 conditions (S9+ / S9-). The CES was set to ±0.5 (CED-0.5) with 200 bootstrap runs to produce the BMDs, and graphical plotting of the CIs was formatted to Log_{10}(BMD) on the x-axis with compound names on the y-axis. Compounds that did not return a BMD CI have a "NA" by the compound name, are highlighted in red, and have the '' symbols in the BMD CI response column. This included 5 fluorouracil (S9±), cyclohexanone (S9-), isobutyraldehyde (S9-), methyl methanesulfonate (MMS) (S9+), o-toluidine (S9-), quercetin (S9-), sodium diclofenac (S9-), and sodium saccharin (S9±).

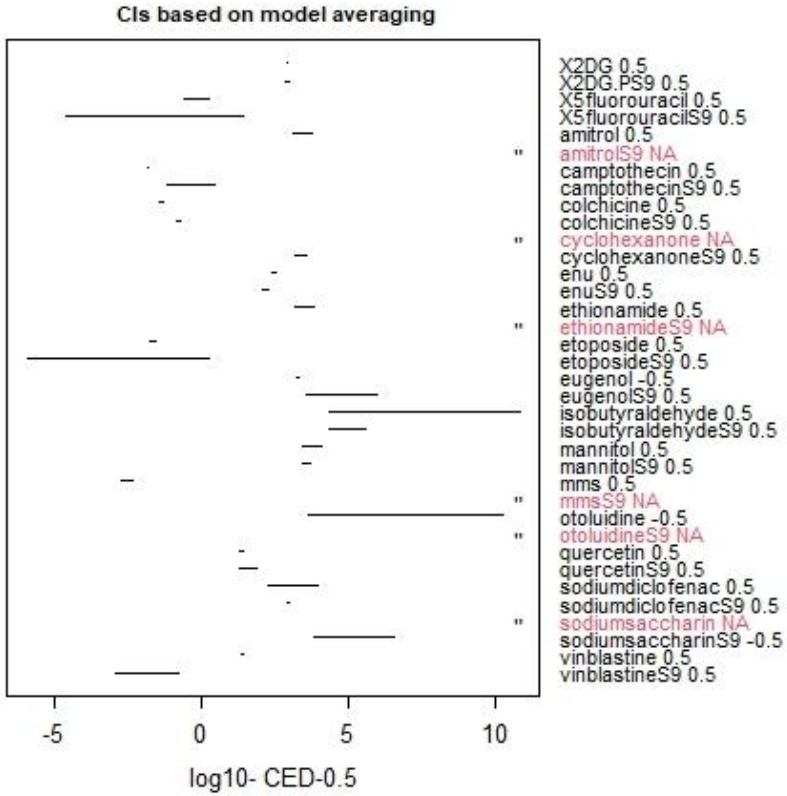


Figure 3.3.1.2 – 90% BMD CIs produced and plotted by PROAST for the Bscl2 endpoint, a reporter of replication stress and ATR mediated DNA damage signalling, across all 18 compounds assessed in the ToxTracker assay, where exponential models 3 and 5 were selected, the concentration was set as the independent variable and the GFP induction was set as the response with the covariate being the S9 conditions (S9+ / S9-). The CES was set to ±0.5 (CED-0.5) with 200 bootstrap runs to produce the BMDs, and graphical plotting of the CIs was formatted to Log_{10}(BMD) on the x-axis with compound names on the y-axis. Compounds that did not return a BMD CI have a "NA" by the compound name, are highlighted in red, and have the '' symbols in the BMD CI response column. This includes amitrol (S9+), cyclohexanone (S9-), ethionamide (S9+), eugenol (S9-), mms (S9+), o-toluidine (S9+), and sodium saccharin (S9-).

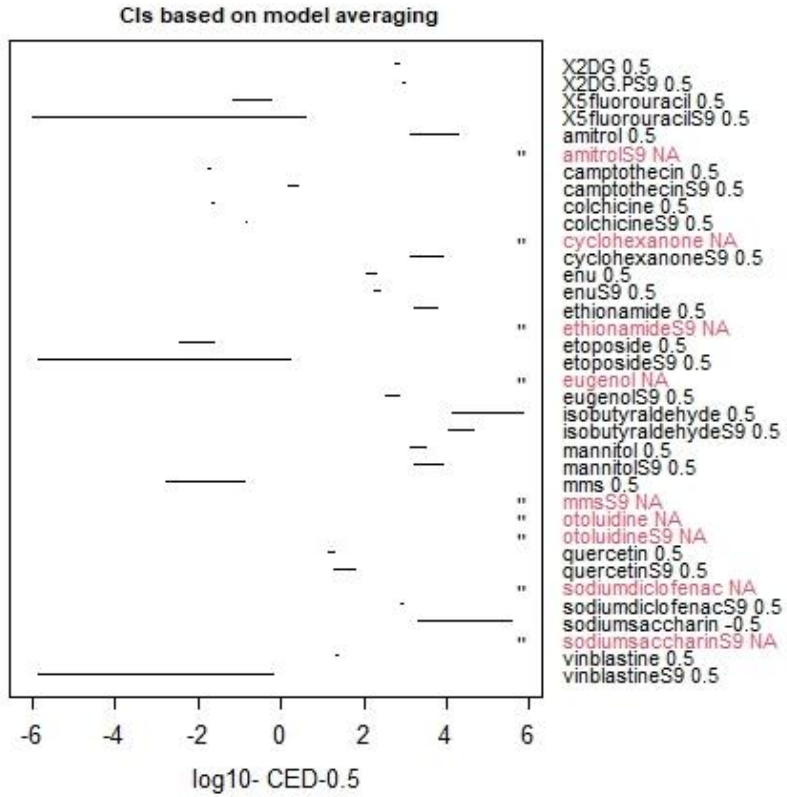


Figure 3.3.1.3 – 90% BMD CIs produced and plotted by PROAST for the Btg2 endpoint, a reporter for p53-mediated cell stress response and cell-cycle perturbation, across all 18 compounds assessed in the ToxTracker assay, where exponential models 3 and 5 were selected, the concentration was set as the independent variable and the GFP induction was set as the response with the covariate being the S9 conditions (S9+ / S9-). The CES was set to ±0.5 (CED-0.5) with 200 bootstrap runs to produce the BMDs, and graphical plotting of the CIs was formatted to Log_{10}(BMD) on the x-axis with compound names on the y-axis. Compounds that did not return a BMD CI have a "NA" by the compound name, are highlighted in red, and have the '' symbols in the BMD CI response column. This includes amitrol (S9+), cyclohexanone (S9-), ethionamide (S9+), eugenol (S9-), mms (S9+), o-toluidine (S9±), sodium diclofenac (S9-), and sodium saccharin (S9+).

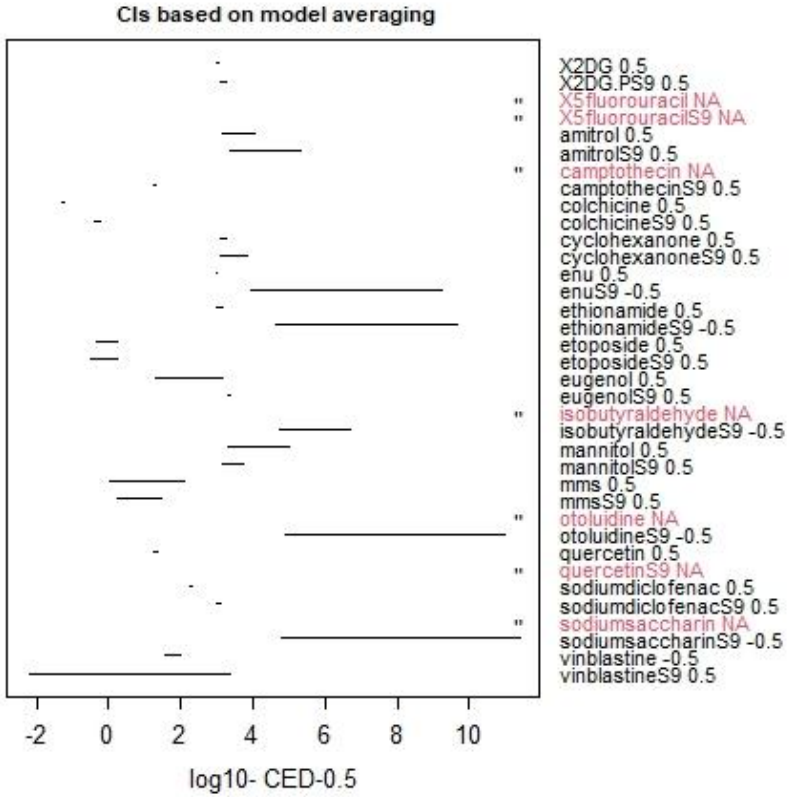


Figure 3.3.1.4 – 90% BMD CIs produced and plotted by PROAST for the Ddit3 endpoint, a biomarker for protein stress response, across all 18 compounds assessed in the ToxTracker assay, where exponential models 3 and 5 were selected, the concentration was set as the independent variable and the GFP induction was set as the response with the covariate being the S9 conditions (S9+ / S9-). The CES was set to ±0.5 (CED-0.5) with 200 bootstrap runs to produce the BMDs, and graphical plotting of the CIs was formatted to Log_{10}(BMD) on the x-axis with compound names on the y-axis. Compounds that did not return a BMD CI have a "NA" by the compound name, are highlighted in red, and have the '' symbols in the BMD CI response column. This includes 5 fluorouracil (S9±), camptothecin (S9-), eugenol (S9+), isobutyraldehyde (S9-), o-toluidine (S9-), quercetin (S9+), and sodium saccharin (S9-).

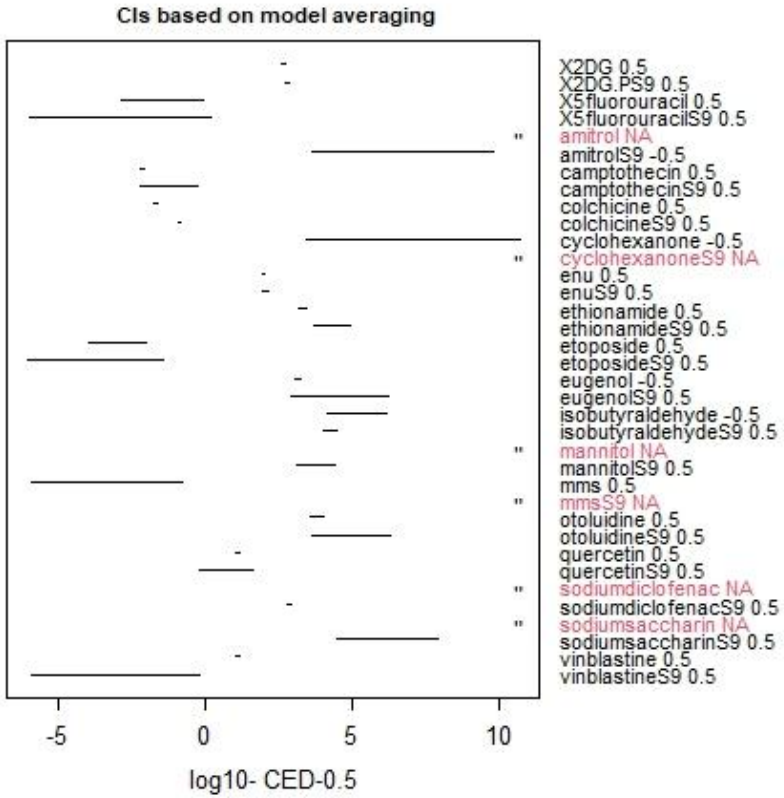


Figure 3.3.1.5 – 90% BMD CIs produced and plotted by PROAST for the Rtkn, a reporter of ATM-mediated DNA damage signalling associated with double-strand break responses, endpoint across all 18 compounds assessed in the ToxTracker assay, where exponential models 3 and 5 were selected, the concentration was set as the independent variable and the GFP induction was set as the response with the covariate being the S9 conditions (S9+ / S9-). The CES was set to ±0.5 (CED-0.5) with 200 bootstrap runs to produce the BMDs, and graphical plotting of the CIs was formatted to Log_{10}(BMD) on the x-axis with compound names on the y-axis. Compounds that did not return a BMD CI have a "NA" by the compound name, are highlighted in red, and have the '' symbols in the BMD CI response column. This includes amitrol (S9-), cyclohexanone (S9+), mannitol (S9-), mms (S9+), sodium diclofenac (S9-), and sodium saccharin (S9-).

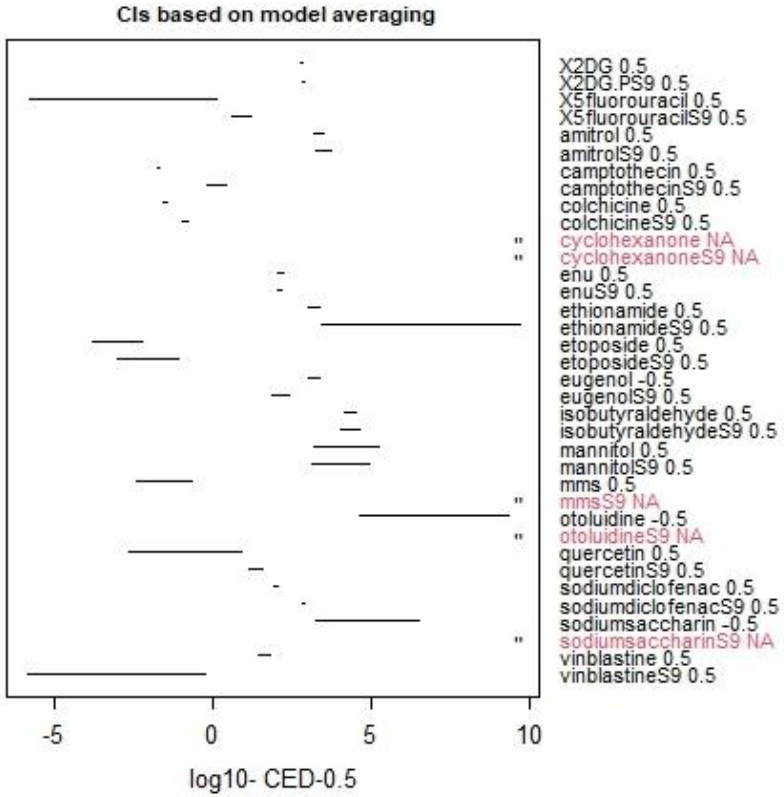


Figure 3.3.1.6 – 90% BMD CIs produced and plotted by PROAST for the Srxn1 endpoint, a biomarker for oxidative stress (ROS), across all 18 compounds assessed in the ToxTracker assay, where exponential models 3 and 5 were selected, the concentration was set as the independent variable and the GFP induction, was set as the response with the covariate being the S9 conditions (S9+ / S9-) and compounds reflecting "NA" outcomes when no dose-response is produced. The CES was set to ±0.5 (CED-0.5) with 200 bootstrap runs to produce the BMDs, and graphical plotting of the CIs was formatted to Log_{10}(BMD) on the x-axis with compound names on the y-axis Compounds that did not return a BMD CI have a "NA" by the compound name, are highlighted in red, and have the '' symbols in the BMD CI response column. This includes cyclohexanone (S9±), mms (S9+), o-toluidine (S9+), and sodium saccharin (S9+).

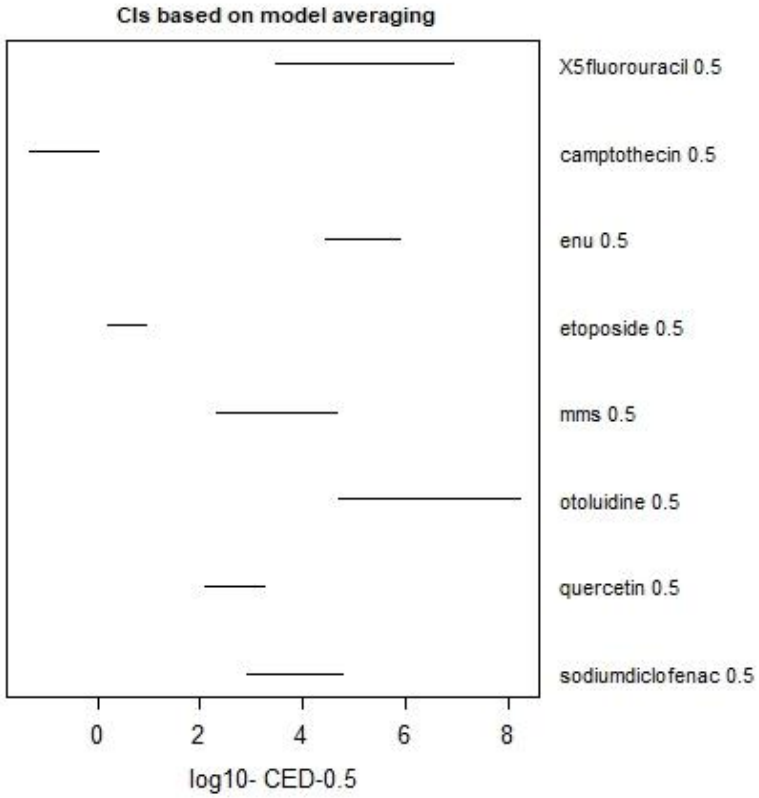


Figure 3.3.1.7 – 90% BMD CIs produced and plotted by PROAST for the TGx-DDI assay across all 18 compounds assessed, where exponential models 3 and 5 were selected, the concentration was set as the independent variable and the fold change was set as the response with the covariate being the S9 conditions (S9+ / S9-). The CES was set to ±0.5 (CED-0.5) with 200 bootstrap runs to produce the BMDs, and graphical plotting of the CIs was formatted to Log_{10}(BMD) on the x-axis with compound names on the y-axis. 8/18 compounds had positive calls for and identified as DDI and, vinblastine/colchicine were classified as aneugens, making them undetectable and irrelevant for the TGx-DDI biomarker.

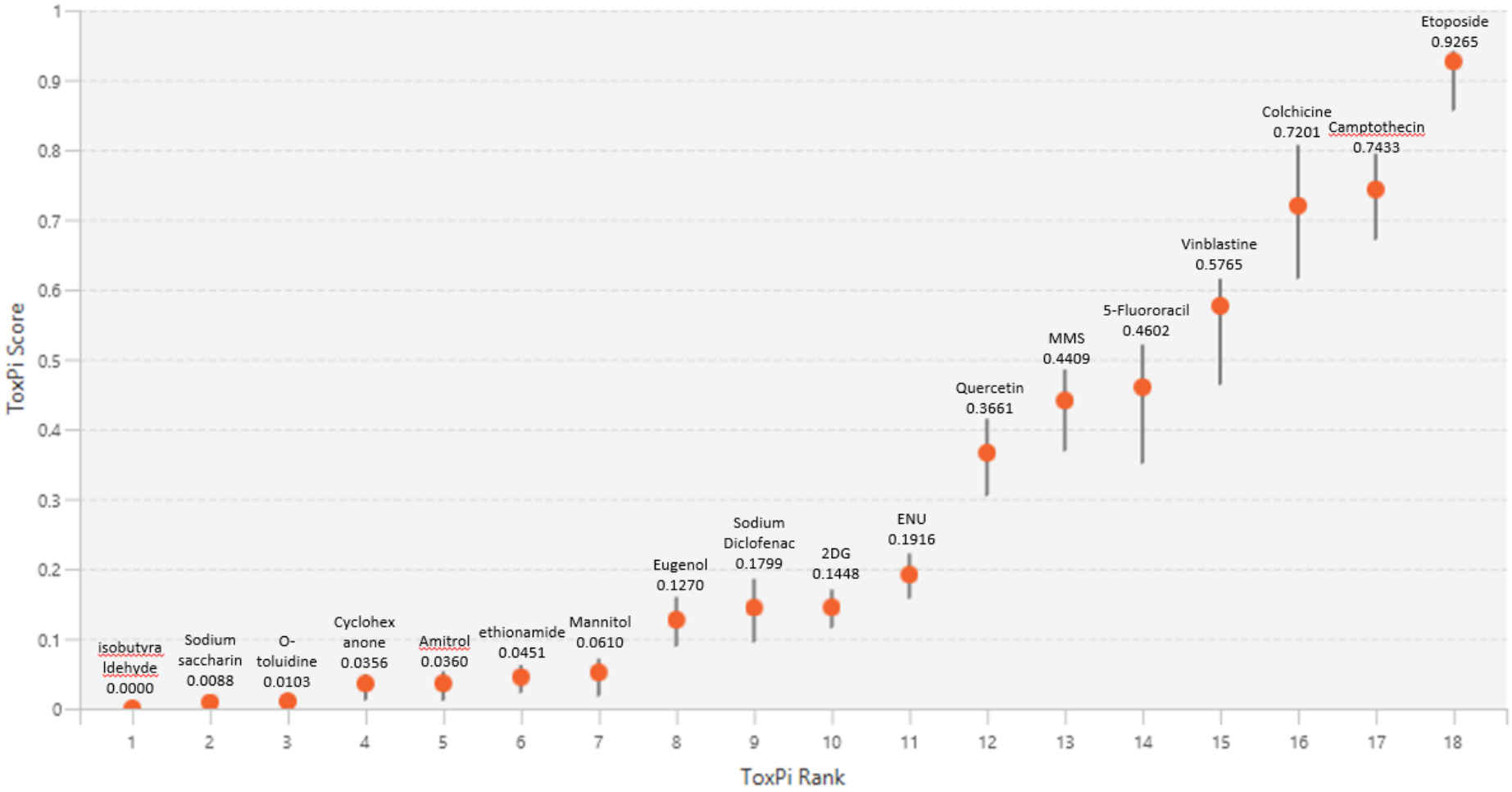


Figure 3.3.2.1 – Graph illustrating the ToxPi ranking derived from BMD data generated in Figures Figure 3.3.1.1 – 3.3.1.7, enabling the comparison of the 18 test compounds (1 = lowest potency and 18 = highest potency). ToxPi scores ranging between 0-1 where the higher score translates to higher genotoxic potency and the error bars represent the CI of each compound.

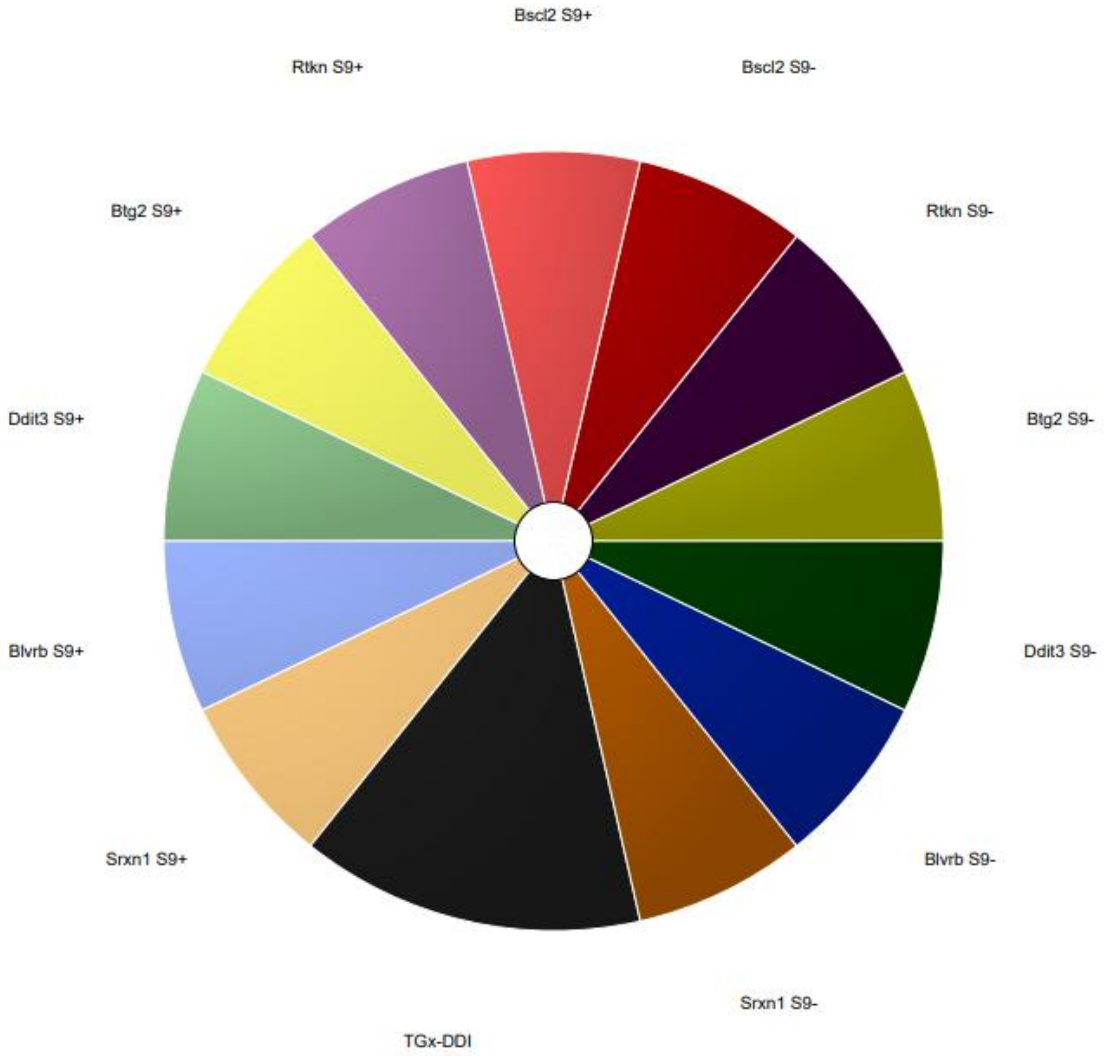


Figure 3.3.2.2 – ToxPi profile key illustrating each slice with their respective endpoints for each endpoint in the Toxtracker and TGx-DDI Assays including ATM/ATR mediated DNA damage (Bscl2 and Rtkn respectively), p53 mediated cell stress response (Btg2), oxidative stress (Blvrb and Srxn1), protein damage (Ddit3), and DDI (TGx-DDI) in the when metabolically activated (S9+) and inactivated (S9-).

etoposide

Score = 0.9265 [0.8582, 0.9391]
Rank = 18 [18, 18] (out of 18)

Btg2 S9- 1.0000 [0.9140, 1.0000]
Rtkn S9- 1.0000 [0.9773, 1.0000]
Bscl2 S9- 0.8734 [0.8613, 0.8864]
Bscl2 S9+ 0.9957 [0.7828, 1.0000]
Rtkn S9+ 1.0000 [1.0000, 1.0000]
Btg2 S9+ 0.9711 [0.7926, 1.0000]
Ddit3 S9+ 0.9605 [0.6941, 0.9605]
Blvrb S9+ 1.0000 [1.0000, 1.0000]
Srxn1 S9+ 0.9716 [0.8750, 1.0000]
TGx-DDI 0.7370 [0.7166, 0.7644]
Srxn1 S9- 1.0000 [1.0000, 1.0000]
Blvrb S9- 0.9717 [0.9661, 0.9782]
Ddit3 S9- 0.7536 [0.7185, 0.7932]

camptothecin

Score = 0.7433 [0.6734, 0.7919]
Rank = 17 [16, 17] (out of 18)

Btg2 S9- 0.9557 [0.8242, 1.0000]
Rtkn S9- 0.8792 [0.7743, 1.0000]
Bscl2 S9- 0.8938 [0.8694, 0.9196]
Bscl2 S9+ 0.7351 [0.6444, 0.7351]
Rtkn S9+ 0.7614 [0.7502, 0.7694]
Btg2 S9+ 0.6077 [0.4752, 0.7461]
Ddit3 S9+ 0.6348 [0.4242, 0.6411]
Blvrb S9+ 0.5068 [0.4282, 0.6377]
Srxn1 S9+ 0.6304 [0.5194, 0.7283]
TGx-DDI 1.0000 [1.0000, 1.0000]
Srxn1 S9- 0.8017 [0.7181, 0.9093]
Blvrb S9- 1.0000 [1.0000, 1.0000]
Ddit3 S9- 0.0000 [0.0000, 0.0000]

colchicine

Score = 0.7201 [0.6174, 0.8042]
Rank = 16 [16, 17] (out of 18)

Btg2 S9- 0.9415 [0.8111, 0.9861]
Rtkn S9- 0.8136 [0.7213, 0.9188]
Bscl2 S9- 0.8263 [0.8108, 0.8427]
Bscl2 S9+ 0.8068 [0.6060, 0.9720]
Rtkn S9+ 0.7150 [0.6151, 0.8563]
Btg2 S9+ 0.7945 [0.6063, 1.0000]
Ddit3 S9+ 1.0000 [0.6747, 1.0000]
Blvrb S9+ 0.6766 [0.5514, 0.8854]
Srxn1 S9+ 0.7949 [0.6184, 0.9784]
TGx-DDI 0.0000 [0.0000, 0.0000]
Srxn1 S9- 0.7721 [0.6922, 0.8749]
Blvrb S9- 0.9398 [0.9356, 0.9434]
Ddit3 S9- 1.0000 [1.0000, 1.0000]

vinblastine

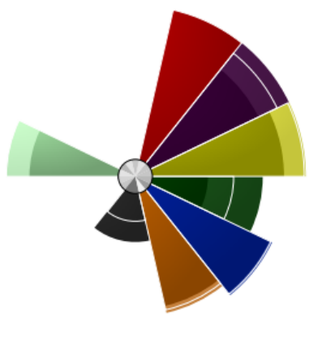
Score = 0.5765 [0.4654, 0.6128]
Rank = 15 [15, 15] (out of 18)

Btg2 S9- 0.4418 [0.3808, 0.4626]
Rtkn S9- 0.4202 [0.3792, 0.4656]
Bscl2 S9- 0.4026 [0.3941, 0.4116]
Bscl2 S9+ 1.0000 [0.8750, 1.0000]
Rtkn S9+ 0.8915 [0.7380, 1.0000]
Btg2 S9+ 1.0000 [0.8656, 1.0000]
Ddit3 S9+ 0.8465 [0.1875, 1.0000]
Blvrb S9+ 0.8991 [0.7307, 1.0000]
Srxn1 S9+ 1.0000 [0.8659, 1.0000]
TGx-DDI 0.0000 [0.0000, 0.0000]
Srxn1 S9- 0.3363 [0.3172, 0.3608]
Blvrb S9- 0.4455 [0.3988, 0.4858]
Ddit3 S9- 0.3880 [0.3830, 0.3925]

5fluorouracil

Score = 0.4602 [0.3527, 0.5184]
Rank = 14 [13, 14] (out of 18)

Btg2 S9- 0.7923 [0.7454, 0.7923]
Rtkn S9- 0.8643 [0.6628, 1.0000]
Bscl2 S9- 0.6418 [0.5924, 0.6883]
Bscl2 S9+ 0.8925 [0.5302, 1.0000]
Rtkn S9+ 0.8621 [0.6672, 1.0000]
Btg2 S9+ 0.9324 [0.6950, 1.0000]
Ddit3 S9+ 0.0000 [0.0000, 0.0000]
Blvrb S9+ 0.0000 [0.0000, 0.0000]
Srxn1 S9+ 0.5072 [0.4254, 0.5736]
TGx-DDI 0.0584 [0.0000, 0.1020]
Srxn1 S9- 0.8335 [0.6194, 1.0000]
Blvrb S9- 0.0000 [0.0000, 0.0000]
Ddit3 S9- 0.0000 [0.0000, 0.0000]



mms

Score = 0.4409 [0.3706, 0.4827]
Rank = 13 [13, 14] (out of 18)

Btg2 S9- 0.9905 [0.8574, 1.0000]
Rtkn S9- 0.9143 [0.7786, 1.0000]
Bscl2 S9- 1.0000 [1.0000, 1.0000]
Bscl2 S9+ 0.0000 [0.0000, 0.0000]
Rtkn S9+ 0.0000 [0.0000, 0.0000]
Btg2 S9+ 0.0000 [0.0000, 0.0000]
Ddit3 S9+ 0.7287 [0.5783, 0.7287]
Blvrb S9+ 0.0000 [0.0000, 0.0000]
Srxn1 S9+ 0.0000 [0.0000, 0.0000]
TGx-DDI 0.1808 [0.0000, 0.3157]
Srxn1 S9- 0.7792 [0.7528, 0.7997]
Blvrb S9- 0.8682 [0.8596, 0.8783]
Ddit3 S9- 0.5298 [0.3618, 0.7194]

quercetin

Score = 0.3661 [0.3066, 0.4125]
Rank = 12 [12, 12] (out of 18)

Btg2 S9- 0.4621 [0.4059, 0.4744]
Rtkn S9- 0.4185 [0.3801, 0.4604]
Bscl2 S9- 0.4106 [0.4104, 0.4108]
Bscl2 S9+ 0.4115 [0.3431, 0.4395]
Rtkn S9+ 0.4766 [0.4123, 0.5221]
Btg2 S9+ 0.4052 [0.3399, 0.4592]
Ddit3 S9+ 0.0000 [0.0000, 0.0000]
Blvrb S9+ 0.3753 [0.3272, 0.4553]
Srxn1 S9+ 0.4323 [0.3571, 0.4979]
TGx-DDI 0.2840 [0.1868, 0.3565]
Srxn1 S9- 0.6865 [0.4926, 0.8374]
Blvrb S9- 0.0000 [0.0000, 0.0000]
Ddit3 S9- 0.4794 [0.4507, 0.5049]

enu

Score = 0.1916 [0.1593, 0.2196]
Rank = 11 [11, 11] (out of 18)

Btg2 S9- 0.3034 [0.2792, 0.3034]
Rtkn S9- 0.2923 [0.2677, 0.3187]
Bscl2 S9- 0.2353 [0.2339, 0.2366]
Bscl2 S9+ 0.3067 [0.2388, 0.3557]
Rtkn S9+ 0.2825 [0.2522, 0.3255]
Btg2 S9+ 0.2781 [0.2193, 0.3383]
Ddit3 S9+ 0.0032 [0.0000, 0.0042]
Blvrb S9+ 0.0697 [0.0098, 0.1056]
Srxn1 S9+ 0.3078 [0.2406, 0.3770]
TGx-DDI 0.0000 [0.0000, 0.0000]
Srxn1 S9- 0.2611 [0.2434, 0.2840]
Blvrb S9- 0.2055 [0.1707, 0.2356]
Ddit3 S9- 0.1361 [0.0753, 0.1899]

2DG

Score = 0.1448 [0.1172, 0.1677]
Rank = 10 [9, 10] (out of 18)

Btg2 S9- 0.2013 [0.1805, 0.2022]
Rtkn S9- 0.1989 [0.1901, 0.2063]
Bscl2 S9- 0.1679 [0.1651, 0.1708]
Bscl2 S9+ 0.1846 [0.1430, 0.2151]
Rtkn S9+ 0.1725 [0.1537, 0.1989]
Btg2 S9+ 0.1688 [0.1304, 0.2098]
Ddit3 S9+ 0.1903 [0.1417, 0.1903]
Blvrb S9+ 0.1441 [0.1218, 0.1813]
Srxn1 S9+ 0.1826 [0.1427, 0.2237]
TGx-DDI 0.0000 [0.0000, 0.0000]
Srxn1 S9- 0.1637 [0.1491, 0.1826]
Blvrb S9- 0.1223 [0.0516, 0.1833]
Ddit3 S9- 0.1305 [0.0704, 0.1836]

sodiumdiclofenac

Score = 0.1442 [0.0967, 0.1832]
Rank = 9 [9, 10] (out of 18)

Btg2 S9- 0.0000 [0.0000, 0.0000]
Rtkn S9- 0.0000 [0.0000, 0.0000]
Bscl2 S9- 0.1424 [0.0122, 0.2649]
Bscl2 S9+ 0.1788 [0.1347, 0.2149]
Rtkn S9+ 0.1673 [0.1508, 0.1907]
Btg2 S9+ 0.1800 [0.1407, 0.2210]
Ddit3 S9+ 0.2227 [0.1540, 0.2227]
Blvrb S9+ 0.1460 [0.1275, 0.1768]
Srxn1 S9+ 0.1847 [0.1434, 0.2277]
TGx-DDI 0.1163 [0.0000, 0.2030]
Srxn1 S9- 0.2805 [0.2548, 0.3135]
Blvrb S9- 0.0000 [0.0000, 0.0000]
Ddit3 S9- 0.2842 [0.2359, 0.3270]

eugenol

Score = 0.1270 [0.0908, 0.1570]
Rank = 8 [8, 8] (out of 18)

Btg2 S9- 0.0000 [0.0000, 0.0000]
Rtkn S9- 0.1303 [0.1191, 0.1364]
Bscl2 S9- 0.1142 [0.1106, 0.1176]
Bscl2 S9+ 0.0377 [0.0000, 0.0558]
Rtkn S9+ 0.0790 [0.0000, 0.1349]
Btg2 S9+ 0.2163 [0.1834, 0.2418]
Ddit3 S9+ 0.1496 [0.1049, 0.1496]
Blvrb S9+ 0.2639 [0.2212, 0.3351]
Srxn1 S9+ 0.3023 [0.2668, 0.3202]
TGx-DDI 0.0000 [0.0000, 0.0000]
Srxn1 S9- 0.1148 [0.1008, 0.1257]
Blvrb S9- 0.0791 [0.0096, 0.1416]
Ddit3 S9- 0.2905 [0.1583, 0.4398]

mannitol

Score = 0.0516 [0.0196, 0.0687]
Rank = 7 [6, 7] (out of 18)

Btg2 S9- 0.1154 [0.0888, 0.1256]
Rtkn S9- 0.0000 [0.0000, 0.0000]
Bscl2 S9- 0.0424 [0.0000, 0.0823]
Bscl2 S9+ 0.0792 [0.0699, 0.0792]
Rtkn S9+ 0.0634 [0.0000, 0.1083]
Btg2 S9+ 0.0774 [0.0336, 0.0975]
Ddit3 S9+ 0.1284 [0.0604, 0.1309]
Blvrb S9+ 0.0658 [0.0000, 0.1052]
Srxn1 S9+ 0.0717 [0.0000, 0.1095]
TGx-DDI 0.0000 [0.0000, 0.0000]
Srxn1 S9- 0.0565 [0.0000, 0.1005]
Blvrb S9- 0.0223 [0.0215, 0.0230]
Ddit3 S9- 0.0000 [0.0000, 0.0000]

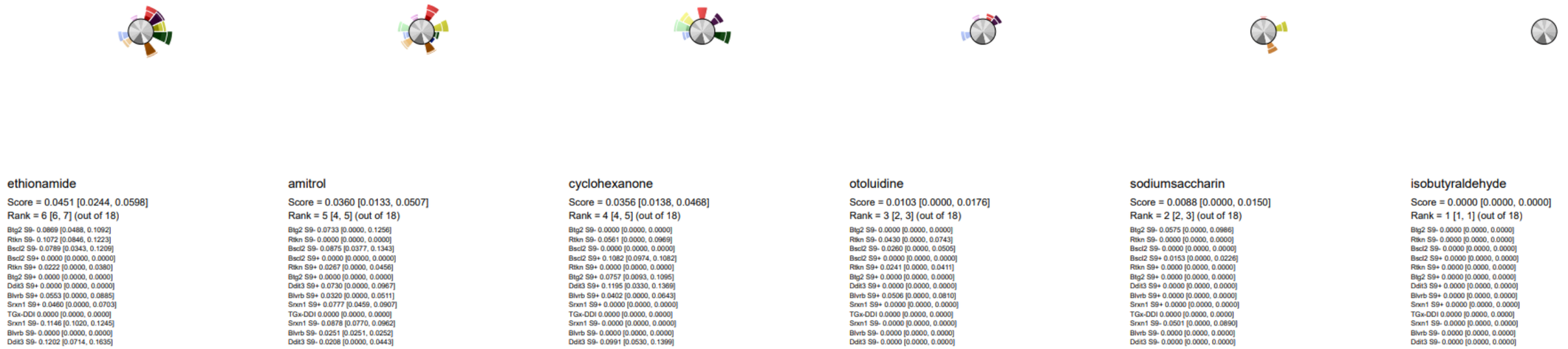


Figure 3.3.2.3 – ToxPi profiles (with bootstrap CI) generated from BMD data for each chemical, showing the biomarkers and their ToxPi scoring for each end point. Each slice represents a biomarker at a given time period in the presence/absence of S9. Shading in the centre of the circle represents uncertainty for each slice (darker = more uncertain). The radius of each slice (from the centre of each circle to the either the edge or white line towards the edge of the slice) represents the potency of each biomarker and the widths the user defined weightings. The brighter regions presented in each is the bootstrap CI where outer edge shows the upper CI, and the inner edge is the lower CI for each slice.

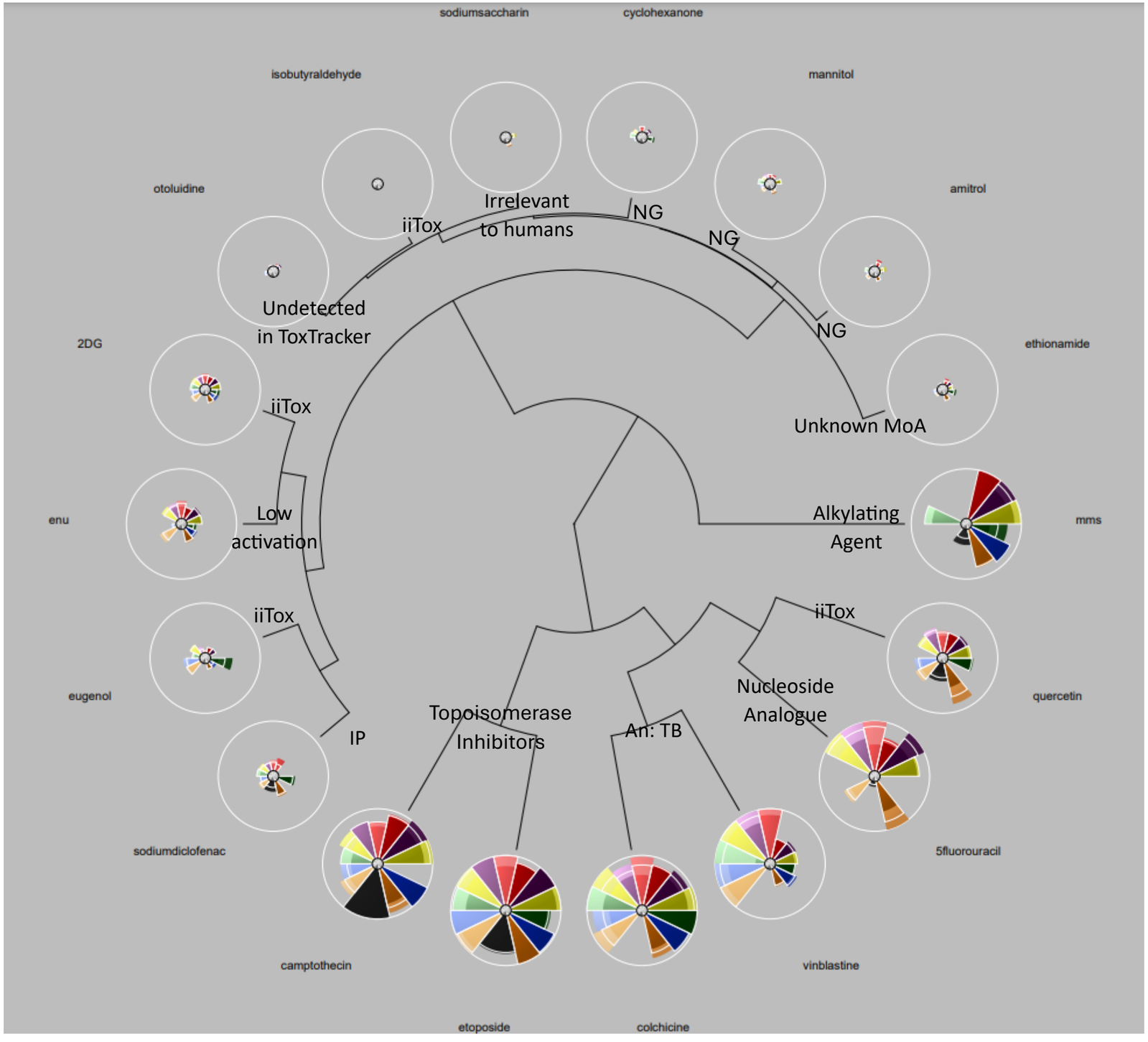


Figure 3.3.2.4 – ToxPi profiles hierarchical clustering generated using complete AGNES clustering on modelled BMDL and BMDU data, each color-coded based on their genotoxic potency in the presence/absence of S9 to show how each of the 18 chemicals group together based on similarity in responses across the ToxTracker and TGx-DDI endpoints. The cluster highlights irrelevant positives (IP), tubulin binding aneugens (An: TB), incorrectly identified compounds in the toxtracker assay (iiTox) and non-genotoxicants (Non-G).

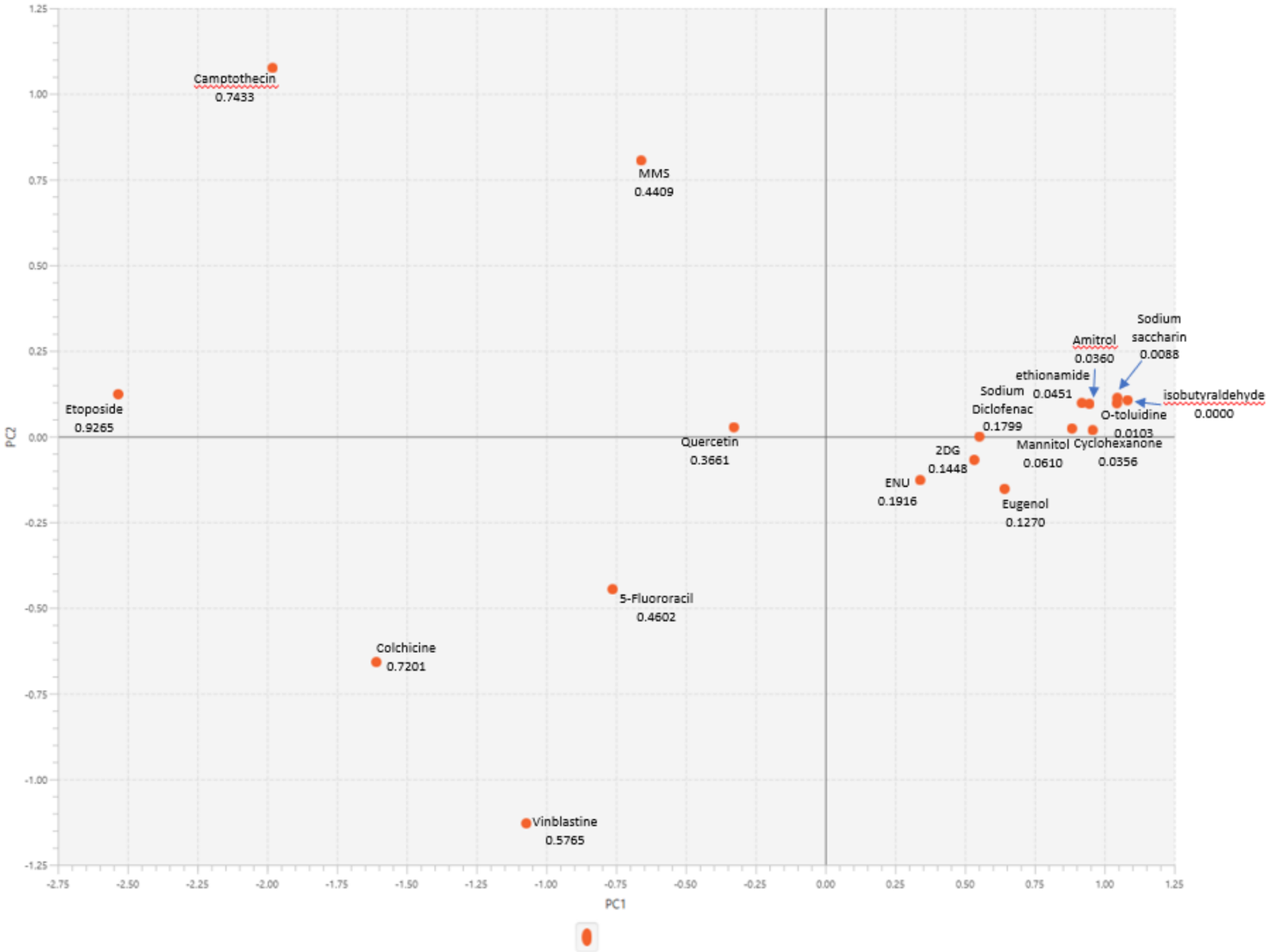


Figure 3.3.2.5 – ToxPi K-means clusters generated using PCA and divisive clustering algorithms to group compounds on a 2D plane defined by the first (PC1) and second (PC2) principal components, representing greatest sources of variability in the data set using a K-value (K=1 for exploratory purposes to avoid forced partitioning). PC1 reflects overall ToxPi potency, with more potent compounds positioned to the left, while PC2 captures biomarker variation, separating compounds with greater S9− potency (below the axis) from those with greater S9+ potency (above the axis). Compounds with similar potency and biomarker response profiles cluster in closer proximity to each other.

3.4 – Discussion

This chapter presents a discussion of novel genotoxicity testing strategies designed to improve chemical classification and potency assessment. The ability for the combined ToxTracker assay and TGx-DDI biomarker approach to offer a highly sensitive and mechanistic insight are explored, while addressing previous limitations in specificity and reducing the reliance on *in vivo* models, in line with the 3Rs principle (Russel and Burch, 1959). This section also considers any implications surrounding the tests design and the benefits, challenges, and future directions of this emerging homogenized testing system, using the results to illustrate their potential in enhancing safety and protecting human health. The development of these pipelines, which integrate both *in vitro* assays and computational models, marks a significant advancement in the detection of genotoxic compounds.

3.4.1 – Pipeline design Limitations

The results from section 3.3 clearly display that both PROAST and ToxPi are highly effective in enhancing comparative analysis. These tools excel in ranking compounds based on genotoxicity, determining endpoint potencies, and organising data into clusters or groups. This approach enhances understanding of compound-specific response magnitudes across endpoints through effective visualisation, thereby improving data interpretation and facilitating comparison within complex assay datasets. However, as explained earlier in Chapter 2, the reliance on quantitative data is a limitation, as it cannot provide numerical information and is dependent on modelling algorithms. ToxPi has additional limitations, including restricted features that prevent users from customising parameters or selecting specific endpoints and compounds. As a result, advanced filtering and data curation are necessary and unavoidable in this chapter.

Another issue with ToxPi is its reliance on relative comparisons through its minimum-maximum normalisation process. This can lead to inconsistent results when analysing compounds (such as clastogens and aneugens) against unrelated biomarkers. The choice of endpoints at the start of the study may cause compounds to show unexpectedly high or low potency. For example, a compound might appear highly potent in one study but much less potent in a similar experiment. In the original study by Allemang et al. (2020), several compounds were incorrectly classified in the ToxTracker assay, either as false positives or false negatives, which contradicts the existing literature. These misclassifications include:

1. Sodium diclofenac: Incorrectly identified as genotoxic due to oxidative stress at high concentrations. At elevated doses, secondary stress-related reporter activation may have driven a false-positive classification rather than true DNA damage.
2. Eugenol: Misclassified as genotoxic despite being primarily an antioxidant that activates srxn1. Reporter activation may reflect redox modulation rather than direct genotoxicity. Additionally, as a complex material reporter positive in some *in vitro* and *in vivo* studies, its context-dependent behaviour may contribute to inconsistent classification.
3. Quercetin: Incorrectly categorised, potentially due to its mixed/pleiotropic MoA. Although, negative *in vivo,* quercetin can modulate multiple stress and signalling response pathways *in vitro* making it difficult to classify as this produces reporter activation patterns that resemble genotoxic responses. This highlights the importance of using multiple complementary assays to distinguish primary genotoxic mechanisms from secondary stress-related effects.
4. 2DG: Incorrectly identified as genotoxic, as it is a non-genotoxicant. Its positive response may reflect metabolic stress or secondary cellular disruption rather than direct DNA damage, leading to reporter activation unrelated to genotoxic mechanisms.
5. ENU and o-toluidine: Both demonstrated lower-than-expected activation, with o-toluidine going undetected in the ToxTracker assay. This may reflect limitations in metabolic activation capacity, insufficient sensitivity of specific reporters or assay-specific constraints in detecting certain mechanistic MoAs.
6. Isobutylaldehyde: Although associated with oxidative stress, it failed to activate srxn1 or blvrb reporters in ToxTracker. Given its negative result *in vivo* comet and MN assay, this may indicate insufficient intracellular stress signalling under the existing experimental paradigm or highlight differences between *in vitro/in vivo* potency qualities.
7. MMS: Demonstrated greater potency than eugenol, with its well-established role as a direct-acting N7-alkylating mutagen. However, the relative potency differences observed between the two compounds likely reflects assay-specific sensitivity, modelling parameters or context-dependent biological activity, which has shown variable *in vitro* and *in vivo* responses.

These discrepancies highlight the limitations in the design of the chosen pipeline and the shortcomings of the original study's ToxTracker assay results. While the pipeline has utility, this is only true when compounds are correctly identified and aligned with existing literature. This raises concerns for the analysis of unknown compounds, emphasising the need for additional reference compounds with stable MoAs to ensure accurate processing of BMD metrics. Although the approach may have some mechanistic value as a screening tool, its cost and complexity may limit its practicality considering the availability of the Ames test, which could be used for initial screening. Future studies must focus on generating robust datasets to prevent inconsistencies that could skew the results away from what is supported by prior research.

3.4.2 – PROAST BMD individual and collective outputs

For majority of the analysis, PROAST effectively generated appropriate BMDs for both the ToxTracker and TGx-DDI assays, managing the BMD production process efficiently. PROAST remained relatively stable and fulfilled its purpose by producing BMD CIs within the bounds of these assays, though some aspects of the process still warrant discussion.

First, the TGx-DDI assay could only perform BMD analysis on known DDI compounds. While non-DDI compounds had sufficient data to produce BMD curves, to prevent false or irrelevant positives, the analysis was limited to compounds such as 5-fluorouracil, camptothecin, ENU, etoposide, MMS, o-toluidine (DDI at one concentration of 9330 μM), quercetin, and sodium diclofenac. Next, the ToxTracker assay required a minimum of five dose concentrations for each reading, as the data it generated was not as extensive as the 64-gene data produced by TGx-DDI. This made the TGx-DDI assay more likely to yield stable results during the bootstrapping process, as more data points were available, though stability was not guaranteed.

Additionally, the TGx-DDI figures required two transformations: normalisation (adding 1 to each fold change value to avoid negative values, which PROAST cannot process) and control readings (setting the dose at 0 μM to a fold change value of 1.000). As the datasets displayed a lot of variation based on non-DDI/raw fold change figures, this variability likely reflects the variability the diversity of compound MoAs given the differences in raw fold change figures across all 64 genes. Unlike non-DDI probability values (an indicator of DDI), raw fold

changes provide a direct consistent measure of transcriptional response, enabling a more comparative assessment across compounds with different MoAs. Therefore, raw data is required to assist in the correct modelling of the BMD and any summary/induction results would skew the data from the expected points/produce increasing separation with the raw GFP induction/fold change figures; this was only an issue for TGx-DDI.

Furthermore, due to the data-intensive nature of TGx-DDI assay results means that BMDExpress (a bioinformatics tool data used to produce BMDs by analysing toxicogenomic datasets), however in this instance, the R Gui version was implemented for BMD analysis, which may introduce further variation to the data. However, these differences suspected that this would be minimal as both formats use all data across all genes to produce BMDs in a similar process with only difference being that BMDExpress streamlined this process, despite both findings likely being virtually identical. Strict guidelines were followed regarding CES sizes, model selection, and bootstrapping, as any changes, particularly to CES, could result in either successful BMD generation or no dose – response findings. Finally, some compounds remained undetected at higher concentrations due to cytotoxicity, leading to inflated BMD values. This was caused by reduced GFP induction at these higher levels, which skewed the model and potentially masked lower BMDs that would indicate higher potency.

3.4.3 – ToxPi scoring ranks and profiles

ToxPi had successfully converted BMD results figures generated by PROAST and was able to relatively display the rankings of the 18 compounds addressed in this study. Despite the ToxPi displaying the potency rankings in an articulate and condensed manner, the rankings themselves might involve some issue due to the relative nature of the assay, particularly for compounds with their MoA existing outside of the bounds of the used assay and are unable to be detected. For example, vinblastine and colchicine were unable to be detected for the TGx-DDI endpoint, meaning their potency is likely to be skewed downwards. A feature that could circumvent this issue would be an option where the endpoints chosen for each compound could be selected so for aneugenic compounds, the TGx-DDI endpoint could be removed for the compound's assessment and still compared to all other compounds, preventing the overall skewing of known potent compounds for the mere virtue of their MoA being undetectable. However, for newer chemicals, endpoint selection would need to follow a predefined, mechanism-agnostic rule, such as excluding reporters that consistently show no activation across a defined activity threshold. Such an approach would avoid post hoc tailoring while

allowing comparisons to be made on biologically responsive endpoints instead of penalising chemicals whose MoAs fall outside of the captured reporter assays. Accordingly, such a framework would require systematic evaluation prior to implementation.

Next, despite how non-genotoxic and genotoxic compounds separated quite well into their respective ranks, some compounds with known genotoxic properties, such as O-toluidine, which was not detected as a DDI at their highest concentrations due to cytotoxicity, which would increase its BMD value, producing a lower-than-expected ToxPi score, skewing it towards similar ToxPi scores to non-genotoxic compounds. This effect could be mitigated by introducing cell viability as a covariate using a cytotoxicity-adjusted modelling approach, incorporating additional doses with smaller concentration increments that do not induce excessive cytotoxicity, or modifying exposure conditions to extend the measurable sub-cytotoxic range and improve BMD estimation. Usually O-toluidine would demonstrate increasing DDI capabilities at higher concentrations and non-DDI at lower ones, however, cytotoxicity at the top concentration would incorrectly identify o- toluidine as non-DDI concentrations – unlike camptothecin, ENU, etoposide, and MMS displayed similar levels of cytotoxicity but being correctly identified as DDI. Furthermore, compounds such as MMS (oxidative stress) obtaining higher ToxPi compared to eugenol (mixed MoA) suggests incorrect identification by ToxTracker is producing lower relative potencies for compounds that are incorrectly identified. Eugenol appears to significantly activate the Rtkn and Srxn1 endpoints at higher concentrations in ToxTracker, suggesting DNA damage and oxidative stress properties, however, the Srxn1 endpoint activation should not occur due to its antioxidant properties, potentially arising from pro-oxidant effects at higher concentrations; the original Allemang study does address this point (Pramod, Ansari, & Ali, 2010).

Although the study clearly highlights the distinction between S9- and S9+ conditions, however, greater emphasis was placed on the combinational aspects of the assays and their results. Additionally, incorrectly identified endpoints means that some compounds (i.e. eugenol) profiles may not be an accurate representation of what the literature suggests surrounding its mode of action. Despite this, the analysis of metabolic activation would still be considered as significant as they further add to the understanding of these compounds through their profiles, bringing us a greater understanding of their features and properties; a necessary process in the development of our knowledge surrounding these genotoxic/non-genotoxic compounds. For each BMD dataset, values were transformed as outlined in

Chapter 2, scaling BMDs between 0.0001 and 10,000 μM (the biologically tested range). This transformation was applied for visual scaling purposes to reduce the disproportionate influence of extreme values or highly skewed endpoints on the overall dataset.

Lastly, a 2-fold (14.3%) weighting was applied to the TGx-DDI endpoint, following a similar approach to the method used in (Fortin et al., 2023), where endpoints are weighted based on their relative significance within that assay. In this case, the 2-fold weighting was justified by the TGx-DDI assay's aggregation of 64 genes into a single endpoint combined with its ability to identify DDI/non-DDI compounds holding more significance, in contrast to the ToxTracker assay, where individual endpoints from that assay were weighted at 1-fold (7.1%) each. Although the 2-fold weighting for TGx-DDI was chosen for this study, alternative weightings could be considered. However, the lack of a standardized numerical quantification system within ToxPi complicates the selection of other values. Any changes to the assigned weightings would require comprehensive justification and remain open to scrutiny. This challenge is further compounded when attempting to establish consistent weightings across multiple assays. Therefore, further work is needed to develop a robust framework for assigning appropriate weightings across different assays, whether they are used individually or in combination.

Finally, Topoisomerase II clastogens like etoposide and camptothecin exhibit higher ToxPi values due to their lower BMDs, likely influenced by the lower concentration ranges tested and enhanced activation at these lower doses to mitigate cytotoxicity, particularly in aneugens. This skews ToxPi scores towards 1.000, suggesting uniform endpoint activation, which is not the case. However, raw dose-response data indicate variability in biomarker activation, which ToxPi may not accurately capture since its scores are relative to each dataset. When compounds with distinct MoAs are analysed, variation in BMD confidence intervals can distort relative scaling, which may reduce sensitivity for identifying very potent cytotoxic or mutagenic compounds when cytotoxicity reduces the dose–response data used in modelling or alternatively exaggerate apparent potency depending on the composition of the dataset.

3.4.4 – ToxPi Hierarchical/Kmeans Clustering

Clustering analysis was performed using ToxTracker, TGx-DDI data, and BMD data through

ToxPi. This revealed that compounds with distinct MoAs and higher potency clustered together. For instance, colchicine and vinblastine, both tubulin-binding aneugens, were clearly separated from topoisomerase I/II inhibitors, such as camptothecin and etoposide. However, other compounds exhibited more sporadic clustering patterns. For example, isobutylaldehyde, despite its oxidative stress properties, does not activate the Blvrb/Srxn1 biomarkers in TK6 cells, aligning with the fact it does not produce a measurable transcriptional activation response. Similarly, 2DG, a known glycolysis inhibitor, showed endpoint activation patterns similar to sodium diclofenac, with both activating the srxn1 (oxidative stress) and Rtkn (DNA damage) pathways. Nevertheless, experimental MN and TGx-DDI assays identified them as non-DDI, aligning with previous studies that classify sodium diclofenac as a non-genotoxic compound. Despite this, sodium diclofenac did show Rtkn/Srxn1 activation (particularly srxn1 at higher doses in aquatic environments) and DDI activity across multiple concentrations in additional MN tests (not processed in this section), though it had a lower ToxPi score than 2DG. Misclassified compounds that did not activate key endpoints related to their known MoA, or activated unrelated endpoints, skewed the clustering, pulling compounds away from their true groupings.

Quercetin was misclassified as non-DDI by ToxTracker and TGx-DDI at certain concentrations. However, at higher concentrations, it has been observed to induce DDI. Despite its mixed MoA, including topoisomerase II inhibition and pro/antioxidant properties, quercetin surprisingly clustered closer to aneugenic tubulin binders rather than topoisomerase I/II inhibitors. This could be due to the similarity in profiles with aneugenic compounds and its lower DDI projection at the non-DDI dose. Quercetin's challenging identification, as noted by Boots et al. (2007), adds to this complexity, as it behaves differently across models. For example, it has been known to test positive *in vitro* and negative *in vivo*.

Initially, hierarchical clustering was applied using only ToxTracker endpoints, resulting in many compounds being incorrectly clustered. However, adding the TGx-DDI endpoint, which provided further insights into compound potency, improved clustering accuracy. Despite the addition of the TGx-DDI reporter improving clustering performance, it is not sufficient to override classification errors arising from the original experimental findings producing BMD CIs and ToxPi profiles consistent with the MoAs of those chemicals. K-means clustering is a standard output, but the PC2 metric requires further analysis as it varies significantly depending on the selected endpoints. PC1, on the other hand, tends to align

more consistently with the ToxPi score. This clustering approach seems to align with the hierarchical system, effectively grouping compounds based on their ToxPi scores and profiles. Adding MN endpoints would offer even deeper insights into the MoA and DDI potential of compounds as it would provide information on whether any mutations led to physical manifestations of MN formation, indicating genotoxicity. While this was not included in the current analysis, the original study Allemang et al., (2020) aimed to identify where such skewing issues were likely to arise.

3.4.5 – Improvements

This section covers potential improvements that would significantly enhance the robustness and clarity of future clustering analyses and increase the overall effectiveness of compound classification. To improve future analyses, several potential enhancements can be made. One important adjustment involves revising the TGx-DDI endpoint weighting. Currently, this is based on a priority system but switching to a more robust and standardised weighting figure could enhance the accuracy of compound classification and clustering outcomes. This refinement would better reflect the relative importance of different endpoints and improve the overall reliability of the data. In addition, a more refined selection of compounds related to oxidative stress biomarkers could improve the clustering results. In future studies, selecting a series of reference compounds with clearly defined and distinct MoAs would help enhance hierarchical and k-means clustering in both TK6 and HepaRG cells, allowing for more precise grouping of compounds based on their mechanisms of action. This step would help reduce the sporadic clustering observed in some analyses. Another key improvement would be the inclusion of *in vitro* MN data as an endpoint in the analysis to measure the physical manifestation of chromosomal damage measure that compliments ToxTracker and TGx-DDI gene expression to enhance the ability to strengthen the ranking, cluster discrimination and robust profiling of compounds. However, an additional approach would be to perform ToxPi analysis by incorporating compounds with a similar MoAs. This could be highly beneficial in providing additional activation insight across endpoints whilst avoiding this skewing effect brought on by the same dataset having compounds with very high/very ToxPi score profiles where compounds either have all their endpoints activated or none at all. At the same time, the strength of ToxPi lies in integrating multiple complementary datasets, so including mutation and MN endpoints alongside reference compounds with distinct MoAs could further improve mechanistic discrimination in the clustering and help balance contributions across diverse endpoints across ToxPi profiles.

Incorporating *in vitro* MN provides additional mechanistic insights by directly measuring chromosomal damage by providing information on physical MN manifestation. This complements ToxTracker biomarkers by detecting compounds that significantly compromise chromosomal integrity or lead to chromosomal genotoxicity as some compounds trigger stress without producing MN. This enables MN data to distinguish between compounds that cause transcriptional DNA damage and those that produce chromosomal alterations, improving our understanding of compound MoA. Furthermore, it may be beneficial to remove S9 metabolic activation data from the analysis. As discussed earlier, this data does not add significant value to the current framework and could be excluded to streamline the results without negatively impacting the overall conclusions. Some issues identified in the clustering and identification processes are tied to experimental discrepancies from the original Allemang et al., (2020) study. However, these inconsistencies did not interfere with the capabilities of tools like PROAST and ToxPi in summarising complex datasets into clear, digestible figures. These tools successfully presented the experimental results in a way that allowed for effective data interpretation, despite the experimental challenges. Lastly, the study design itself could be optimised to maximise information extraction and enhance the assessment of compound potency. By refining experimental protocols and adjusting endpoint selection, future analyses could yield more accurate and insightful results, enabling more precise evaluations of compound activity. These improvements in combination would help to address the limitations in the current analysis, resulting in more reliable clustering, better compound differentiation, and clearer insights into compound mechanisms and potency.

3.5 – Chapter conclusions

The results of the data analysis performed on the ToxTracker and TGx-DDI assays demonstrate the ability for two separate toxicogenomic assays to provide information on a wide array of endpoints by combining data from both reporter-based and expression profiling tests. Although TGx-DDI did not identify all DDI-positive compounds, it contributed additional transcriptional context that supported mechanistic interpretation and clustering analyses. These methods provide a highly sensitive and mechanistic insight into compound classification while addressing limitations in specificity and reducing reliance on *in vivo* models, aligning with the 3Rs principle (Replacement, Reduction, and Refinement) (Russel and Burch, 1959).

This testing paradigm enables identification of genotoxic potential and MoA classification in human-relevant *in vitro* systems, supporting hazard screening and prioritisation. As a result, only compounds meeting clearly defined concern criteria would progress to *in vivo* testing, thereby reducing unnecessary animal studies and avoiding progression of compounds with minimal or uncertain genotoxic potency. The combination of *in vitro* assays and computational models like PROAST and ToxPi represents a significant progression in hazard identification and potency assessment of genotoxic compounds. Despite certain limitations, these tools effectively rank compounds, determine endpoint potencies, and simplify comparative analysis. By enabling earlier and more quantitative identification of genotoxic chemicals, this approach supports risk-based prioritisation and informed decision-making during drug development, reducing the likelihood that false negative genotoxic compounds are not overlooked and advance to later-stage testing or clinical evaluation.

However, challenges remain, including issues with endpoint selection and the need for more standardised and robust weighting systems in TGx-DDI. Misclassifications of certain compounds, such as sodium diclofenac and quercetin, point to the necessity for a more refined selection of reference compounds with distinct MoA to enhance clustering accuracy. Additionally, while the incorporation of TGx-DDI data improved clustering outcomes, certain errors persisted, suggesting the need for further refinement in endpoint integration. Future studies should consider removing less informative endpoints, such as S9 data, and focus on incorporating additional relevant endpoints, such as *in vitro* MN data, to provide deeper insights into compound MoA. Addressing these experimental discrepancies and optimising study design will allow for more precise evaluations of compound potency and improve the accuracy of clustering and profiling results. This approach allows for the detection of genotoxic compounds that would normally test negative in other *in vitro* genotoxic tests assays (i.e. AMES/MN) supporting a weight-of-evidence assessment and reducing the likelihood or false positives during hazard assessment.

Ultimately, this chapter underscores the potential of integrating mechanistic assays and computational models in genotoxicity testing, by providing earlier, mechanistic insight into compound hazards, improving prioritisation, and supporting more informed safety assessments, marking a step forward in protecting human health. These findings lay the groundwork for continued refinement and development in the field of genotoxic compound analysis, offering a promising path for future research and application.

Chapter 4 – Developing Benchmark Dose-Based Analytical Pipelines for the iScreen High-Content Imaging Platform Mechanistic Classification of Aneugenic and Clastogenic Compounds in Genotoxicity Assessment

4.1 Introduction

As briefly discussed in chapter 2.4.3.2, using the example of diethylnitrosamine (DEN), some assays implement systems where MoA overlap may likely occur, and the testing system implemented may detect various instances of compounds activating endpoints similar/related indirectly related their respective MoA. In some cases, detecting genetic events, such as DNA strand breaks or DNA adducts, requires sensitive methods using specialised probes, antibodies, or microbes to characterise challenging and highly sensitive/specific mechanisms (OECD., 2016). Additionally, a testing system focused on detecting gene mutations and chromosomal aberrations *in vitro* can provide mechanistic insight and support the identification and prioritisation of genotoxic hazards, addressing limitations of traditional scoring systems in early drug discovery, particularly when screening large numbers of compounds. Flow cytometric automation and image analysis increases the capabilities of these traditional systems to analyse certain endpoints, however, a newer mechanistic insight into specific chromosomal aberration could provide further information into carcinogenic compounds and allow for a robust automated testing environment for a large series of compounds to be compiled for increased dataset information extraction in a time-efficient manner (Bryce et al., 2010; 2013). Additionally, it provides an avenue for the implementation of robotics systems in liquid management devices and Artificial Intelligence (AI) image analysis software to increase quality of analysis while simultaneously handling large-scale chemical screenings efficiently.

The development of the iScreen high content imaging platform was designed to provide a high content multiplexed imagining platform using a series of antibodies and imaging tools to analyse the ability of compounds to induce some degree of chromosomal aberrations through the incorporation of centromere detection in MN, DNA damage response and MN fragmentation (Sun et al., 2022). This testing is essential in the evaluation of potential DNA damaging chemicals that could be precursors to carcinogenesis by providing a reliable high-throughput method to detect such damage. Changes to the chromosomes through deletion, insertion and substitution of the DNA sequences may result in chromosomal malformation and fragmentation, making these changes to the chromosomes key indicators of genomic

instability. While such effects are often assessed within the standard testing battery, it remains important to examine mechanistically, particularly in cases where certain MoAs may not be fully captured by individual DNA damage detection methods/assays (Li et al., 2019; Wilson et al., 2021). Furthermore, this allows for early-stage detection of these potential carcinogens further developing toxicological studies and therapeutic treatments of these compounds.

Current high content imaging technologies developed an automated 96-well *in vitro* assay analysis technique, such as the MultiFlow/MicroFlow assays. These systems utilise flow cytometry software platforms such as BD FACSDiva alongside BD flow cytometers to image and quantitatively analyse cell quantities, fold changes or fluorescence induction through the application's respective data acquisition, analysis and gathering tools available. The iScreen assay employs immunofluorescent labelling and image-based quantification to detect chromosomal abnormalities by visualising MN, which indicate chromosomal fragments or whole chromosomes not incorporated into daughter nuclei (Campbell et al., 2014; Sun et al., 2022). This approach enables high-resolution imaging, allowing for precise detection of indicators of genomic instability. These abnormalities, such as whole chromosome loss or chromosomal fragments, are key markers of genotoxicity and are closely associated with cancer development. Therefore, these endpoints are crucial for identifying genotoxic agents and classifying their MoA.

The utilisation of the iScreen assay provides significant advantages through its high-throughput capabilities making it more efficient as opposed to traditional methods to produce automated and quantifiable chromosomal aberration results rapidly whilst retaining reproducibility. The detection system remains highly sensitive, enabling the identification and quantification of key genotoxic events, such as DNA strand breaks, gene mutations, chromosomal fragmentation and whole chromosome loss, thereby allowing targeted regions and damage-response pathways to be assessed. The highly broad applicability of the iScreen assay allows for strong evaluation of high number of compounds and detecting highly specific MoAs simultaneously providing a comprehensive view of compound *in vitro* genotoxicity (Sun et al., 2022).

For this chapter, the iScreen assay was utilised to assess 10 compounds with mixed MoAs providing a cohesive system/pipeline which translates complex iScreen datasets into efficient visualised models/figures for the purposes of maximising information extraction and

comparatively analysing potential genotoxic compounds based on MoA and AOP using a multiplexed image analysis platform. The A375 cell line was chosen for this assay due to its rapid growth, which facilitates high-throughput screening, and its stable background MN frequency, which allows for accurate high-content imaging—making it a robust model for evaluating anti-cancer compounds. A comprehensive approach adopted from the iScreen, MultiFlow, MicroFlow (MicroFlow assay using TK6 cell lines) and centromere detection staining, in combination, provided an extensive analysis into the MoA behind these compounds to discriminate them into their respective categories: aneugens (aurora kinase inhibitors/tubulin binders), clastogens and non-genotoxic compounds. CENPA antibody staining was used to label kinetochore regions allowing for the discrimination between MN with whole chromosomes or chromosomal fragmentation. Additionally, γ -H2AX and spot/cell figures were used to mark the presence of dsDNA to test for clastogenic MoA through the detection of dsDNA breaks. Finally, to detect and distinguish between aurora kinase inhibiting and tubulin binding aneugens, the MPM2 and p-H3 endpoints were used: an increase in in mitotic cell activity observed in tubulin binders and the dephosphorylation of p-H3 alongside polyploidy induction were used as an indicator for aurora kinase inhibiting activity. While the standard *in vitro* MN assay provides a regulatory approved screening system to identify manifestations of genotoxic damage, it is incapable of distinguishing between clastogenic aneugenic MoAs without additional testing (which adds additional cost/time). The iScreen assay addresses this by integrating additional multiparametric imaging endpoints – including p-H3, MPM2, polyploidy, yH2AX and CENPA to provide further mechanistic insight into compound's genotoxicity, providing highly sensitive and specific mechanistic insight to assist in early phase drug development. This designs a landscape for safe and efficient future drug development tests for new/existing compounds.

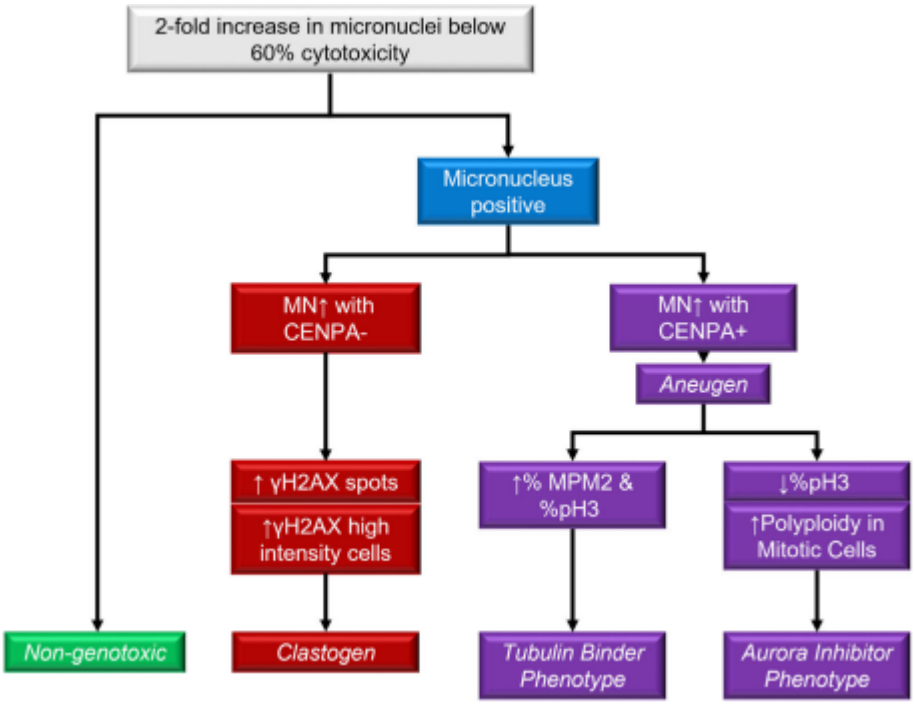


Figure 4.1.1 – Flowchart demonstrating the identification of each compound's classification based on their MoA using the endpoints across the iScreen assay (adapted from (Sun et al., 2022)).

4.2 Materials and Methods

This chapter outlines the materials and methods within the iScreen assays to curate and addresses the data processing performed summarising the Sun et al., (2022) paper and Litron Laboratories testing kits.

4.2.1 Cell lines, culturing and treatment

The cell line chosen for the iScreen assay are A375 cell lines, providing a well-defined malignant melanoma cell line with human origin and wild type p53, which are relevant to *in vitro* modelling to study cancer cell proliferation, drug resistance and MoAs due to their resistance to therapies, rapid growing and stable MN background frequency (Sun et al., 2022). The only instance where a different cell line is applied would be the MicroFlow assay where TK6 cells were used for reasons explained in previous chapters. A375 cells (from ATCC CRL-1619^{TM}) were grown in Dulbecco's Modified Eagle Medium (DMEM) (Gibco) comprising of 10% fetal bovine serum, penicillin (50 units/ml), 0.05 mg/ml streptomycin, and 2mM L-glutamine. Cells are seeded into 96-well plates at 5000 cells per well for 24-hour/ 7500 cells for 4-hour treatments. Cells are incubated for 24-hours prior to treatment in 5% CO_2 at 37°C. All compounds are exposed to each cell line at 4/24-hours for across each endpoint after preparation in 1% v/v DMSO through serial dilutions over a range of tested concentrations using serial dilution. Although the assessment of these compounds using S9 is possible, metabolic activation was not observed for the iScreen assay as the primary aim for this chapter was to detect and distinguish between aurora kinase inhibitors, tubulin binders (both aneugens) and clastogens. Therefore, the addition of S9 activated compounds would add additional metabolic variables beyond the scope of this initial mechanistic validation system (although the incorporation of S9 would be valuable for future work).

Table 4.2.1.1 – Chemicals used in this investigation alongside their classifications and tested concentration range adapted from Sun et al., (2022).

Compound	**CASRN, Sigma Aldrich**	**MoA classification (Sources)**	**Tested Concentration range (µM)**
AMG900	945595-80-2	Aneugen, Aurora Kinase Inhibitor (Payton et al., 2010)	0 – 0.100

Baraserntib	722544-51-6	Aneugen, Aurora Kinase Inhibitor (Yang et al., 2010)	0 – 0.500
Colchicine	64-86-8	Aneugen, Tubulin Binder (Kirkland et al., 2016)	0 – 0.100
Cisplatin	15663-27-1	Clastogen (Kirkland et al., 2016)	0 – 150.000
Dasatinib	302962-49-8	Clastogen (U.S. FDA, 2006)	0 – 0.225
Etoposide (ETO)	33419-42-0	Clastogen (Kirkland et al., 2016)	0 – 10.000
Mitomycin C (MMC)	50-07-7	Clastogen (Kirkland et al., 2016)	0 – 10.000
Nocodazole	31430-18-9	Aneugen, Tubulin Binder (Hendriks et al., 2016)	0 – 10.000
Taxol	33069-62-4	Aneugen, Tubulin Binder (Hendriks et al., 2016)	0 – 1.000
Tunicamycin	11089-65-9	Non-Genotoxic, Apoptosis Inducer (Guha et al., 2017)	0 – 50.000

4.2.2 Imaging Dataset Curation

Cells fixed with 2% polyoxymethylene in PBS solution at room temperature for the preservation of cell structures and immobilisation of proteins to ensure that cellular fixation can occur without damaging proteins. Once mixed with cold methanol, cell fixation dehydrates cells, denatures proteins, retains cellular morphology and preserve phospho-epitopes (a protein region that contains phosphorylated amino acids as a post translation modification to assist in the regulation of protein functionality) to enhance antibody accessibility and preserve antigen. After fixation, permeabilization with PBS solution (+-0.05% Titron X-100 detergent) allows for antibodies to penetrate and receive access to intracellular antigens target proteins located within cells. Additionally, blocking buffer (1% bovine serum albumin in PBS (0.05% Titron X-100)) blocked cells and diluted antibodies.

Next, primary antibody preparation took place where for the 4-hour timepoint, the primary antibody solution contained:

- Rabbit phospho-histone γH2AX
- Mouse Anti-phosphoSer/Thr-pro MPM2
- Rat p-H3

For the 24-hour timepoint, the primary antibody solution contained:

- Rabbit phospho-histone γH2AX
- Mouse CENPA

Cells are incubated with their respective primary antibodies for 2-hours before washing with PBS three times at room temperature before labelling. Subsequently cells were labelled with secondary antibody solution for 1-hour at room temperature in a dark environment. The cells were labelled with:

- Alexa Fluor® 488 mouse
- Alexa Fluor® 555 rabbit
- Alexa Fluor® 647 rat
- Hoechst 33342 (1µg/ml)

The 24-hour plates received a final labelling with conjugated MPM2-Cy5 for 1-hour in the dark followed by PBS washes for conjugation and finally, cell plates are prepared for image analysis after treatment with 200 µL of PBS solution (containing sodium azide). An Operetta CLS system (PerkinElmer), equipped with single band emissions filters, was used to acquire images using the four fluorescence channels based on the four cell labels applied and no antibody controls were required during imaging as the protocol developed ensures no fluorescence emission crosstalk detection, meaning no compensatory circumstances/conditions were required. Saturation was avoided through LED power management and PeriScan technology during image acquisition eliminated highly cytotoxic cell wells (<1000 cells), reducing imagine times. Plates were image scanned at 5X objective and analysed using Harmony^R 4.9 software (global whole image) to acquire cells/well figures which were subsequently, non-cytotoxic cells were scanned at 20X for 35 fields/well and a z-stacking covering 3mm over 4 planes.

For 4-hour treatments, the Hoechst 3342 channel was utilised for nuclei segmentation and regions were created based on nuclei re-segmentation resizing by 30 pixels from the outer border removing any border objects. The p-H3 and MPM2 channels were calculated, and thresholds were used to identify potential tubulin binders, where MPM2-positive cells were excluded to identify interphase cells. The high intensity γH2aX nuclei were isolated based on thresholds while γH2aX spot quantification in interphase cells was performed through spot detection. Overall, this means that outputs for the 4-hour treatment includes readings for p-

H3-positive nuclei, MPM2-positive nuclei, high intensity γH2aX nuclei and γH2aX spots/cell.

For 24-hour treatments, all readings mentioned, except for p-H3, are analysed (MPM2-positive nuclei, high intensity γH2aX nuclei and γH2aX spots/cell), with additional endpoints. MN were characterised by CENPA intensity, the same threshold used to determine CENPA-positive MN. This means the new endpoints include 24-hour cytotoxicity, MN, CENPA-positive MN , γH2aX-positive MN and MN-positive γH2aX. All results have their concentrations and respective endpoint outputs recorded which are further used before BMD analysis is performed to normalise these readings. For both MN and cytotoxicity readings, the MicroFlow assay was applied, using TK6 cells, following the methodologies described in earlier chapters to obtain their appropriate percentages/figures using the *in vitro* Microflow® Kit reagents (Litron Laboratories, Rochester, NY, USA) following protocols described by (Bryce et al., 2013; Avlasevich et al., 2011 & Litron Laboratories., 2018). To retain response relevancy, any compounds exhibiting cytotoxicity figures ≤60%/relative cell survival rate <40% were excluded from further analysis as described similarly in chapter 3.2.3.1 due to excessive cytotoxicity.

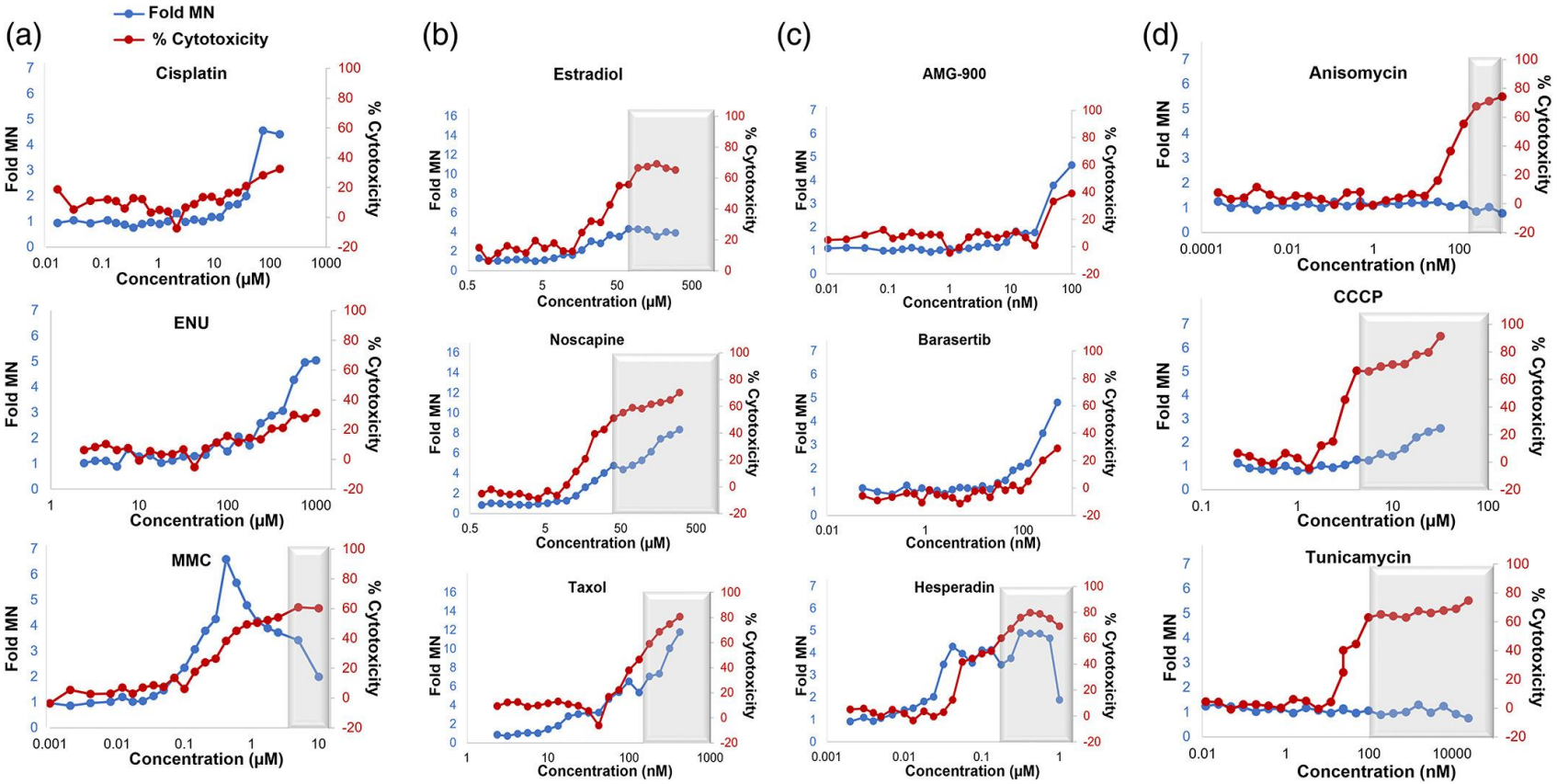


Figure 4.2.2.1 – Induction of MN and cytotoxicity from various compounds in A375 cells with red representing the rising/falling percentage of cytotoxicity with the increasing concentrations, where the y-axis showing the cytotoxicity percentage is on the right side of each graph. Additionally, the blue line represents the fold change increase/decrease in MN induction increase above the control (y-axis on the right) where (a) clastogens (cisplatin, ENU, MMC) (b) tubulin binders (estradiol, noscapine, taxol) (c) aurora kinase inhibitors (AMG900, barasertib, hesperadin) (d) non-genotoxicants (anisomycin, CCCP, tunicamycin). The grey region towards the higher concentrations represented the excessive cytotoxicity region where any concentration exceeding in ≥60% had their results removed from further analysis. Data shown are adapted from Sun et al., (2022).

4.2.3 Working in PROAST

As alluded to in previous chapters, each dataset gathered requires transformation and careful consideration for the type of data generated before any further processing commences. BMD as a normalisation step for dose-response data to gather a fixed figure based on a variety of response figures combined with concentrations provides the best assessment for these analyses across multiple assays. To generate these BMDs, the original dose-response figures were imported into PROAST (with the removal of any concentrations with a given response of 0) as csv files with dose being actively compared to all responses across all endpoints. Following this, a similar procedure as described earlier in this thesis (chapters 2.2.5.1 and 3.2.3) with the use of the R Gui PROAST with continuous individual data option being selected and “exponential models 3 or 5 from various model families" modelling was applied where concentration was set as the independent variable and each endpoints’ response figure set as the dependent variable with an absence on any covariates being selected. The CES was similarly set to +0.5 above the control with 200 bootstrap runs assisting in the generation of 90% BMD CI values, producing a BMD graph for each compound containing log10(dose) and log10(response) for the x and y axis axes as shown in figure 2.2.5.2.1 and once all BMDs are curated and processed individually through PROAST, an output as described in figures 3.3.1.1 – 3.3.1.7, is produced where each successfully ran BMD CI endpoint is simultaneously curated into a single graph with the log10(BMD CIs) being compared on the x-axis and each compound being displayed on the y-axis. This was performed for each endpoint and upon completion, BMD CI figures were further collected ready for processing through ToxPi where transformations to the dataset allowed for further visualisations to commence.

4.2.4 ToxPi Data Visualisation

Once PROAST gathered the BMD data across all 4/24-hour endpoints, the data was collated in the format of a CSV file and the 0.0001 – 10000 μM BMD limits are imposed on the original dataset, as discussed in the previous chapters, to aid in data visualisation purposes. Once all endpoints have been gathered, the CSV file is imported into ToxPi, all compounds are added to the analyses and $-\log_{10}X$ transformations are applied to each endpoint, where all ToxPi slices comprised of each biomarker’s respective BMD CIs (BMDL and BMDU). Finally, in this chapter, as the weighting is a major factor in the analyses, the weightings across the endpoints were modified using the ToxPi provided functionality. The purposes of

these weightings being applied are due to the biological significance of each biomarker, especially regarding their MoA, sensitivity and specificity. Endpoints vary highly in their relevance and due to their varying biological processes/mechanisms that they are referring to regarding their genotoxicity. Therefore, endpoints with different MoA (dsDNA damage, oxidative stress, etc), sensitivity/specificity (within each assay), and regulatory standards should have their biological endpoints weighted differently based on such findings. As no formal system has been established for the iScreen assay, an approach derived from the Fortin et al., (2023) paper performed by Health Canada will be indicted. Although the system used was not specifically referenced in the paper, the approach was able to be distinguished and applied. Therefore, a ranking system based on relative endpoints significance within each assay was applied, focused on weighting endpoints relative to each assay instead of individual biomarkers themselves. The weights and percentages applied to each reporter in ToxPi are as follows:

- MN - 3-fold (15.8%)
- CENPA - 2-fold (10.5%)
- H2AX, MPM2 & H3 - 1-fold (5.3%)

Although, as a weighting system is not formally established for most assays, this would mark a significant starting point for how weighting could be applied to endpoints across assays using ToxPi in the future as an important tool for such approaches toward genotoxic compound assessment. Figure 4.1.1 further demonstrates this hierarchy of endpoint significance through the initial identification of MN positive cells for genotoxic/non-genotoxic compound characterisation, then observations of CENPA presence/absence within the micronuclei, and lastly, γH2AX, MPM2 & H3 endpoints for specific genotoxic MoA identification. Finally, the BMD CIs are produced by generating the ToxPi profiles using the "compute bootstrap CI" option, where ToxPi ranks give an indication of relative potency, the ToxPi profiles are colour coordinated (figure 4.3.2.1) and the Mcquitty algorithm for hierarchical clustering is selected.

4.3 Results

This chapter covers the results produced by both PROAST and ToxPi, following the protocols addressed in the methods section for visualised outputs of the iScreen assay to outline the advantages of using ToxPi over traditional BMD analysis for dataset dimensionality reduction and observe whether iScreen could discriminate between

compounds based on their collective MoA using a list of 10 known aneugens/clastogens. Furthermore, the weighting system implemented across various endpoints is also observed for its ability to appropriately set endpoint weight based on their mechanistic properties, sensitivity, specificity, regulatory standards and overall biological relevance.

.

The assessment of dose-response relationships is a cornerstone in toxicological evaluation, providing critical insights into the potency and MoA of compounds. For this purpose, BMD modelling was performed using the R GUI version of PROAST to model benchmark dose (BMD) values for a series of compounds, utilising continuous individual data and selecting exponential models 3 or 5 from various families of models. The concentration served as the independent variable, while the response of specific endpoints was the dependent variable. A CES of ±0.5 above the control was applied, supported by 200 bootstrap runs to generate 90% BMD CIs. These analyses produced BMD plots for each compound, where log10(dose) and log10(response) were represented on the x- and y-axes, respectively, as exemplified in Figure 2.2.5.2.1. Subsequently, endpoint-specific BMD CIs were aggregated into a single visual output to facilitate compound comparisons, followed by further transformation for visualization in ToxPi. This methodological framework provides a robust basis for evaluating compound-specific potency and MoA, especially with biomarkers such as MN, CENPA, γH2AX, MPM2, and H3. Below, we present a table with definitions of all reporters alongside the detailed results for each compound, highlighting the endpoints that successfully produced BMD CIs and those that did not, along with interpretations of the biological implications.

Table 4.3.1 – iScreen Endpoint Nomenclature and Biological Interpretation.

iScreen Endpoint	Duration	Context/Cell population	Biological relevance
MN	24hr	All cells	Frequency of MN (general genotoxicity).
H2AXposMN	24hr	MN-positive cells	MN positive cells containing fragments from dsDNA damage signal.
H2AXapo	4/24hr	Apoptotic cells	The dsDNA damage signalling occurring in cells undergoing apoptosis.
nuclearH2AX	4/24hr	Main nucleus (all cells)	The dsDNA damage response (double-strand breaks) in the primary genome.
MPM2pos	4/24hr	Mitotic cells	Marker of mitotic phosphorylation. Indicates cells in mitosis (M-phase).
H3pos	4hr	Mitotic cells	Marker of mitotic entry and progression.
MNposCENPA	24hr	MN-positive cells	MN containing centromeres (whole chromosomes).
MNnegCENPA	24hr	MN-positive cells	MN lacking centromeres (i.e., the MN contains acentric fragments).
MNposH2AX	24hr	MN-positive cells	H2AX mediated dsDNA damage in the nucleus of cells that have formed MN.
MNnegH2AX	24hr	MN-negative cells	The dsDNA damage in the nucleus of cells that have not formed micronuclei.

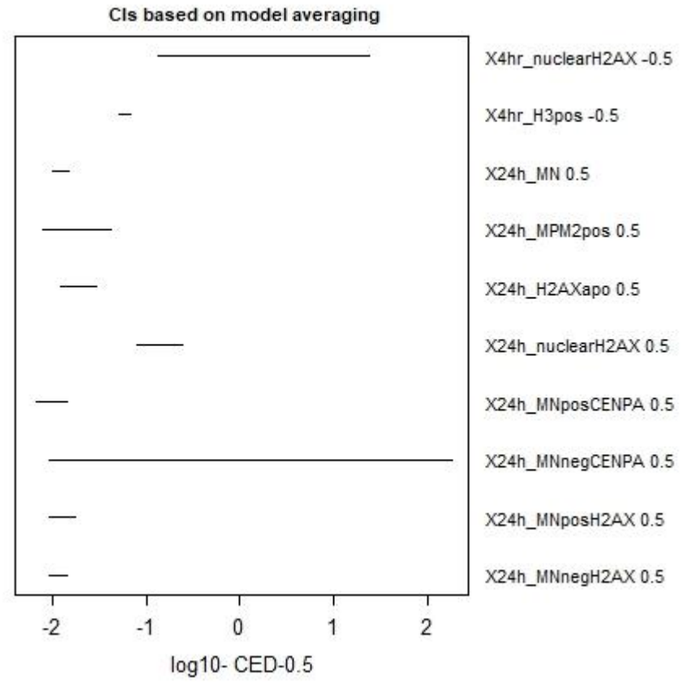


Figure 4.3.1.1 – Results showing PROAST produced and plotted 90% BMD CIs for AMG900 across 10/13 endpoints that successfully produced BMD CIs with a CES set to ±0.5 and 200 bootstrap runs plotted on a Log10(BMD) x-axis with 4/24-hour endpoints on the y-axis. These endpoints cover MN identification, CENPA presence/absence in cells and γH2AX, MPM2 & H3 biomarker responses. The specific endpoints that did not produce any BMDs include: 24hr H2AXposMN, 4hr H2AXapo, and 4hr MPM2pos. These imply a lack of response in endpoints covering dsDNA damage (clastogens) and mitotic cell activity (tubulin binding aneugens), correlating with the knowledge that AMG900 is an aurora kinase inhibiting aneugen, despite some response produced by a series of γH2aX biomarkers.

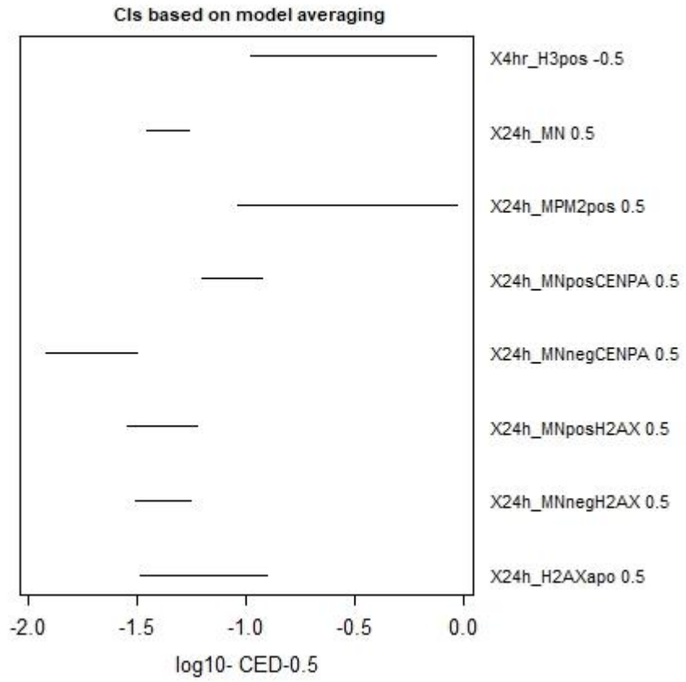


Figure 4.3.1.2 – Results showing PROAST produced and plotted 90% BMD CIs for baraserntib across 8/13 endpoints that successfully produced BMD CIs with a CES set to ±0.5 and 200 bootstrap runs plotted on a Log10(BMD) x-axis with 4/24-hour endpoints on the y-axis. These endpoints cover MN identification, CENPA presence/absence in cells and γH2AX, MPM2 & H3 biomarker responses. The specific endpoints that did not produce any BMDs include: 24hr H2AXposMN, 4hr H2AXapo, 4/24h_nuclearH2AX, and 4hr MPM2pos implying a lack of response in dsDNA breaks (dsDNA) and mitotic cell activity (tubulin binding aneugens) biomarkers, indicative of the aneugens aurora kinase inhibiting MoA that baraserntib possesses.

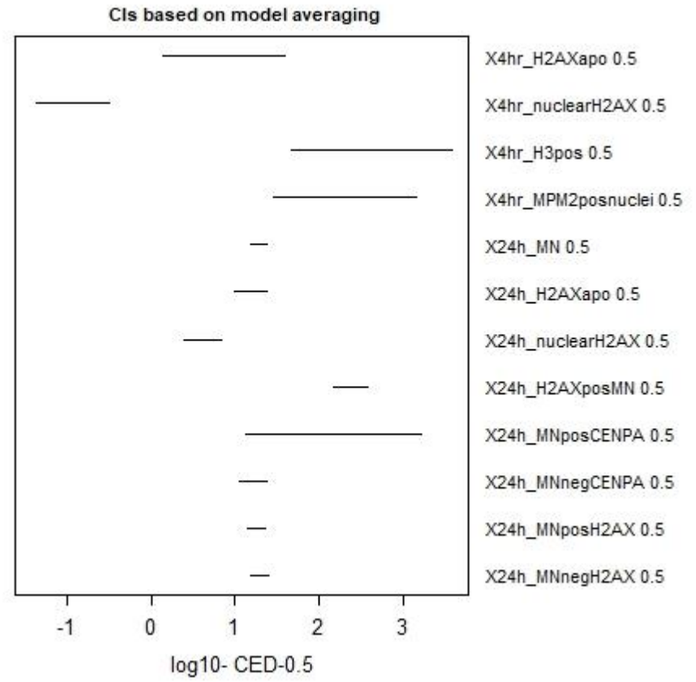


Figure 4.3.1.3 – Results showing PROAST produced and plotted 90% BMD CIs for cisplatin across 12/13 endpoints that successfully produced BMD CIs with a CES set to ±0.5 and 200 bootstrap runs plotted on a Log10(BMD) x-axis with 4/24-hour endpoints on the y-axis. These endpoints cover MN identification, CENPA presence/absence in cells and γH2AX, MPM2 & H3 biomarker responses. The specific 24hr MPM2pos endpoint did not produce a BMD implying a lack of response in mitotic cell activity (tubulin binding aneugens) biomarkers, which is to be expected as cisplatin has a clastogenic MoA. However, despite the unexpected activated for endpoints p-H3/MPM2/CENPA+, they demonstrated larger BMD CIs and higher BMD values implying lower potencies across those endpoints, meanwhile γH2AX related endpoints see narrow CIs with lower BMD figures suggesting higher activation/potency at those endpoints.

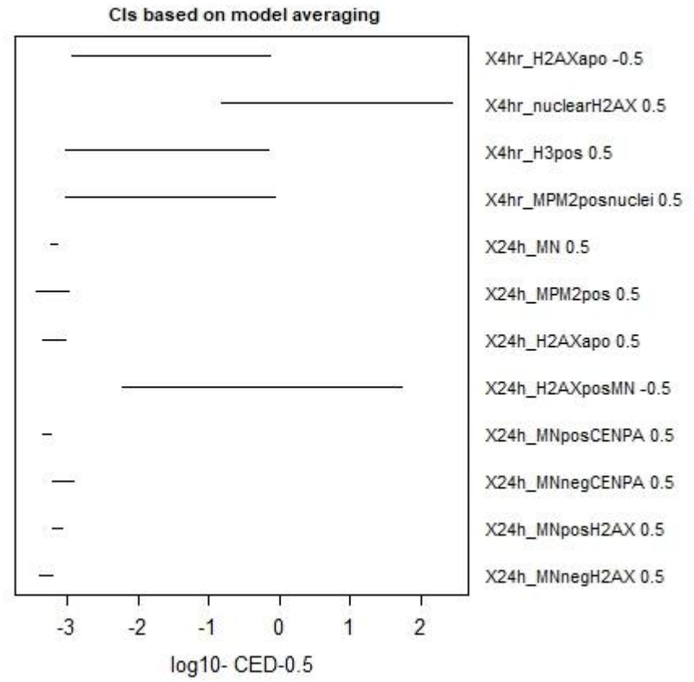


Figure 4.3.1.4 – Results showing PROAST produced and plotted 90% BMD CIs for colchicine across 12/13 endpoints that successfully produced BMD CIs with a CES set to ±0.5 and 200 bootstrap runs plotted on a Log10(BMD) x-axis with 4/24-hour endpoints on the y-axis. These endpoints cover MN identification, CENPA presence/absence in cells and γH2AX, MPM2 & H3 biomarker responses. The 24h_nuclearH2AX endpoint did not produce a BMDs, suggesting that a lower response from γH2AX biomarker within the nuclei, correlating to its primary MoA as a tubulin binding aneugen, however, a response across other γH2AX endpoints remains unexpected. The lower BMD values across most endpoints further imply a higher potency for colchicine.

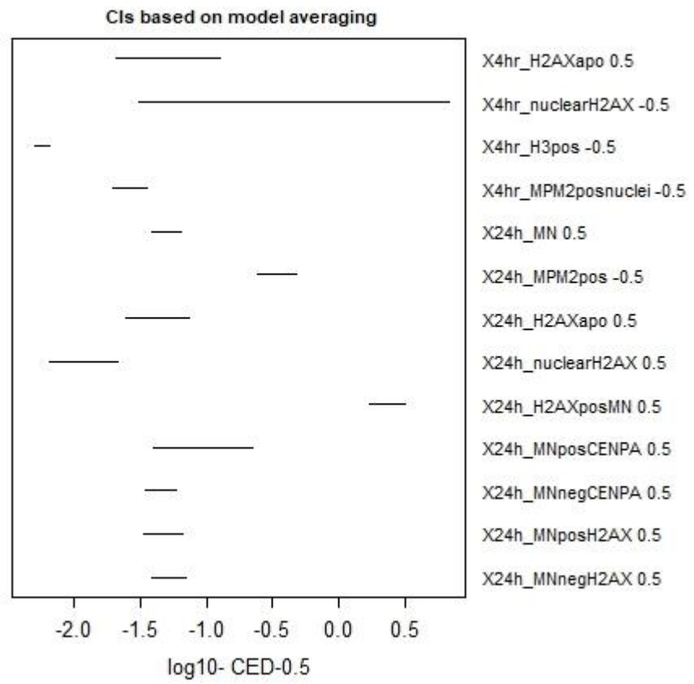


Figure 4.3.1.5 – Results showing PROAST produced and plotted 90% BMD CIs for dasatinib across all 13 endpoints that successfully produced BMD CIs with a CES set to ±0.5 and 200 bootstrap runs plotted on a Log10(BMD) x-axis with 4/24-hour endpoints on the y-axis. These endpoints cover MN identification, CENPA presence/absence in cells and γH2AX, MPM2 & H3 biomarker responses. As dasatinib clastogenic MoA activates the γH2AX endpoint as expected, it further unexpectedly triggers responses from CENPA, MPM2 & H3 endpoints, likely due to external factors outside of dasatanib's primary MoA activating endpoints such as high potency inducing generalised cell stress producing unrelated endpoint responses. Furthermore, the nuclearH2AX endpoint has a wider CI suggesting more uncertainty surrounding the γH2AX endpoint; a key endpoint for clastogenic MoA.

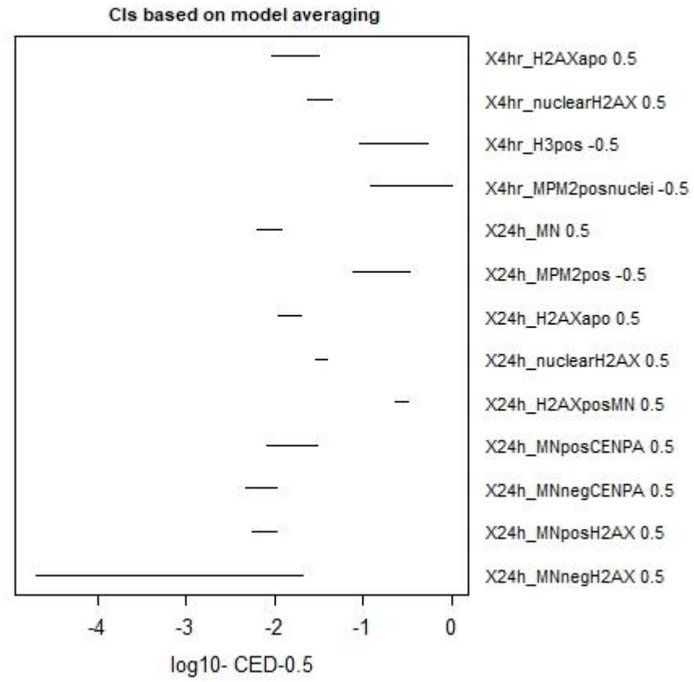


Figure 4.3.1.6 – Results showing PROAST produced and plotted 90% BMD CIs for ETO across all 13 endpoints that successfully produced BMD CIs with a CES set to ±0.5 and 200 bootstrap runs plotted on a Log10(BMD) x-axis with 4/24-hour endpoints on the y-axis. These endpoints cover MN identification, CENPA presence/absence in cells and γH2AX, MPM2 & H3 biomarker responses. ETO has a known clastogenic MoA activates the γH2AX endpoint as expected, despite the activation of CENPA, MPM2 & H3 endpoints to lesser extents.

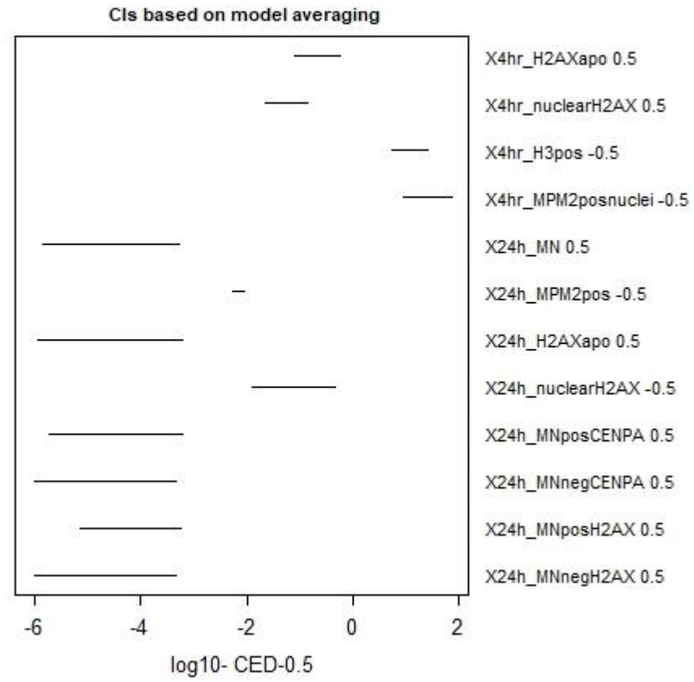


Figure 4.3.1.7 – Results showing PROAST produced and plotted 90% BMD CIs for MMC across 12/13 endpoints that successfully produced BMD CIs with a CES set to ±0.5 and 200 bootstrap runs plotted on a Log10(BMD) x-axis with 4/24-hour endpoints on the y-axis. These endpoints cover MN identification, CENPA presence/absence in cells and γH2AX, MPM2 & H3 biomarker responses. The missing endpoint is unexpectedly 24h_H2AXposMN, despite the primary clastogenic MoA of MMC, which may be down to several biological/experimental issues such as compound stability or cell line insensitivity to that endpoint.

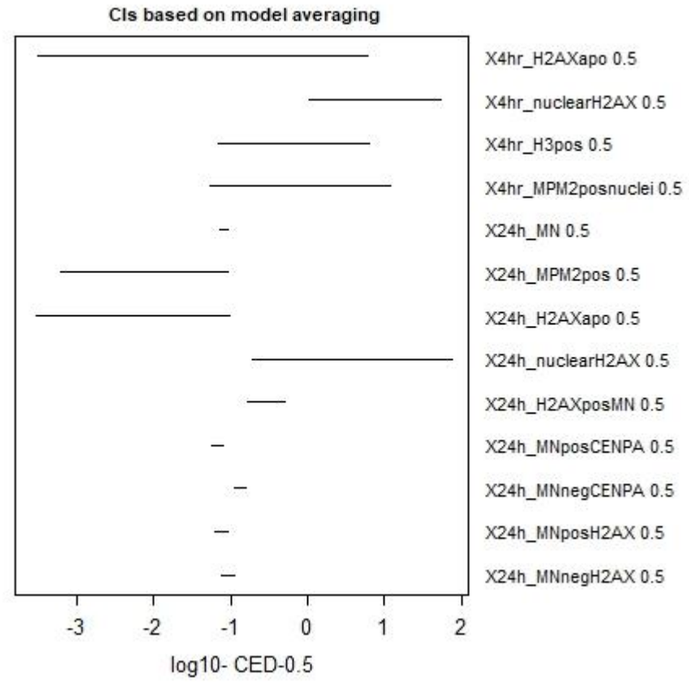


Figure 4.3.1.8 – Results showing PROAST produced and plotted 90% BMD CIs for nocodazole across all 13 endpoints that successfully produced BMD CIs with a CES set to ±0.5 and 200 bootstrap runs plotted on a Log10(BMD) x-axis with 4/24-hour endpoints on the y-axis. These endpoints cover MN identification, CENPA presence/absence in cells and γH2AX, MPM2 & H3 biomarker responses. It activates key endpoints for its primary aneugenic tubulin binding MoA, such as CENPA, MPM2 and H3, however, it appears to also activate other unrelated endpoints, such as γH2AX.

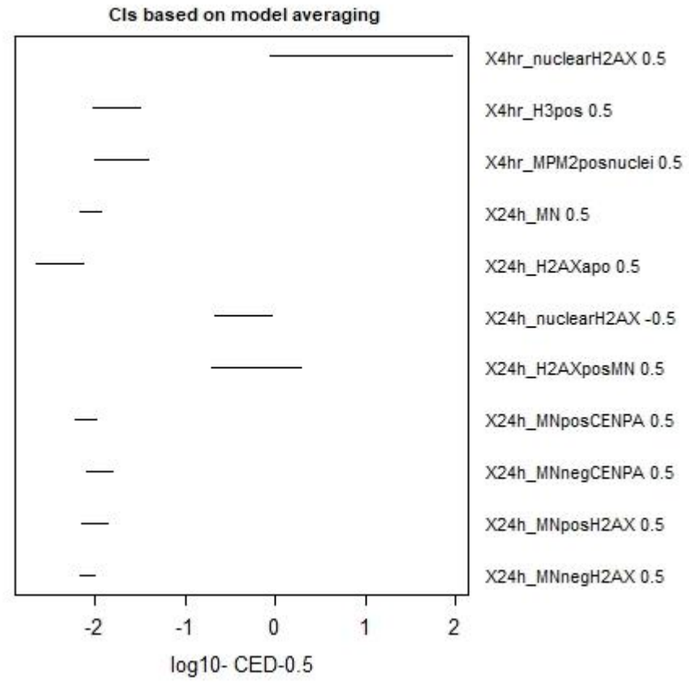


Figure 4.3.1.9 – Results showing PROAST produced and plotted 90% BMD CIs for taxol across 11/13 endpoints that successfully produced BMD CIs with a CES set to ±0.5 and 200 bootstrap runs plotted on a Log10(BMD) x-axis with 4/24-hour endpoints on the y-axis. These endpoints cover MN identification, CENPA presence/absence in cells and γH2AX, MPM2 & H3 biomarker responses. Taxol has a known tubulin binding aneugen and appears to mostly activates endpoints related to its MoA (CENPA, H3 and MPM2), however, other endpoints such as γH2AX are also somewhat activated. Furthermore, 4h_H2AXapo endpoint produced no BMD, however, unexpectedly, the 24h_MPM2pos produced no BMD, which is inconsistent with its MoA.

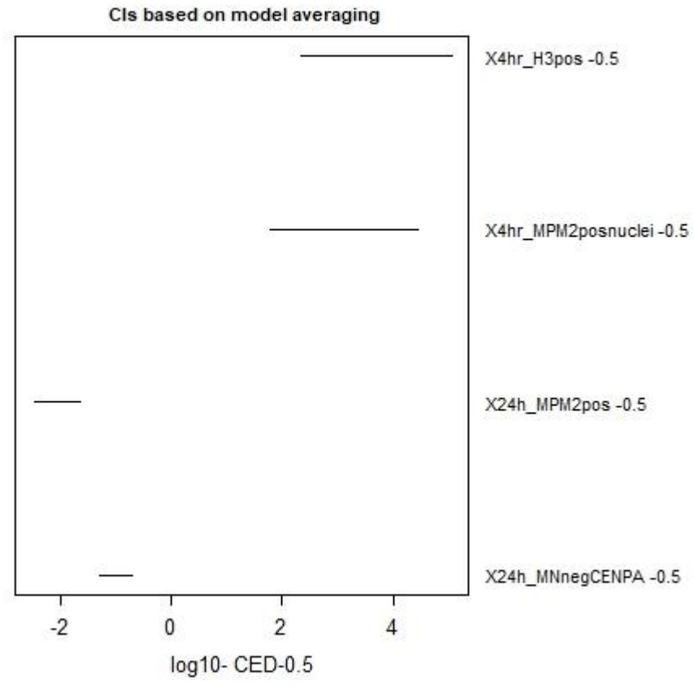


Figure 4.3.1.10 – Results showing PROAST produced and plotted 90% BMD CIs for tunicamycin across 4/13 endpoints that successfully produced BMD CIs with a CES set to ±0.5 and 200 bootstrap runs plotted on a Log10(BMD) x-axis with 4/24-hour endpoints on the y-axis. These endpoints cover MN identification, CENPA presence/absence in cells and γH2AX, MPM2 & H3 biomarker responses. Tunicamycin is a non-genotoxic compound and appears to activate 24h_MNnegCENPA, 24h_MPM2pos, 4h_MPM2posnuclei and 4h_H3pos endpoints, however, as no BMD is produced for the 24_hr MN endpoint, it is assumed that no micronuclei has formed, suggesting that no chromosomal damage/mis-segregation events had occurred within the cell samples and any BMD readings produced are likely due to certain types of DNA mutations/damage that does not result in major alterations to the chromosomes. Hence, tunicamycin may induce some degree of DNA damage but not any that induced numerical/structural chromosomal damage.

The BMD modelling results, as illustrated in Figures 4.3.1.1 to 4.3.1.10, provide a comprehensive overview of the potency and mechanistic activity of each compound across multiple endpoints. The analysis revealed distinct patterns of activation and inhibition, consistent with the known MoAs of the compounds. For instance, aurora kinase inhibitors such as AMG900 and barasertib demonstrated responses consistent with their aneugenic activity, while clastogens like cisplatin and etoposide (ETO) activated γH2AX-related endpoints as expected. However, unexpected activations in endpoints unrelated to the primary MoA were observed for some compounds, such as dasatinib and nocodazole, suggesting potential off-target effects or generalized cellular stress responses. Together, these findings underscore the utility of BMD modelling for high-resolution analysis of compound-specific toxicological profiles and their MoA. Further analysis using tools like ToxPi will enable deeper insights into the relative potency and toxicity profiles of these compounds, aiding in potency/hazard assessment and mechanistic toxicology studies.

The integration of ToxPi analysis into the dose-response assessment framework offers a multidimensional visualization of biomarker-specific toxicological profiles. By incorporating endpoint weightings, compound rankings, and clustering methodologies, ToxPi enables an enriched understanding of the relationships between biomarkers, mechanisms of action (MoA), and compound potency. This section details the development of ToxPi profiles for the 10 tested compounds. Endpoint weightings were derived from fold changes (Section 4.2.4) to ensure biologically relevant scaling. Compounds were ranked based on genotoxic potency, followed by hierarchical and PCA-based clustering to delineate groupings based on MoA and biomarker activation patterns. Together, these analyses provide a systematic framework for interpreting BMD data in a visually accessible manner while retaining critical statistical rigor. These analyses below reveal distinct groupings of compounds based on their MoA and biomarker activation patterns, providing insights into their genotoxic potential and biological activity.

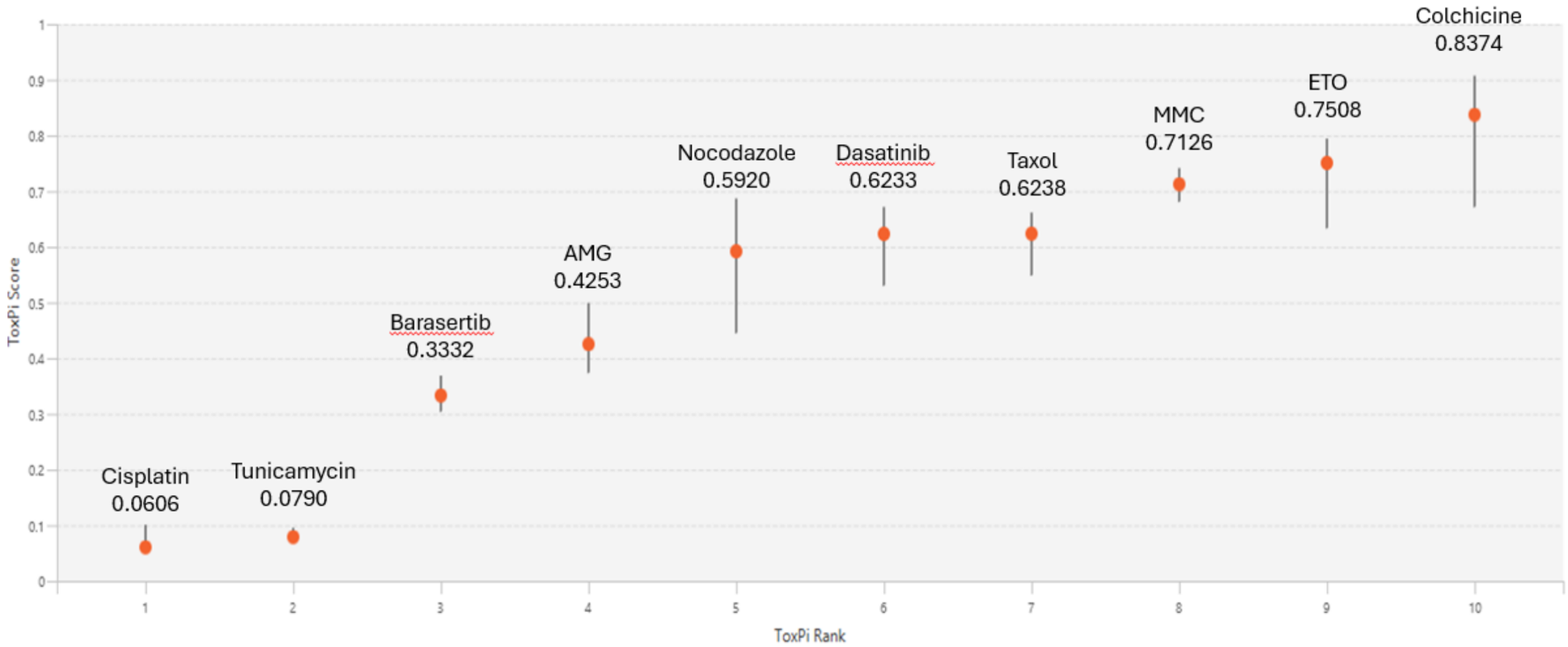


Figure 4.3.2.1 – ToxPi ranking of compounds based on relative genotoxic potency in the iScreen assay (1 = lowest potency, 10 = highest potency). Scores range from 0-1, with higher scores indicating greater potency in this assay. Error bars represent the confidence interval. The ranking reflects performance across the specific endpoint panel used here; for instance, cisplatin scored lowest despite its known human toxicity, highlighting that this ranking is assay-specific and not a measure of overall hazard.

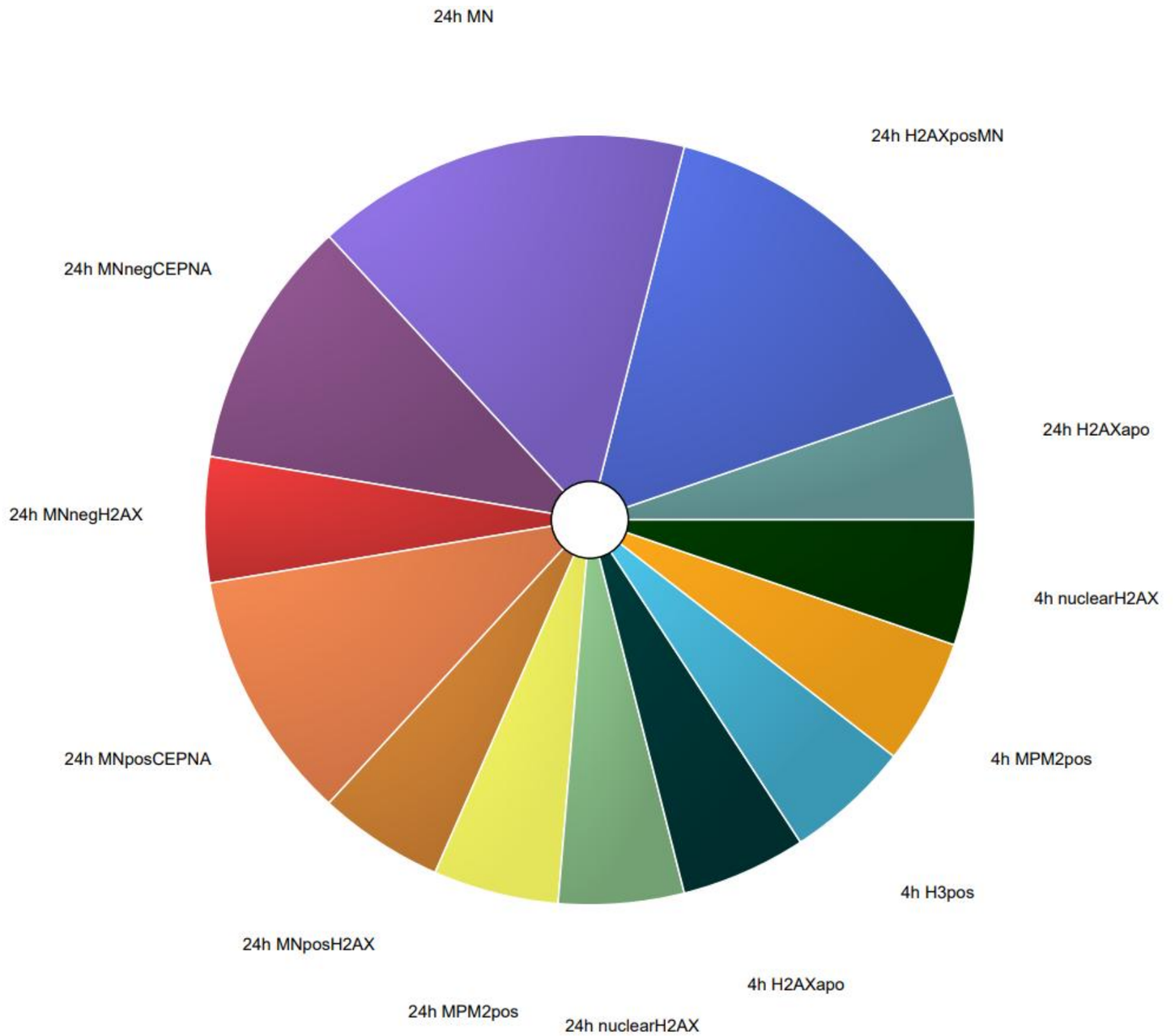


Figure 4.3.2.2 – ToxPi profile key illustrating each biomarker examined in the iScreen assay as an individual slice with corresponding colours. Weightings reflect biological significance as described in section 4.2.4: the MN endpoint received the highest weighting at 3-fold (15.8%) as the primary genotoxicity marker representing the physical manifestation of chromosomal damage or whole chromosome loss. CENPA endpoints received a 2-fold weight (10.5%) , reflecting their critical role in distinguishing aneugenic (centromere-positive) versus clastogenic (centromere-negative) mechanisms. A baseline weight of 1-fold (5.3%) was assigned to the remaining endpoints (γH2AX, MPM2, and H3) to enable specific mechanism of action (MoA) identification while acknowledging their supporting role relative to the core genotoxicity endpoints.

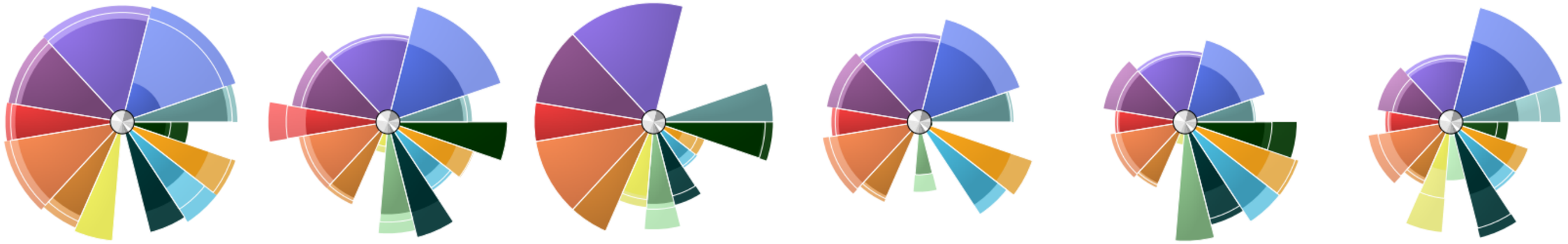


Figure 4.3.2.3 – ToxPi profiles (with bootstrap CI) generated from BMD data for each chemical, showing the biomarkers and their ToxPi scoring for each end point. Each slice represents a biomarker at a given time. Shading in the centre of the circle represents uncertainty for each slice (darker = more uncertain). The radius of each slice (from the centre of each circle to the either the edge or white line towards the edge of the slice) represents the potency of each biomarker and the widths the user defined weightings. The brighter regions presented in each is the bootstrap CI where outer edge shows the upper CI, and the inner edge is the lower CI for each slice.

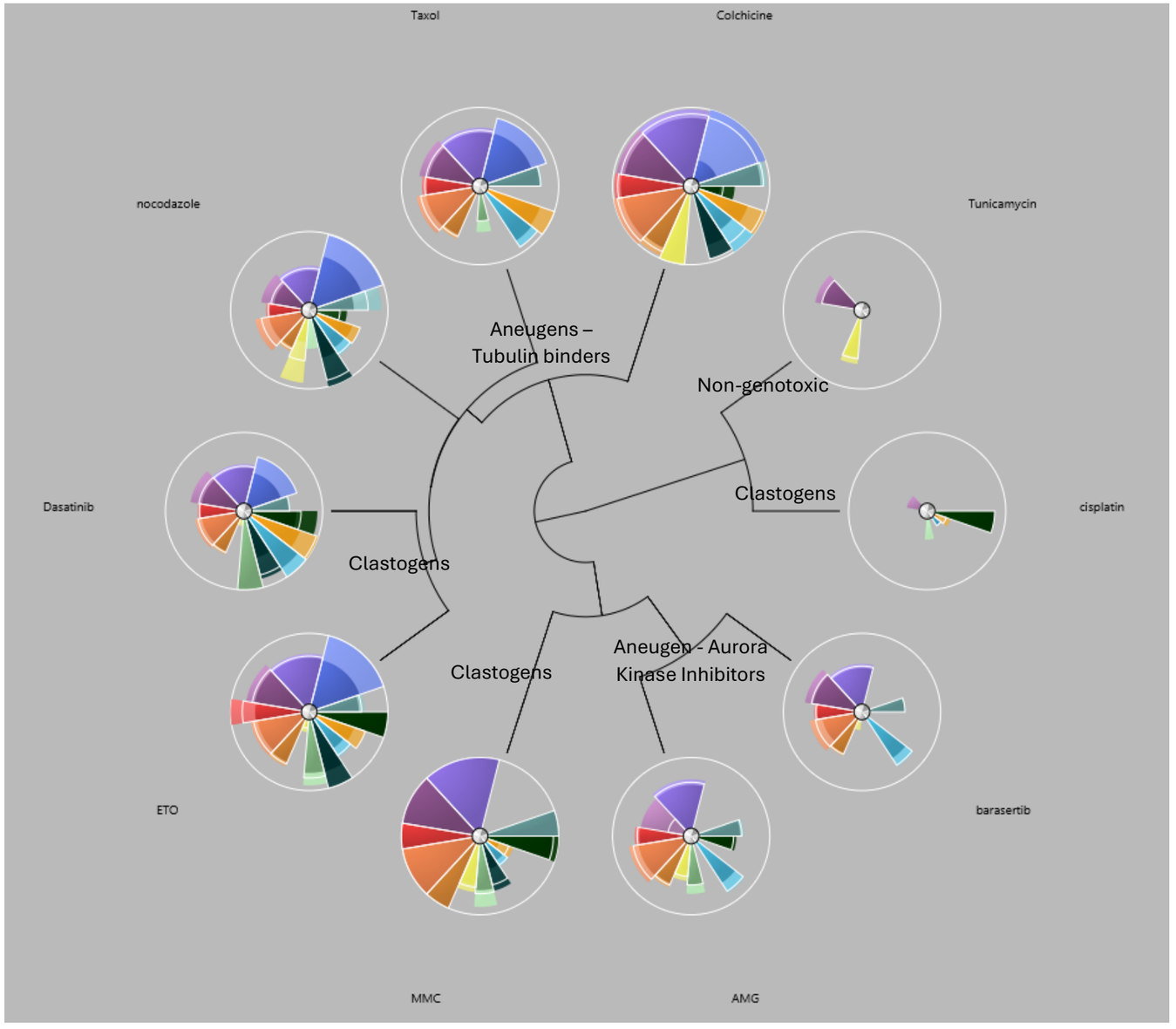


Figure 4.3.2.4 – Mcquitty hierarchical clustering of ToxPi profiles was performed using complete AGNES clustering to identify chemical groupings, each color-coded based on their biomarker activation and MoA. ToxPi separated the compounds into their known subgroupings: clastogens, weak clastogens, aneugenic tubulin binders, aneugenic aurora kinase inhibitors and non-genotoxicants.

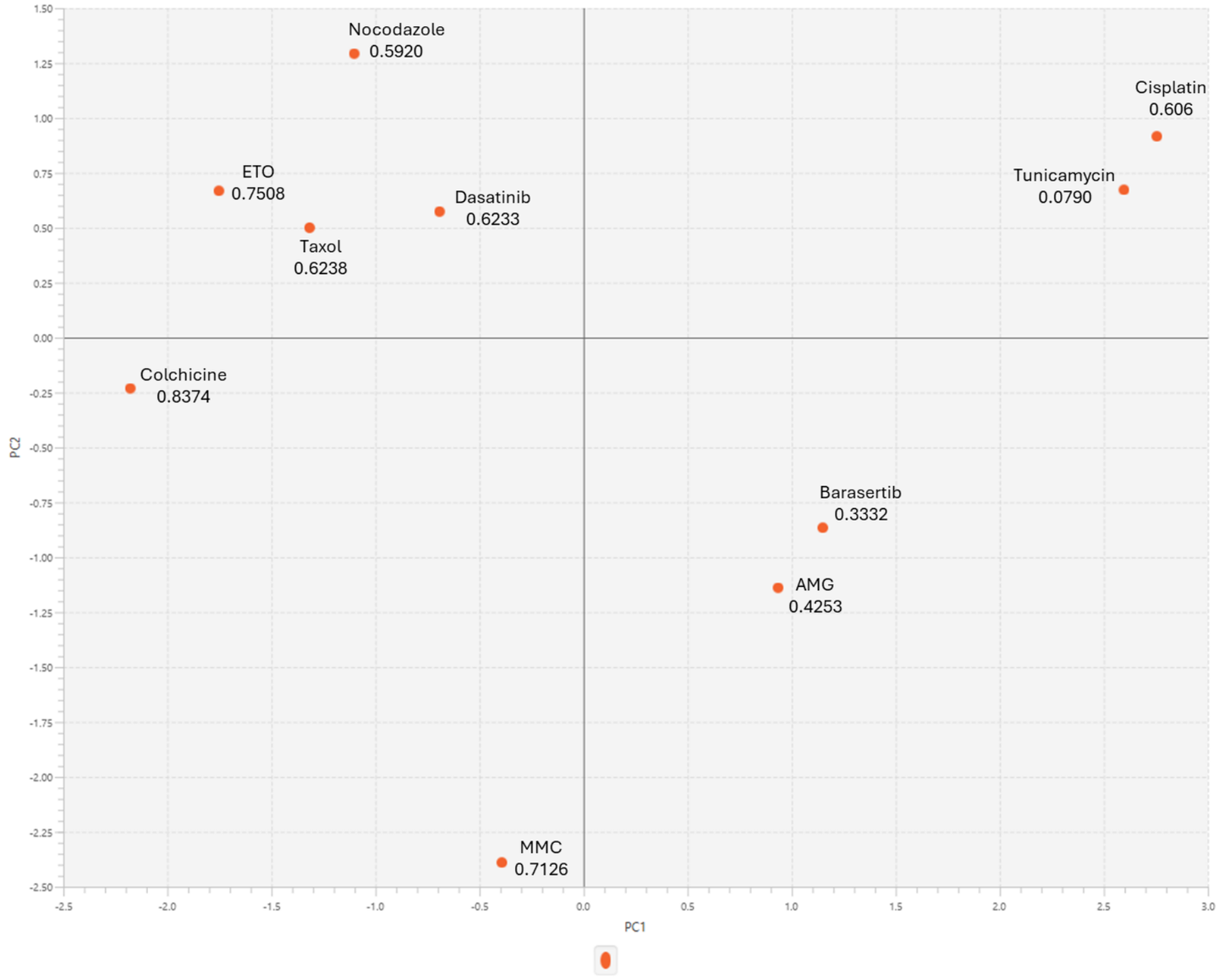


Figure 4.3.2.5 – ToxPi K-means divisive/PCA clustering where each compound is grouped based on variation across the ToxPi scores/genotoxicity and biomarker sensitivity. Unlike previous chapters, as a fixed cluster number was not set, the compounds were instead manually observed, and compounds were grouped based on PC1 and PC2 analyses. For this analysis, compounds with higher potencies had more negative PC1 values and compounds clustered within proximity to each other similarly to the hierarchical clustering performed in figure 4.3.2.4 explaining why MMC is significantly distanced from other compounds.

The ToxPi results presented in figures 4.3.2.1 – 4.3.2.5 offered a mutli-dimensional understanding of genotoxicity across the 10 compounds analysed by presenting their potency rankings, their biomarker activation profiles and MoA clustering with biomarker weight application considerations. Figures 4.3.2.1 displays a key which shows the ToxPi rankings, showing which compounds, across the selected iScreen biomarkers, has the highest to lowest relative potencies and figures 4.3.2.2 – 4.3.2.3 displays the allocation and degree of biomarker activation demonstrated by each compound. Additionally, figure 4.3.2.4 displays the AGNES hierarchical clustering of the 10 analysed compounds, separating them out by their general MoAs, predominantly into four categories, including clastogens, aneugens and non-genotoxicants with cisplatin being an exception which clustered with tunicamycin likely due to its lowered biomarker activation and overall ToxPi scoring. Hence, only cisplatin (clastogens) clustered incorrectly with tunicamycin (non-genotoxicant) likely due to lower overall ToxPi scores and response/potency across the selected endpoints. Meanwhile, figure 4.3.2.5 displays the K-means clustering revealed distinct groupings based on genotoxicity and biomarker sensitivity, with compounds exhibiting higher potencies clustering together and outliers like MMC being significantly distanced from others. The discussion section below further addresses these findings to rationalise the meaning behind each of the figures and refine our understanding of the biological/statistical conditions and circumstances associated with these findings across the compounds analysed.

4.4 Discussion

This section discusses the outputs surrounding the novel iScreen assay figures, particularly around the generation of BMD CI graphs using PROAST to analyse 13 distinct endpoints to distinguish between the varying MoA for each 10 compounds with known clastogenicity, aneugenicity (tubulin binding/aurora kinase inhibiting), and non-genotoxic properties. Furthermore, this section alludes to ToxPi's ability to comparatively assess the same subset of compounds using their unique ranking, profiling and clustering algorithms to assess whether the implementation of it provides additional figures to further complement the existing PROAST outputs providing additional mechanistic/potency information to assist in the classification of these compounds through dataset dimensionality reduction. Overall, the combination of these techniques attempted to produce overall analysis profile of each of the 10 compounds for comparative assessment purposes using iScreen assay figures/outputs.

4.4.1 BMD CI charts

The BMD provides an excellent normalisation approach towards dose-response data when regarding the iScreen assay and PROAST was successfully capable of depicting BMD CIs regions across multiple endpoints for multiple compounds analysed in this investigation. Figures 4.3.1.1 – 4.3.1.10 clearly displays each biomarker within at 4/24-hours and allows for any significant/insignificant differences between endpoints to be distinguished from overlapping metrics. It can accurately depict 90% CI BMD potency range at which a compound, marking each compound's MoS using its true original experimental detection metrics. Despite this, it is incapable of comparatively analysing both endpoints and compounds simultaneously, a limitation resulting in compounds requiring superimposing multiple compounds assessment metrics to be overlapped for comparison, which would be challenging as the ranges of where each compound exhibits a BMD result may vary between the compounds, potentially introducing some degree of confusion when analysis must be drawn. Furthermore, it does not allow for comparison across compounds to be clearly depicted and for dataset dimensionality to be reduced, limiting its analytical capacity.

Other issues may involve endpoints across multiple compounds that are expected to activate not providing a dose-response curve, meaning no BMD CIs are generated. One reason could be that some compounds could potentially require biotransformation to form highly reactive metabolites which contribute toward chemically induced genotoxicity (Attia., 2010). However, for this chapter, the compounds tested in the iScreen assay did not require metabolic activation to exhibit their genotoxic effects and require no further metabolic activation as it brought no significant difference to their potencies. A limitation of the iScreen assay is that it detects clastogens and aneugens utilising MN formation, however, does not detect pure mutagens that induce point mutations without causing chromosomal damage. Such detection techniques require complementary assays for detection (e.g. Ames test) to detect point mutations linked to specific genetic damage events. Furthermore, the incorporation of additional endpoints within the iScreen assay could aid in extracting as much genotoxic potency/mechanistic information as possible to compensate for these detection deficiencies with non-specific toxicity/off target effects interfering with known pathways being additional considered (for example using toxtracker assay biomarkers blvrb/srxn1 to detect oxidative stress). Finally, a limitation of this approach is that BMD modelling is sensitive to data quality. Compounds that produced high variability between replicates or

exhibited poor curve fit were excluded, as their BMD confidence intervals would be too wide for meaningful interpretation and would also increase CI widths if the background exhibited excess variation.

Despite this, in a few instances some compounds produced heightened responses at unexpected endpoints. For these results, non-specific toxicity from generalised cell stress may result in activation of unexpected biomarkers, despite the mitigation provided through the 60% cytotoxicity limits imposed on the original dataset. These compounds may also interact with non-DNA targets, kinases, etc leading to signal transduction changes that activate pathways unrelated to DNA damage but lead to detection within the assay. Additionally, despite that all compounds did not have oxidative stress as their primary MoA, they may induce further ROS production either indirectly or as a secondary effect from their primary MoA, resulting in genotoxicity due to the contributed oxidative stress these compounds cause (Di Ianni et al., 2022, Møller et al., 2014). Finally, unlike ToxPi, all endpoints could not have additional weightings parameters placed on them and no manipulation of the original CSV dataset would allow for this to currently be feasible. Therefore, the lack of weighting application could bring the significance of these results into questions as, despite the success in mapping individual figures that remain true to the integrity of the original BMD CIs generated through the use of raw dose-response information, the lack of weighting parameters means that all biomarkers are treated as equals in terms of significance, which would not constitute the biological reality of how these endpoints are marked based on their comparative relevance. Therefore, additional analysis software platforms, such as ToxPi, would require implementation to further apply theses weightings and ensure these shortcomings are addressed.

4.4.2 ToxPi outputs

Once successful BMDs were generated using PROAST, their outputs CI values were placed into a CSV file and imported into ToxPi, where the final figures 4.3.2.1 – 4.3.2.5 produced for each endpoint, once the necessary transformations as described in 4.2.4 were made to the PROAST dataset BMD outputs and each CI bound is imported into a CSV file and placed into each individual slice. Once the $-\log_{10}(x)$ transformation was successful applied of across all endpoints, weighting was applied to each endpoint according to section 4.2.4, and unlike

PROAST, successful generations of output figures were generated that prioritised the biological relevance of each endpoint regarding genotoxicity.

The 3-fold (15.8%) weighting was applied to the MN endpoint as the chromosomal damage biomarker to highlight whether exposure to a compound would induce MN development and overall lead to genotoxicity to some degree. Although biomarkers may induce genetic alterations, the manifestation of MN remains a key factor in determining whether this has caused structural or numerical chromosomal damage. This allows clastogens and aneugens (MN+) to be distinguished from compounds that do not produce such chromosomal effects (MN-), though MN- compounds may still be genotoxic via other mechanisms (e.g., point mutations). The CENPA biomarkers had 2-fold (10.5%) application of weighting to consider the presence of centromere regions within the micronuclei for the detection of chromosomal malsegregation and aneuploidy, making it as a key endpoint for the detection of clastogens (CENPA-) to be distinguished from aneugens (CENPA+). The remaining endpoints (γH2AX, MPM2 & H3) all have 1-fold (5.3%) weightings applied as they are all confirmatory biomarkers that further support the MN and CENPA findings, confirming the presence of clastogens (γH2AX) and tubulin binding (MPM2+ & H3+)/aurora kinase inhibiting (H3- & polyploidy+) aneugens. The endpoint fold change order was determined by the original Sun et al., (2022) paper, with the specific figures selected following the logic applied by Health Canada in their Fortin et al., (2023) paper, however, as no formal weighting figures exist for these specific endpoints, a simplified approach was adopted: MN-primary endpoints were weighted 3-fold, CENPA endpoints 2-fold, and all others 1-fold. This integer scale was required by ToxPi software and reflects relative priority (3 = highest) for this proof-of-concept validation. Despite this, any numerical values will always be heavily scrutinised as exact numerical values are difficult to distinguish and require much more exploration, however the weighting system applied appears to be a significant step in allowing for future work to assist in the development of a precise numerical system that could be applied for the iScreen and other assays. The potency of compounds for ToxPi ranking is influenced by the selection of MoA endpoints (which are chosen based on the specific MoA targeted within each assay), the compounds included, the weighting of endpoints, experimental data, and the mathematical framework behind the ToxPi score calculation.

Additionally, unexpected higher/lower endpoint projections and undetectable MoAs may contribute toward a decrease of overall ToxPi scoring for compounds and endpoint projections as they may have specific MoAs that are not covered by the selected assay. For

example, cisplatin is a known clastogenic compound with atypical alkylating/cross-linking properties which is expected to have increased induction for the γH2AX endpoint. However, as all other endpoints (except MN) would have lower projections, this is likely to contribute to the lower overall ToxPi ranking, making it appear as if it has the lowest potency despite its significantly higher “4hr nuclear H2AX” projection (Kirkland et al., 2016). Cisplatin primarily produces DNA lesions through N7 guanine/adenine adducts and crosslinks, mechanisms not directly measured by the iScreen assay, which detects yH2AX as a general DNA damage reporter and MN chromosomal damage. Therefore, these specific lesions may require DNA replication or repair to manifest as detectable double-strand breaks or micronuclei, potentially explaining the lower biomarker activation observed. Therefore, cisplatin's reduced detection reflects endpoint specificity rather than an absence of clastogenic activity.

Detection issues, as described in previous chapters, could further produce lower projections within the compound itself. The iScreen method relies on endpoints like MN formation, γH2AX for DNA double-strand breaks, and centromere protein staining to differentiate genotoxic mechanisms. While cisplatin is a known clastogen, it induces DNA cross-linking rather than the extensive fragmentation associated with some other clastogenic agents with more pronounced γH2AX responses or greater MN formation. Additionally, aneugenic compounds require the completion of mitosis prior to MN generation, hence why assay timing will impact potency and 24hr treatments would be required to capture both clastogenic and aneugenic effects, although some aneugens require longer expression times to produce greater MN induction. However, utilising targeted detection techniques or combining different assays with expanded endpoints and targeted detection techniques could mitigate these limitations, as demonstrated in Chapter 3, where ToxTracker endpoints with detection issues for compounds were able to mitigate these concerns. ToxPi’s operation is also inherently relative where this additional incorporation of additional endpoints would alter the ToxPi ranking, profiling and clustering outcomes.

4.4.3 – improvements

To resolve the ambiguous clustering of Mitomycin C and sharpen the distinction between MoA classes, the current dataset would benefit from a strategic expansion and a refinement of the analytical lens. The key is not merely to add more compounds, but to introduce specific "anchoring" agents that share MMC's precise mechanism of DNA cross-linking, such as additional cross-linkers like cisplatin, to complement the existing clastogens. This would

force the algorithm to recognise a distinct "cross-linker" sub-cluster within the broader clastogen family, pulling MMC away from aneugens or compounds with mixed response profiles. Simultaneously, incorporating a wider spectrum of mechanistically nuanced agents, such as topoisomerase II inhibitors (e.g. doxorubicin) to represent other clastogen subtypes and spindle poisons (e.g., paclitaxel) to solidify the aneugen cluster, would create more well-defined biological groupings. Crucially, this should be paired with a shift from standard Euclidean distance to a correlation-based distance metric, which focuses on the shape of the response profile rather than its magnitude. This ensures that MMC and other DNA-damaging agents cluster together based on their shared mechanistic patterns, even if their phenotypic potency varies. Separately, it should be noted that direct comparisons of potency rankings between aneugens and clastogens are complicated by experimental concentration ranges. Aneugens, such as colchicine, were tested at lower concentration ranges due to cytotoxicity limits, which can result in higher relative ToxPi scores that may overrepresent their potency compared to clastogens tested across wider ranges. Including additional compounds with matched concentration windows or applying normalization strategies would help address this in future analyses.

Additionally, certain MoAs may appear as more potent in this assay simply because they activate more biomarkers more strongly. This means the addition of endpoints that capture underrepresented mechanisms (e.g. crosslinking agents such as cisplatin) would help balance this bias, allowing for more equity regarding relative potency analysis across different MoA classes. Furthermore, additional prevision in weighting selection could provide additional precision and refinement of the ToxPi profiles and how each endpoint is accurately weighed according to the biological literature. However, to make such a system requires additional experimentation and a wider regulatory imperative to help steer such guidelines into a realistic direction that will maximise the degree of information extracted by legitimising the numbers chosen for each endpoint weight.

A fixed procedure surrounding the handling of iScreen data in processing is required to develop guidelines which accurately visualise the iScreen datasets in a controlled manner whilst retaining information that directly correlates to the experimental detection findings while simplifying the figures produced. Implementing such refinements would require extended iScreen experimentation with increased data collection. Protocol development would need to address experimental artifacts, unexpected biomarker responses, replicate integration, and criteria for non-responsive compounds. Additionally, the protocol must have

an established criteria for interpreting compounds that only demonstrate positive responses at high concentration with unclear dose-response relationships – a scenario that may indicate threshold effects, artifacts related to cytotoxicity or genuine genotoxicity requiring further investigation**.** Establishing such a standardised procedure and incorporating these enhancements will not only improve the robustness and accuracy of the iScreen analysis but also ensure its applicability in generating biologically meaningful and reproducible insights for genotoxicity assessment.

4.5 – Chapter conclusions

This chapter overall evaluated the iScreen high-content imaging platform *in vitro* genotoxicity screening approach and its ability to mechanistically distinguish and classify clastogens, aneugens and non-genotoxic compounds. The iScreen assay is an automated and advanced high-throughput imaging system that identifies specific biomarkers that detect chromosomal instability with heightened accuracy and reproducibly/assay throughput (Sun et al., 2022). Sensitive immunofluorescent labelling and endpoint specific imaging provided strong detection methodology and a nuanced insight into compound MoA for genotoxic/non-genotoxic compounds to be distinguished and comparatively analysed (Campbell et al., 2014: Sun et al., 2022). The integration of the MicroFlow and MultiFlow assays further enabled this chapter to demonstrate the potential of iScreen to provide aneugenicity screening – a critical gap that exists in early-drug discovery where the standard MN assays alone cannot differentiate anuegens from clastogens without additional testing. This platform's multiplexed high-content imaging abilities allowed for the assessment of multiple MoAs within dose-response datasets to assess DNA damage with superior sensitivity and efficiency compared to traditional genotoxicity assays. The MN assessment of the MicroFlow assay allowed for the screening of genotoxicants/non-genotoxicants and the incorporation of targeted antibody markers—such as γH2AX for double-strand breaks, CENPA for chromosomal fragments, and MPM2 for mitotic activity—allowed for the assay could readily distinguish between aneugenic and clastogenic MoA in a streamlined manner enabling viable high-volume screening options from the increased efficiency reducing analysis times.

A major advantage of the iScreen assay lies in its ability to classify compounds by MoA especially in differentiating aurora kinase inhibiting and tubulin binding aneugens and clastogens**.** By using endpoints such as MN formation and CENPA, the assay the assay not

only detects chromosomal damage, but it also distinguishes whether it results from aneuploidy (whole chromosome loss) or chromosomal fragmentation, offering an improvement over traditional MN assays that may lack unless further testing is performed. This enables rapid classification of genotoxic MoAs in a single assay, providing mechanistic insight into chromosomal alterations that would typically require additional follow-up assays to characterise This method also allowed for the quantification of these changes across varied concentrations and time points, adding robustness to the results and enhancing the dataset's depth for dose-response analysis.

To complement the iScreen imaging data produced, PROAST and ToxPi provided analytical tools which contributed to the multidimensional profiling of genotoxic potency. The dose-response analysis with PROAST produced benchmark dose (BMD) estimates, providing a quantitative basis for comparing potency across endpoints, enabling the normalisation of response data and the establishment of confidence intervals for BMD values, offering a statistically strong metric of genotoxic response across a diverse compound list. Additionally, ToxPi facilitated the visualization of genotoxic compounds into ToxPi rankings, profiles, and hierarchical/k-means clusters based on biomarker potency and MoA. Furthermore, the weighing of critical endpoints such as MN formation, CENPA, and γH2AX/MPM2/H3 allowed for ToxPi to provide a biologically informed prioritisation of biomarkers, which further refined the mechanistic classification of tested compounds. This visualisation approach also highlighted the variations in endpoint responses across different MoAs, revealing an assay’s potential to provide novel insights into compound toxicity profiles. In early drug development, this information can inform risk assessment by flagging potential genotoxicity mechanisms early, enabling compound prioritisation before advancing to more regulatory tests.

However, the iScreen platform encountered several challenges, particularly in endpoint specificity and compound sensitivity. Some compounds failed to produce expected dose-response curves, possibly due to factors like compound stability, required metabolic activation, or specific MoA that the assay could not detect. Additionally, non-specific cellular stress led to unexpected biomarker activations in certain compounds, highlighting the importance of refining assay conditions to minimise off-target effects – situations where reporters are activated as a secondary effect as opposed to a primary one (e.g. non-genotoxicants inducing yH2AX activation due to cytotoxicity rather than direct DNA damage). The weightings applied to the ToxPi profiles, while valuable, lacked an established

system for precisely determining endpoint significance and while the study provided an initial framework for biomarker weighting based on relative biological relevance, further research is required to validate these weightings and potentially standardise them for broader use across multiple assays. Developing a standardized weighting approach based on empirical data from a range of genotoxicity assays could provide greater consistency and improve the interpretive power of ToxPi profiles.

This study suggests several avenues for future development of the iScreen assay through the expansion of the panel of compounds and biomarkers to cover a wider range of MoA could enhance predictive modelling through improvements in compound potency while enhancing the assay's capability to detect unique genotoxic mechanisms, such as oxidative stress or specific kinase interactions. Integrating complementary assays, such as ToxTracker, could provide an additional layer of mechanistic information, particularly for compounds with atypical MoA. Additionally, the incorporation of artificial intelligence and machine learning algorithms could further optimise data analysis, reducing the potential for human bias in endpoint weighting and increasing the throughput and accuracy of mechanistic classification.

Overall, the iScreen high-content imaging platform represents a significant advancement in *in vitro* genotoxicity testing using its ability to automate and quantify DNA damage events with mechanistic precision, offering a substantial improvement over traditional methods, making it a valuable tool for early-phase drug development. The integration with PROAST and ToxPi indicates the potential of iScreen to perform detailed mechanistic analysis, enabling sophisticated classification of aneugenic and clastogenic compounds by their MoA. By addressing the challenges identified in this study and pursuing further methodological advancements, the iScreen platform holds promise for becoming a cornerstone tool in the future of genotoxicity testing, supporting safer and more effective drug discovery and future drug development.

Chapter 5: Benchmark Dose and high-throughput-toxicokinetic modelling for *in vitro* to *in vivo* extrapolation (IVIVE) to generate administered equivalent doses (AEDs); a potential alternative to animal testing.

5.1 Introduction

The future of genotoxicity testing models is highly dependent on the continued effort of a scientific collective that pursues the potential development and evolution of testing assays using the 3Rs principle (Replacement, Reduction, and Refinement) to find solutions to current drug testing methodologies (Russell and Burch, 1959; Sewell et al., 2016**).** Currently, a critical step in this process is maximising the information extracted from these genotoxicity datasets to retain dataset integrity and translate raw data into interpretable findings that directly reflect potency/genotoxicity, reducing the need for follow-up testing. Additionally, the process and means to extract this data is attempting to make this as streamlined and straightforward as possible to extract the most relevant drug dose-response information that provides the greatest insight into drug mechanisms, reactions, and MoA. This both strengthens the accuracy of safety assessment whilst simultaneously reducing/minimising the use of animal testing due to ethical concerns surrounding the practice within industries such as pharmaceuticals, agriculture, cosmetics, food, and chemicals. Genotoxicity assessment focusing on the genetic aspects surrounding these concerns, hence why testing models that focus on detecting and identifying mutagenic, clastogenic and aneugenic compounds have been under continued development, however, their reliance on animal models remains and continued international pressure/demands to develop robust alternatives continue to arise (Kavlock et al., 2018). The standardised testing battery described in chapter 1 continues to remain a staple for drug assessment despite the reliance of *in vivo* testing on low-throughput strategies and procedures. Recently, assays such as the MultiFlow, MicroFlow, ToxTracker, etc have increased the throughput capacity at which compounds could be thoroughly assessed. However, it is important to note these are primarily mechanistic tools rather than direct animal model replacements, so despite their success in developing and advancing the landscape for genotoxicity testing assays in hazard identification and MoA understanding, no true alternative animal-free testing methodology has been devised.

The Grimm (2019) paper describes how the U.S. Environmental Protection Agency and Canadian federal government has committed to the eventual elimination of animal/mammal toxicity testing altogether and both have begun to strive towards developing alternative rapid

screening assessment tools to devise such a solution. These *in vitro* NAMs attempt to modernise risk assessment methodologies and prioritise rapid screening techniques to achieve an *in vivo*-free alternative to current testing systems. ToxTracker, MultiFlow and MicroFlow assays are all examples of NAMs that effectively measure DNA damage reporter signals and allow for the identification of genotoxicants and their respective MoA, offering a promising toolset that scientists can explore to better understand the properties of known and unknown compounds/chemicals. However, it is required to build on this precedence as these *in vitro* assays usually require *in vivo* follow-up tests to improve the utility of their capabilities of direct comparison with human exposure and to correspond with pharmacodynamic and pharmacokinetic properties exhibited by humans and other mammals. Furthermore, kinetic and toxic properties might be displayed *in vitro* but not *in vivo* and testing methods *in vivo* are currently necessary as absorption, distribution, metabolism and excretion are taken into consideration for contributing towards other genetic downstream events that may manifest as genetic mutations/aberrations (ICH S2, 2012).

IVIVE offers a solution to this concern through the direct extrapolation of experimental *in vitro* datasets into *in vivo* figures without the requirement of a battery system while providing a NAM assay that quantitatively and comparatively assesses genotoxicity datasets taking both toxicodynamic and toxicokinetic ADME parameters into account. IVIVE translates *in vitro* genotoxic concentrations, BMD figures, and toxicokinetic figures (μM) into *in vivo* genotoxic AEDs (mg/kg/bw/day) representing the administered dose that requires administration *in vivo* to reach a steady-state concentration in the plasma that induced genotoxicity. This provides significant advancement in genotoxicity by providing an alternative to animal testing for genotoxicity testing using human equivalent extrapolated figures, implementing potency prediction techniques to determine the PoD and for potency and risk assessment purposes.

IVIVE, also known as reverse dosimetry, is a methodology that employs the use of an *in vitro* blood/target tissue concentrations to achieve predicted *in vivo* human doses utilising toxicokinetic ADME modelling (drug action upon administration). It is a dose reconstruction/estimation of environmental exposures that are consistent with biologically relevant monitoring data. It uses cross species extrapolation to transform *in vitro* rodent models into *in vitro* human models and apply IVIVE to obtain *in vivo* human toxicity figures in the form of the AED using toxicokinetic modelling to bridge the gap between exposure and assessment. Chemical specific toxicokinetic modelling would be compared to *in vivo* measurements to evaluate the bias/uncertainties surrounding IVIVE, however this remains

unavailable due to insufficient data, high resource requirements and chemical-specific challenges, such as biological variability and challenges in extrapolation power.

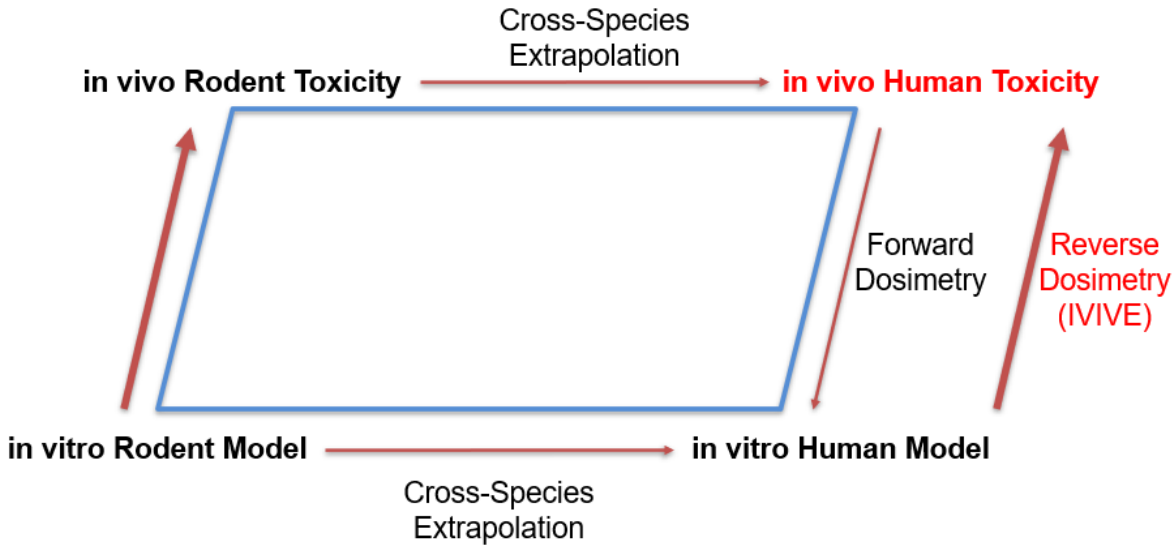


Figure 5.1.1 – A depiction of the reverse dosimetry and cross-species extrapolation applied from *in vitro* rodent models to obtain *in vivo* human toxicity figures produced by (Beal 2022 – Appendix 6).

Additionally, HTTK models provided an *in-silico* framework to estimate the daily dose required in humans to produce a concentration in the body equivalent to the toxic dose *in vitro*, providing direct extrapolation. HTTK simplifies toxicokinetic parameters across a large chemical space and provides greater confidence in modelling implementation despite it producing some uncertainty around specific datapoints.

HTTK modelling involves the application of a series of generic toxicokinetic models to predict *in vivo* measurement estimates, such as tissue or plasma concentrations, from chemical exposure dose-response data. This produces an estimated daily dose required in humans to reach a body concentration equivalent to the concentration observed as toxic *in vitro*; essentially functioning as an extrapolatory tool. Unlike chemical-specific PBPK data, which produces the most accurate chemical-specific estimates but often relies on a multitude of variables requiring extensive assessment to acquire, HTTK has adapted the toxicokinetic models used from pharmaceuticals. This allows for broader application to generate steady-state plasma concentrations (Css) by specifically extracting necessary chemical-specific information – particularly, fraction unbound in plasma (Fub), hepatic clearance (Clint), and the octanol-water partition coefficient (logP).

If accurate figures can be produced when these broader generic models are applied, this approach could be highly beneficial as it would encourage the refinement of existing *in vitro* models to estimate human-equivalent exposure doses of various drugs without the need for highly specific PBPK modelling. These highly specific models reply on the additional extraction of toxicokinetic variables, which can be a time-consuming and resource-intensive process, especially while chemical-specific toxicokinetic models remain unavailable for many compounds.

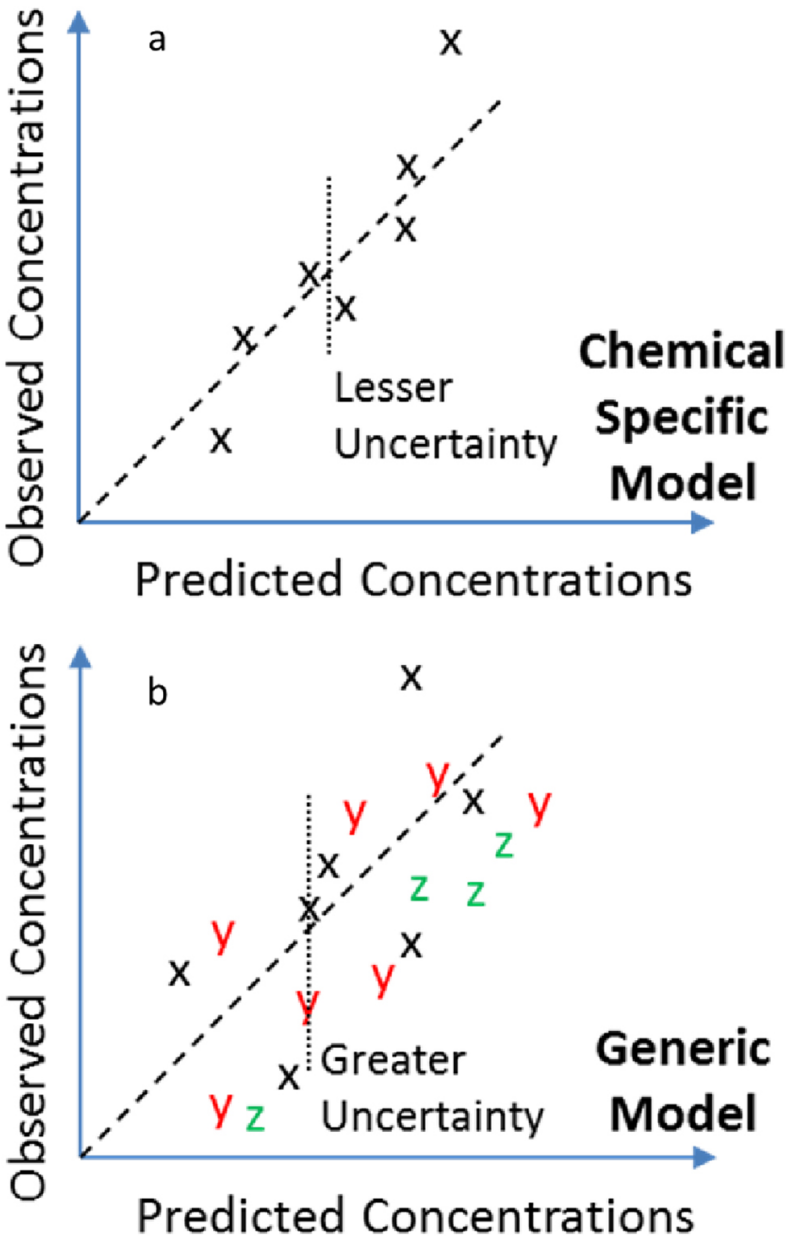


Figure 5.1.2 – An evaluation of both chemical and genetic pharmacokinetic based toxicokinetic models, demonstrating the increased uncertainty surrounding the generic models as opposed to the chemical-specific model. For situations where chemical specific toxicokinetic data remain unavailable, the genetic model still demonstrates the ability to accurately model *in vivo* data despite the increased uncertainty (Produced by Beal, 2022 and adapted form Cohen Hubal et al., 2018).

Generic HTTK models provide good estimates for human and rat *in vivo* measurements for the majority of chemicals to allow for the Css values which can be used to obtain the predicted AED. However, chemicals with HTTK predictions are not on the same order as *in vivo* readings trend to have higher Css values (conservative figure for hazard evaluation). HTTK forecaster screened 10,000 chemicals across 1400 toxicological endpoints and is compatible with quantitative high throughput screening, producing a highly detailed quantitative and qualitative database for chemical profiles. This 1400 endpoints *in vitro* Toxcast assay produces an AC_{50} value which is the concentration at 50% maximum biological response is achieved the *in vitro* assay to assist in translating *in vivo* concentration-response data into *in vivo* dose levels, making up the toxicodynamic factor. With toxicokinetic variables, including plasma binding and metabolic stability, to produce Fub, Clint and logP input variables used to produce the IVIVE figures necessary (Css), which are combined with the toxicodynamic AC_{50} figures to produce final AED prediction value, marking a conservative point of departure aligning with the upper level of human exposure to assign chemicals with the potential to disrupt normal cellular pathways at relevant human exposure levels. Additionally, the octanol-water partition coefficient and molecular weight are included to produce these toxicokinetic parameters and assist in the production of the Css.

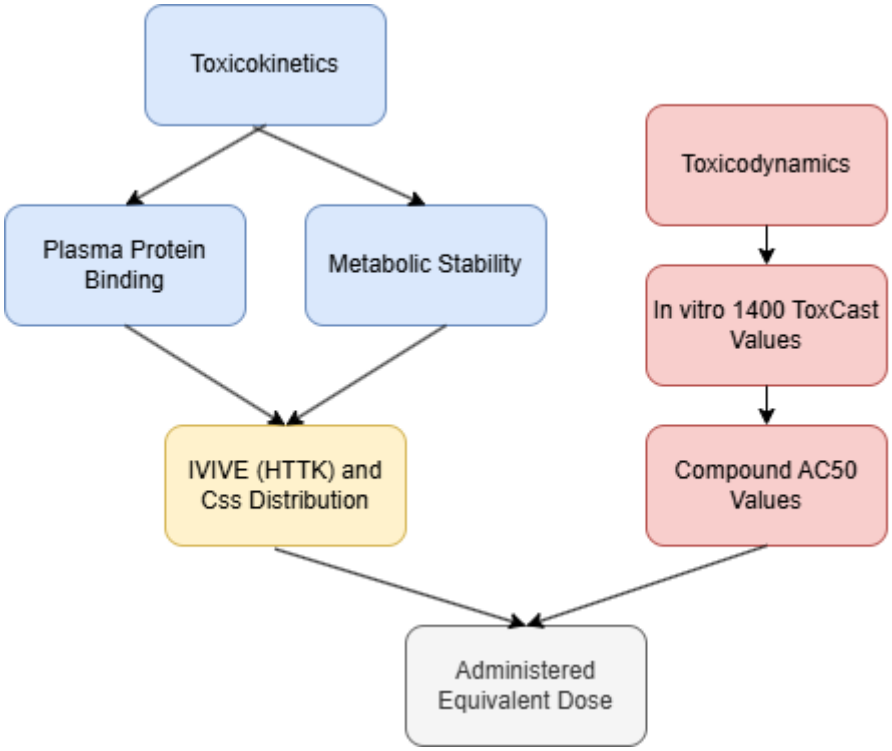


Figure 5.1.3 – The depiction of both toxicodynamic and toxicokinetic parameters combining to produce the AED, adapted from Rotroff et al., (2010). Additionally, it further displays the data provided by ToxCast, *in vitro* generated data, registration documentation and computational modelling (IVIVE application).

For HTTK modelling, the simplest modelling workflow applied used the “3compartmentss steady-state model”, which involves the use of *in silico* applications such as the ADMET predictor/online databases to obtain Fub, Clint and LogP variables which are placed into HTTK database, assisting in the calculation of the Css (Pearce et al., 2017). The metabolic factors and distribution of the Css are considered allowed for the AED to be finalised and calculated. Studies have demonstrated that *in vitro* data allows for this AED determination as a surrogate PoD metric that determines bioactivity exposure ratios and margins of exposure through the division of the determined AED by the estimated human exposure levels (Paul Friedman et al., 2020; Health Canada 2021; Beal et al., 2022.1). However, these studies did not consider each specific genotoxic endpoints within the bounds of an assay, which the Beal et al., (2022) study attempted to explore, and found that there is a case where utilisation of the AED assisting the application of IVIVE to dose-response data within different types of genotoxic assays/NAMs.

This chapter aims to explore this concept further using MultiFlow/MicroFlow/pig-a data produced by Litron Laboratories Dertinger et al., (2019) to investigate the utilisation of pharmacokinetic, pharmacodynamic and HTTK modelling to develop and refine predictive methods of performing IVIVE data analysis on 6 aneugenic/clastogenic compounds to directly link *in vitro* results to *in vivo* potency metrics and advancing potency prediction techniques for use in hazard assessment and PoD determination. From this, a pathway towards the potential development of strong extrapolatory method could assist in the deduction of a future potential alternative to animal testing modelling, reducing the requirement of animal models in genotoxicity studies, thereby lowering/alleviating the costs and ethical concerns surrounding the consequences of such practices. The FDA, Health Canada and EFSA are currently developing the testing surrounding IVIVE frameworks alongside the pharmaceutical industry, further encouraging progression within genetic toxicology to further improve the rate at which new compounds may have their genotoxicities modelled purely relying on their *in vitro* figures to produce an extrapolatory branch leading to the production of realistic conservative *in vitro* figures that can be further extrapolated toward doses.

However, unlike normal, the Css is instead combined with BMD from *in vitro* datasets (using PROAST) to replace AC50 as a normalised dose-response metric to determine the AED figures, despite their similar operation. The advantages of using the BMD as it is less

dependent on experimental doses selection/spacing, considers the entire dose-response curve and demonstrates flexibility in determining the PoD whilst the AC50 is a singular measurement where 50% of the maximum effect occurs, not necessarily considering any low-dose-modelling/biological effects thereby making the BMD a more reliable and practical alternative to facilitate IVIVE through better predictions of human risk to compound exposure (Davis, et al., 2011; Rezvanfar, 2014; EFSA 2017; Paul Friedman et al., 2020; Crump 2011). This would then be compared to *in vivo* values to assess the effectiveness of IVIVE in estimating the PoD through the AED.

Table 5.1 – Definitions for IVIVE and HTTK Abbreviations Table (Pearce et al., 2017).

Variable	Name
1compartment	One compartment model
3compartment	Three compartment model
3compartmentss	Three compartment steady state model
BW	Body weight
Css	Steady-state concentration in the plasma
Clint	*In vitro* intrinsic hepatic clearance
PBTK	Physiologically based toxicokinetic model
AC50	Concentration at 50% maximum activity
Kdose	Constant dose rate
Fub	Fraction unbound, *in vitro* ratio of unbound to total concentration in plasma
Qgfr	Glomerular filtration rate
Qliver	Blood flow to the liver.
Qgut	Blood flow to the gut
Clmetabolism	Whole liver hepatic clearance scaled from Clint
Rblood2plasma	Ratio of the blood concentration of a chemical to the plasma concentration
Qcardiac	Cardiac output/blood flow through the heart and lungs
Ql	Blood flow to the liver
GFR	Glomerular filtration rate
Kgutabs	Gut absorption rate, default of 1 h^{-1}

Clwell-stirred	Hepatic clearance modelled with well-stirred approximation using Clmetabolism
AED	Administered equivalent dose
IVIVE	*In vitro in vivo* extrapolation
HTTK	High throughput toxicokinetics

5.2 – Materials and Methods

5.2.1 - Workflow for Data Handling and Assay Processing

As outlined in chapter 2, the *in vitro* testing systems used here encompass endpoints from the MultiFlow and MicroFlow DNA damage assays, including γH2AX, p53, p-H3, MN, and polyploidy. The benchtop experiments were conducted by Litron Laboratories using the MultiFlow® DNA damage kit (Litron Laboratories, Rochester, NY, USA), assessing p53, γH2AX, and p-H3 endpoints as described in chapter 2 where TK6 cells are analysed at 4/24-hr time points via flow cytometry. The p53/γH2AX endpoints were assessed using propidium iodine median channel fluorescence intensity, expressed as fold-change relative to solvent controls and p-H3/polyploidy was assessed using relative positive and increased DNA content nuclei percentage respectively. Meanwhile, the MN endpoint was obtained through the microflow assay using the *in vitro* MicroFlow® kit reagents (Litron Laboratories, Rochester, NY, USA) to obtain the MN induction determined by SYTOX green and ethidium monoazide fluorescence. Data were organized in tables separating concentrations (μM) from median fluorescence intensity for γH2AX and p53, microbead counting for RNCs, relative population doubling for p-H3 and polyploidy and MN% fold change for MN (Litron Laboratories ., 2018). Measurements were taken highlighting the known MoAs for each chemical in Excel spreadsheets before further processing took place.

Table 5.2.1.1 – A list of compounds investigated during this IVIVE chapter including their respective sources and individual MoA.

Compound	**CASRN, source**	**MoA classification (Sources)**
Cisplatin	15663-27-1, Sigma-Aldrich	Atypical alkylator/cross-linking (Clastogens) (Kirkland et al., 2016)

Cyclophosphamide Monohydrate	6055-19-2, Sigma-Aldrich	Nitrogen mustard (Clastogens) (Kirkland et al., 2016; Rodriguez-Antona and Ingelman-Sundberg., 2006)
Diethylnitrosamine	55-18-5, Sigma-Aldrich	Polycyclic aromatic hydrocarbon (Clastogens) (Yamazaki et al., 1992)
Hydroxyurea	127-07-1, Sigma-Aldrich	Ribonucleotide reductase inhibitor; (Non-Mutagenic Clastogens) (Ware and Dertinger, 2021)
Melphalan	142-82-3, Sigma-Aldrich	Nitrogen mustard-type alkylator (Clastogens) (Dertinger et al., 2012)
Vinblastine Sulphate	143-67-9, Sigma-Aldrich	Tubulin Binder (Aneugen) (Kirkland et al., 2016)

For the *in vivo* endpoints, the rat *pig-a* gene mutation test is implemented using the Microflow® /MutaFlow® testing kit (Litron Laboratories, Rochester, NY), containing dimethyl sulfoxide (Sigma-Aldrich) and heat-inactivated fetal bovine serum (VWR, Radnor), to extract *in vivo* dose-response metrics (Litron Laboratories, 2018). The *in vivo* MicroFlow® PLUS-R flow cytometry kit scores for micronucleated reticulocyte (MN-RET) scoring indicating chromosomal damage, the *in vivo* Rat MutaFlow® Kit (*pig-a* gene mutation test) to assess mutant erythrocytes/reticulocyte using anti-CD59 PE/ anti-CD61-PE markers (MUT-RET), and the *in vivo* MicroFlow® PLUS-RL kit for liver micronuclei scoring (MNHEP) (for detecting genomic instability/damage to hepatocytes) and proliferation (ki-67+) (signifying cellular repair activation due to stress or tissue injury caused by potential genotoxic agents). The protocols are described in (Dertinger et al., 2019), however briefly, rats started treatment at 4 weeks with 3 males and 3 females per treatment group (following IWGT recommendations) and test chemicals were administered orally at 10 mL/kg bodyweight for (cyclophosphamide, diethylnitrosamine, hydroxyurea and melphalan) with exception to cisplatin and vinblastine sulphate which were provided as injections into the intraperitoneal region of each rat at 5 mL/kg bodyweight (Uno et al., 2015).

Each chemical was assessed through one of three designs: 3 days x 1 week (3Dx1W), 3 days x 3 weeks (3Dx3W) and 7 days x 4 weeks (7Dx4W). The 3Dx1W design consisted of exposure for 3 consecutive days with 24-hour intervals between each exposure where, MN-RET data is collected on the 4th day from tail vein blood using the *in vivo* MicroFlow® PLUS-R assay. On the 7th day (four days after the final exposure), the rats had all their blood drained and left lateral liver lobes collected for use in the the *in vivo* MicroFlow® PLUS-RL

assay to assess MNHEP/ki-67+. Meanwhile, a fraction of the same blood sample was used to conduct the *pig-a* test using the *in vivo* Rat MutaFlow® Kit for the MUT-RET endpoint. For the 3Dx3W, a similar approach to the 3Dx1W was followed with consecutive exposures for 3 days, where on the 4th day, tail vein blood was collected for MN-RET analysis. However, unlike the 3Dx1W design, on days 8 – 10, rats were re-exposed to the substance with 24-hour intervals cycles between each point of exposure and was further repeated on days 15 – 17, until blood for both the MNHEP, Ki-67+ and MUT-RET endpoints were gathered on the 21st day (end of the third week). Additionally, for the 7Dx4W exposure, unlike the other designs, rats were exposed to the test chemical across all days of the week for 4 consecutive weeks with 24-hour intervals between each individual exposure. Rat tail vein blood was still collected on the 4th day after exposure for MN-RET analysis and the final blood sample was collected on day 28 (1 day after final exposure) for MNHEP, Ki-67+ and MUT-RET analysis. All this was described in (Dertinger et al., 2019) and further explained and detailed in the instructions provided by each kit (Litron Laboratories., 2018; 2021; 2022).

The MN-RET (and CD71+) frequencies were determined on the 4th day of each experimental design and followed the procedures outlined in Dertinger et al., (2004) and described following the *in vivo* MicroFlow® PLUS-R instruction manual (Litron Laboratories., 2017). Briefly, they describe how the collected tail vein blood sample is mixed with coagulants to prevent clotting and fixed using cold methanol fixative and placed in an ultracold freezer (-75°C to -85°C) which would afterwards be placed into long term storage solution (LTSS). On the day of analysis, compounds are washed out of the fixative and LTSS from the samples using buffer solution with 1% fetal bovine serum (FBS) and following this, the samples are prepared for labelling using a combination of buffer solution, RNase solution, anti-rat CD71 antibody (producing labelling solution 1) and labelling solution 1 with CD61 antibody to produce labelling solution 2. Upon exposure to DNA staining solution (DNA stain with buffer solution), flow cytometry is performed using the FACSCantoII flow cytometer with the FACSDiva™ software where manufacturer instructions are followed alongside plotting adjustments to extract response frequencies, which are converted to response fold change figures relative to the solvent control, which are used to extract the BMD, post-flow cytometric analysis, based on CD71 cell surface markers.

For the extraction of MNHEP, Ki-67+ information, *in vivo* MicroFlow® PLUS-RL was performed on the 4th day, 21st day or 28th day of the analyses as described earlier using rat

liver after the rats are exsanguinated to extract frequency percentages. The left lateral lobe was harvested and placed into cold ice with liver preservation buffer (ethylenediaminetetraacetic acid (EDTA) and Hanks' Balanced Salt Solution (HBSS)) containing DMSO. Once the liver is removed and dried from the preservation buffer, a 1g section of the liver left lobe was removed and added to liver rinsing solution (dH2O and liver rinse), shaken and rewashed. The liver solution is then transferred into erythrocyte cleaning solution (dH2O, liver rinse and EGTA (ethylene glycol tetra acetic acid) solution) and collagenase solution (HBSS. FBS and collagenase solution). Then FBS and HBSS is added, and the liver is chopped with scissors and vortexed allowing for larger pieces to settle and transferring the supernatant to a new tube where it is centrifuged. The mixture is resuspended in HBSS and FBS and washed several times, passed through a strainer and on the day of analysis, after storage, complete lysis solutions 1 (incomplete lysis solution 1, RNase and Ki-67 Antibody eFluor® 660) and 2 (Incomplete Liver Lysis Solution 2 and Liver Nucleic Acid Dye) are added and flow cytometry is performed to analyse compounds using the FACSDiva™ software as described earlier.

For the MUT-RET endpoint, the *pig-a* test is performed using the *in vivo* Rat MutaFlow® kit to extract erythrocyte and reticulocyte information using CD59 as the primary GPI-anchored protein marker identifying erythrocytes and mutated reticulocytes (MUT-RET). Once the whole blood sample from the rat (post-rat exsanguination) at identical time intervals to the MNHEP/Ki-67+ (*in vivo* MicroFlow® PLUS-RL) test (4^{th} day, 21^{st} day and 28^{th} day). To briefly describe the kit instructions, the whole blood sample is combined with anticoagulant solution and aliquoted with rodent blood freezing solution in preparation to be frozen. The frozen samples are then thawed before *pig-a* testing commences using the MutaFlow® Mouse or Rat Blood Thaw and Analysis Kit (Litron Laboratories., 2022). Once thawing is complete, anticoagulant solution and lympholyte® mammal solution is added to enrich erythrocytes and assist in depleting platelet/leukocytes via centrifugation. After erythrocyte-rich pellets are washed and resuspended with buffered salt solution, working antibody solution (anti-CD59-PE) is added to attach the appropriate antibodies to CD59 glycoproteins, and once unbound antibodies are washed off with buffer salt/FBS solution, the cells are incubated in working anti-CD59-PE microbead suspension. Once unbound the anti-PE microbeads are washed off, buffered salt and FBS solution are added again to resuspend the cells, and the remainder of the antibody-labelled cells are applied to Miltenyi magnetic cell separation columns (immunomagnetic depletion) (Miltenyi biotec) and eluates are collected

which are subsequently collected and supernatants aspirated. Cell pellets from the eluates are resuspended in nucleic acid dye and counting beads solution for flow cytometric analysis in the form of relative anti-CD59-PE fluorescence events using FACSDiva™ software as described in the other *in vivo* assays using a FACSCanto II flow cytometer (Dertinger et al., 2011).

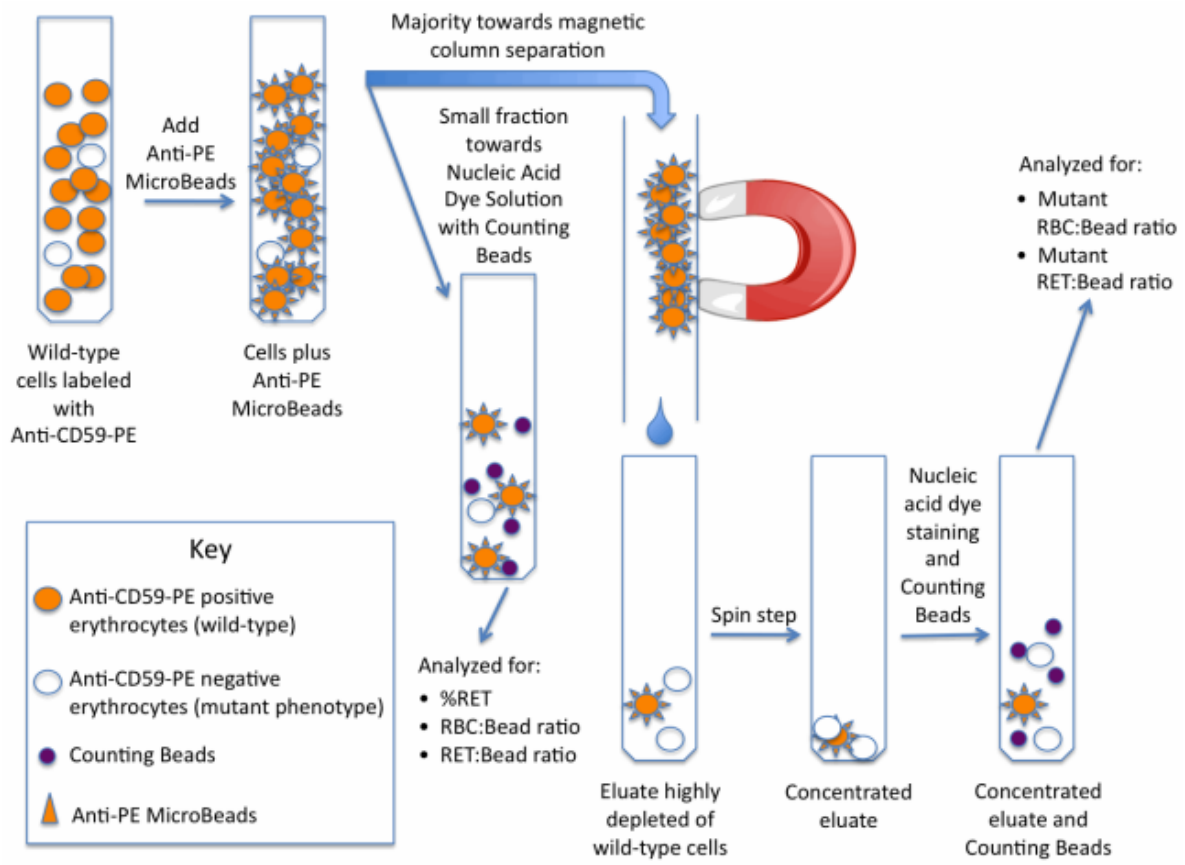


Figure 5.2.1.2 – Diagram depicting the MutaFlow® Mouse or Rat Blood Thaw and Analysis high throughput protocol within the *pig-a* test performed using the *in vivo* Rat MutaFlow® kit. Blood samples are incubated with anti-CD59-PE to label wild-type erythrocytes (PE-positive) and mutant phenotypes (PE-negative). After adding anti-PE magnetic particles and Counting Beads, a portion is analysed by flow cytometry for pre-column cell-to-bead ratios and reticulocytes. The remaining sample passes through a magnetic column, where mutant cells and beads are collected in the eluate. Post-column eluates are stained and analysed to calculate mutant cell frequencies (Litron Laboratories, 2022).

5.2.2 – PROAST procedure to extract BMDs

Once all raw dose-response metrics had been collected in the format of CSV files, for all *in vivo* figures, including Ki-67+, MNHEP, MN-RET, MUT-RET across 3Dx1W, 3Dx3W, 7Dx4W conditions, the BMDs were extracted using PROAST v70 through the EFSA website following the same conditions as stated in previous chapters: continuous individual response data, exponential models model 3 and 5, independent set as the dose/concentration, each response set as the dependent variable, study design/endpoint used as the covariates and CES set at 50% (0.5) with 200 bootstrap runs to produce BMDL and BMDU values. These values were chosen by Litron Laboratories when developing the protocols for *in vivo* datasets for a one-off collaborative project. For *in vitro* datasets, an identical procedure was followed with the exception that the R version of PROAST v70.3 was applied and concentrations where cytotoxicity exceeded 60% for microflow endpoints (MN) and 80% for MultiFlow (polyploidy, p-P3, p53 and γH2aX) endpoints (100 – RNC value below 40% and 100 – relative increased nuclei count (RINC) value below 20% respectively), are excluded from the analyses due to excessive cytotoxicity as alluded to in earlier chapters. Once all BMDs are collected, 0.0001 – 10000 μM transformations (biologically tested range) are applied to both *in vitro* and *in vivo* results before further processing is performed to transform these figures into AED values using HTTK.

5.2.3 - 3compartmentss modelling

The workflow used for HTTK modelling to derive the AED figures (mg/kg-bw/day) follows the 3-compartment-steady-state model (3compartmentss). Toxicokinetic parameters were measured using the plasma protein binding and hepatic clearance (metabolic stability) assays which are combined with the *in silico* ADMET predictor (version 10) for chemicals with missing input toxicokinetic parameters following previous methods outlined by Marc Beal in (Beal et al., 2022.1). These predictions are mostly based on *in vivo* human data which is highly relevant for this investigation. The toxicokinetic data is then ran through the HTTK database to measure these experimental parameters to model the Css distribution using Monte Carlo simulation for HTTK to account for interindividual variation in human populations by considering different demographics and subgroups (sex, age, body weight, ethnicity, and renal function parameters).

3compartmentss equation

$$C_{ss} = k_{dose} \Big/ \left(f_{ub} Q_{gfr} + \frac{(Q_{liver} + Q_{gut}) f_{ub} Cl_{metabolism}}{(Q_{liver} + Q_{gut}) + f_{ub} Cl_{metabolism} \,/\, R_{blood\,2\,plasma}} \right)$$

Figure 5.2.3.1 – The equation for the 3compartmentss steady state model to generate the HTTK predicted Css where the definitions of each component are placed in the table below (Pearce et al., 2017).

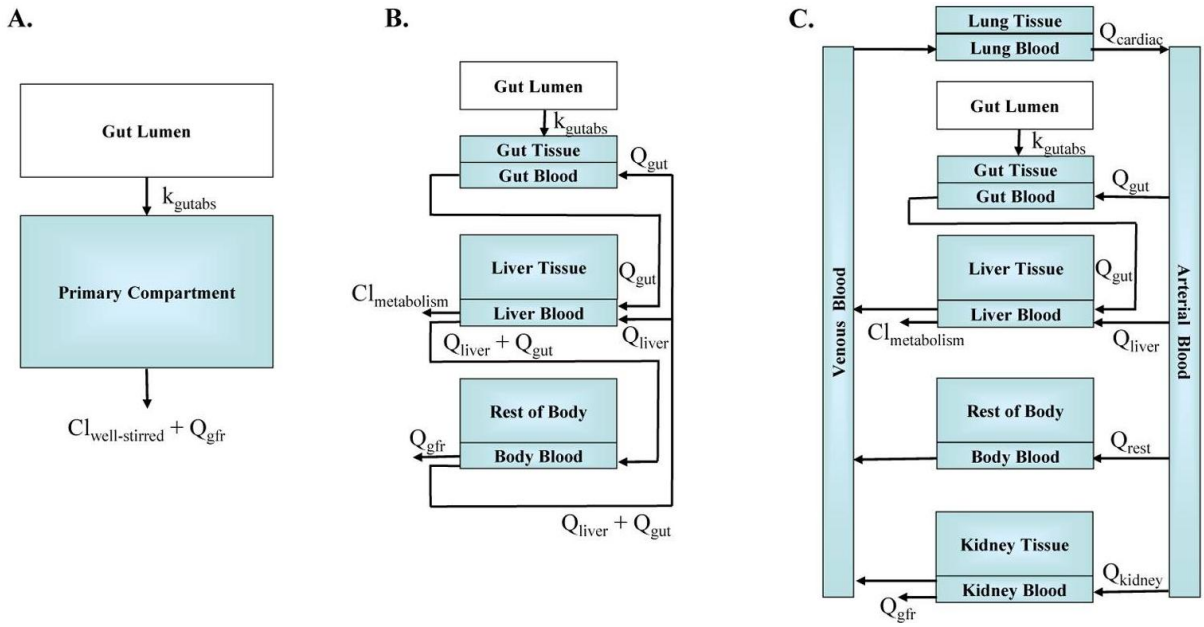
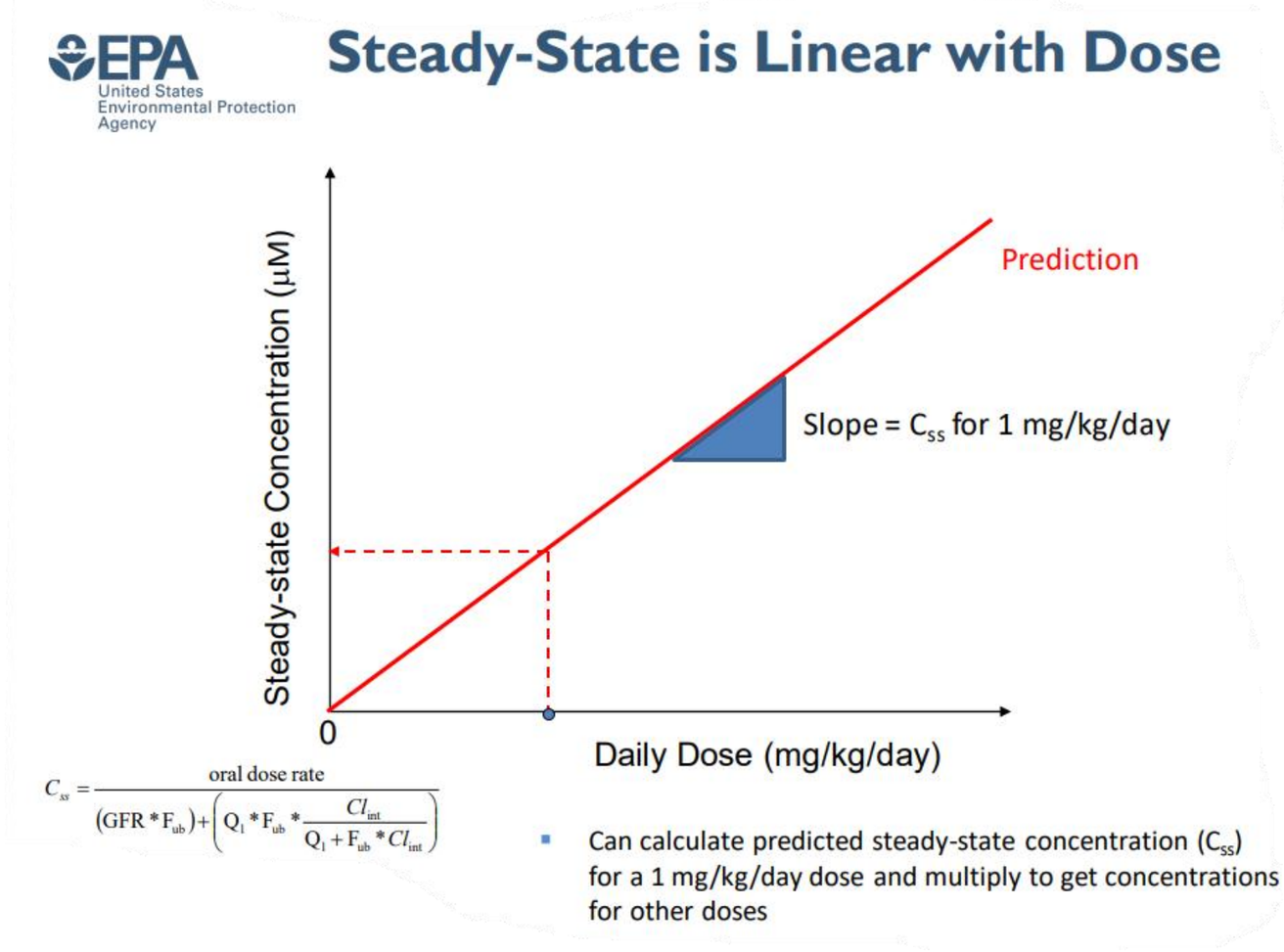


Figure 5.2.3.2 – The (A) 1compartment, (B) 3compartment and (C) PBTK modelling diagrams showing flow from the gut lumen to the blood vessels, gut, lungs, the liver, kidney and rest of the body to predict the Css where definitions of each component are placed in the table below (Pearce et al., 2017).

The Css distribution is calculated using the 3compartmentss equation and as the steady-state is linear with dose, HTTK allows for steady-state IVIVE to be mapped (Wetmore et al., 2012). The prediction gradient is determined comparing daily dose (mg/kg/day) and the Css (μM), which allows for the "Css for 1mg/kg/day" to be determined using the gradient itself (μM ÷ mg/kg/day) the gradient can be multiplied by the daily dose to get the Css for each daily dose. The Css reflects the oral ingestion rate over the renal clearance and hepatic clearance and is generated computationally using the ADMET predictor, which finds the Fub, Clint and LogP variables for each compound as the set toxicokinetic parameters and combine with chemical canonical SMILES and molecular weight to assist in the calculation of the Css using HTTK.



$$C_{ss} = \frac{\text{oral dose rate}}{(GFR * F_{ub}) + \left(Q_l * F_{ub} * \frac{Cl_{int}}{Q_l + F_{ub} * Cl_{int}}\right)}$$

(Passive) Renal Clearance

Hepatic Clearance (Metabolism)

Figure 5.2.3.3 – This graph depicting how the Css (µM) and daily dose (mg/kg/day) can be plotted against each other to generate a linear line, and the gradient (Css ÷ daily dose) is equivalent to the predicted "Css for 1mg/kg/day" (which can also be multiplied to get concentrations for other doses) (Wetmore et al., 2012).

Next, the axes are reversed, which is the reverse dosimetry step, so the new gradient (daily dose ÷ Css) is now equivalent to the mg/kg/day per Css (when simulated dose is 1 mg/kg/day). To obtain the AED, we then divide the bioactive concentration at 50% maximum activity (BMD) by the Css (when simulated dose is 1 mg/kg/day).

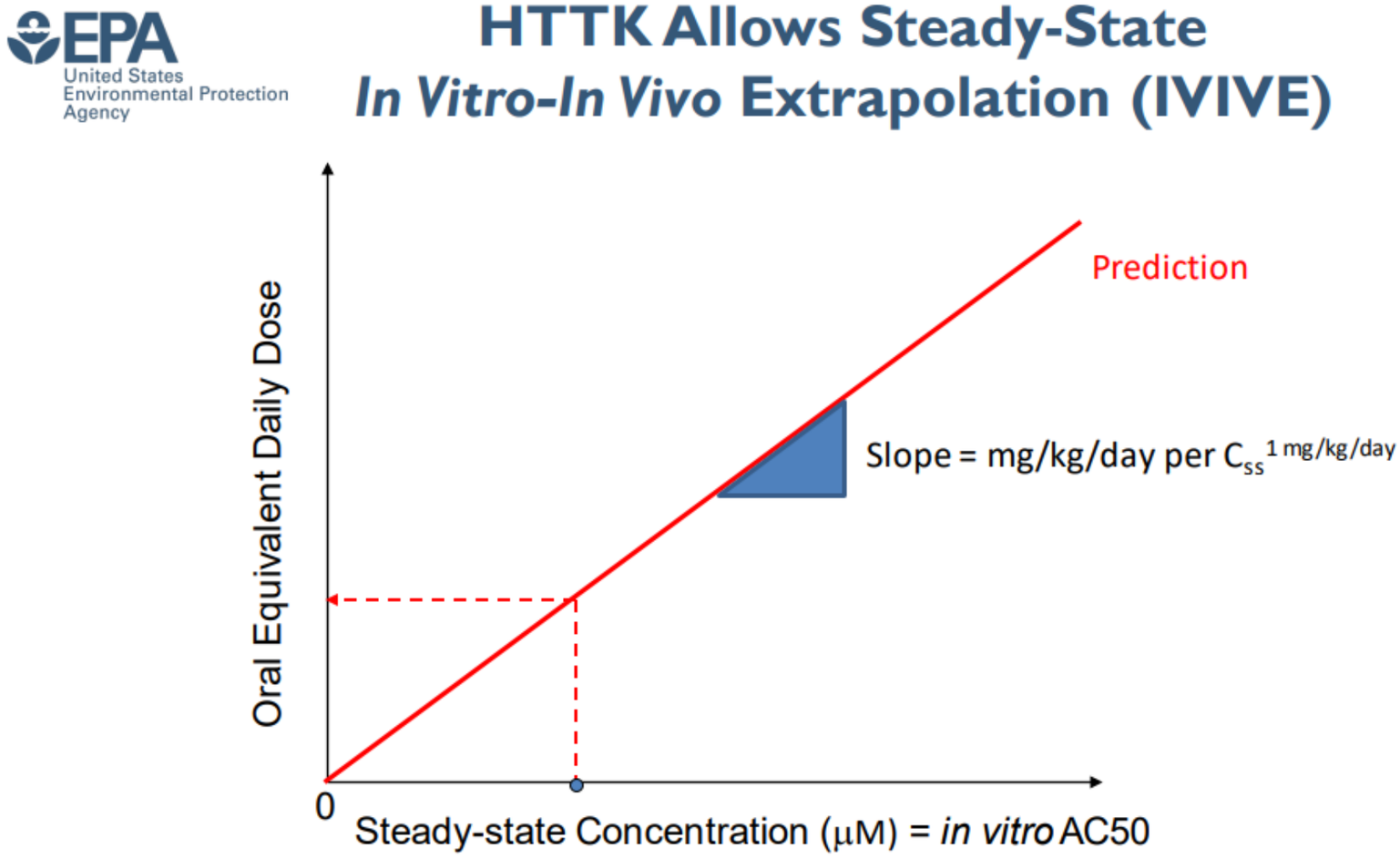


Figure 5.2.3.4 – This graph shows the implementation of reverse dosimetry by swapping the axes, and the equivalent daily dose and Css (*in vitro* AC50) are plotted against each other to generate a slope where the gradient is mg/kg/day per Css for 1mg/kg/day. The final AED is finally found when AC50 ÷ Css (when simulated dose is 1mg/kg/day) (Wetmore et al., 2012). In this chapter, the calculations used are identical with the exception that the *in vitro* AC50 is replaced by the BMD.

HTTK was used to generate the Css for 1mg/kg/day value. This is followed up with the BMD being divided by the Css for 1mg/kg/day figure to generate the final AED for each compound. The HTTK 3compartmentss model can apply the largest number of chemicals specifically those with missing information needed to parameterise the model. Substituting current methods with this new approach methodology (NAM) will enhance prioritisation, rapid screening, and genotoxic chemical assessment, allowing us to further develop strategies to combat limitations and within the *in vitro* paradigm regarding human relevance (considering the variation within a population) and how and eliminating the ethical concerns regarding the *in vivo* approaches to compound analysis.

5.2.4 – Css determination using HTTK to generate the AED

This chapter aims to use the BMD as opposed to the AC50 PROAST is utilised with raw *in vitro* data to produce BMDs which will be used in the production of AED figures and compared to *in vivo* values. Once, the BMDs are collected across all endpoints using the aforementioned conditions and calibrations, they are curated into xlsx file. Next, the ADMET predictor/online drug databases are used and produce Fub, Clint and logP values using the canonical SMILES and molecular weight of each compound to obtain these *in silico* parameter predictions based on mostly *in vivo* human data. Once these variables are found, the Css is found using the HTTK database. Firstly, the toxicological parameters are ordered in the xlsx files in the format drawn as shown below:

Chemical.Name	CASRN	Human.Clint	Human.Fub.plasma	logP	MW
Bisphenol M	13595-25-0	317.51	0.033	5.75	346.46
Bisphenol TMC	129188-99-4	119.65	0.043	5.23	310.43

Table 5.2.2.1 – The xlsx file format adopted for the importation of chemical toxicokinetic parameter within the HTTK database using Bisphenol M and Bisphenol TMC as example chemicals – adapted from the presentation by Beal., (2022).

To perform this, the toxicokinetic parameters were extracted for all compounds tested using the HTTK v2.2.2 database operating in in R Gui. This thereby allowed for these parameters to produce a Css value by inputting them into the programme and allowing for all commands used in the code to be attributed to each toxicokinetic factor for the human species, allowing the Css to be calculated using the “calc_mc_css” function and implementing Monte Carlo

simulations (Beal., 2022). The Monte Carlo simulation is a virtual population generator used within HTTK to account for inter-individual variation within the human population, considering different demographics and subgroups in the sampler by changing, gender, age limit, bodyweight, ethnicity, and renal function parameters (by default, the US population is modelled) (Breen et al., 2022). This means that Css results, in μM, at the 95th percentile of 1000 individuals are returned, using the 3compartmentss model, for the human species, with seeding is set at 1 to prevent randomisation of numerical Css values – mainly with slight variation of the decimals within a number (Breen et al., 2022). This allows for a fixed workflow that does not update every run to be produced as this value is pseudorandom in R Gui – hence a seed set at 1 is implemented. Monte Carlo simulations also vary the liver volume, cell density, blood flow, bodyweight, Qgfr and Clint with a 30% coefficient being implemented for each parameter by default. Finally, to obtain the AED (in μM), the BMD produced for each compound at each endpoint is divided the single seeded Css value produced (AED = BMD ÷ Css): the AED being the pseudo-steady-state-plasma concentration in a simulated dose of 1 mg/kg-bw/day (Breen et al., 2022). This same calculation is performed using the BMD produced across both *in vitro* and *in vivo* datasets, with identical use of the Css for all sets of analyses and conditions.

5.2.5 – ToxPi analysis

Once all AED values are curated for both *in vivo* and *in vitro* datasets, they are transformed into 0.0001-10,000 μM concentrations to aid with the visualisation process for ToxPi as described similarly to BMD analysis using ToxPi in earlier chapters. Once this was complete, all data regarding compounds, endpoints tested both *in vivo* and *in vitro* were compiled into CSV files where each chemical was listed alongside their endpoints composing of the combined conditions and biomarker responses obtained. For the *in vivo* endpoints imported (Ki-67+, MNHEP, MN-RET, MUT-RET), the BMDL and BMDU for each testing condition (3Dx1W, 3Dx3W, 7Dx4W) were placed into individual slices (i.e. one slice is labelled as "3Dx1W Ki-67+" consists of "3Dx1W Ki-67+ BMDL" and "3Dx1W Ki-67+ BMDU" imported into the slice) and the -log10(x) function is applied to all slices. Similarly, for *in vitro* endpoints (γH2AX, p53, p-H3, MN, and polyploidy), each slice consisted of the BMDL and BMDU for each endpoint and -log10(x) transformations were made. Additionally, for both *in vitro* and *in vivo* datasets, no additional weighting percentage transformations were made, and the ToxPi results were calculated using 1000 compute bootstrap runs to produce

the ToxPi CIs across all compounds. For ToxPi ranks, the ToxPi score CIs are calculated for each compound. For hierarchical clustering, the complete clustering algorithm was applied and -means clustering results are further observed and assessed.

5.3 – Results

This section highlights both the PROAST, AED and ToxPi outputs to assess the capacity and capabilities of the implementation of IVIVE platforms to visualise the direct extrapolation of *in vitro* into *in vivo* figures, using the AED as a normalised dose-response metric across 6 aneugenic/clastogenic compounds with varying MoAs. The use of *in vivo* and *in vitro* datasets produced by Litron Laboratories would allow for the exploration of this extrapolatory concept, further driving the development of pathways that would lead to the potential introduction of multiple cost-effective and ethically acceptable alternatives to current existing controversial animal testing methods that are currently implemented for the genotoxic assessments of drugs/chemicals in the pharmaceutical, agricultural, cosmetics, and food industries.

5.3.1 *In vitro* PROAST outputs

The BMD CIs were determined for the six compounds assessed to provide further insight into the dose response analysis across the *in vivo* IVIVE endpoints tested. For the *in vivo* endpoints, the study design and endpoint were used as covariates, with a CES 0.5 and 200 bootstrap iterations were used to produce the BMD CIs. *In vitro* datasets were similarly processed but excluded concentrations where cytotoxicity exceeded 60% for microflow endpoints (MN) or 80% for MultiFlow endpoints (e.g., polyploidy, p-P3, p53, and γH2AX) to mitigate excessive cytotoxicity interference.

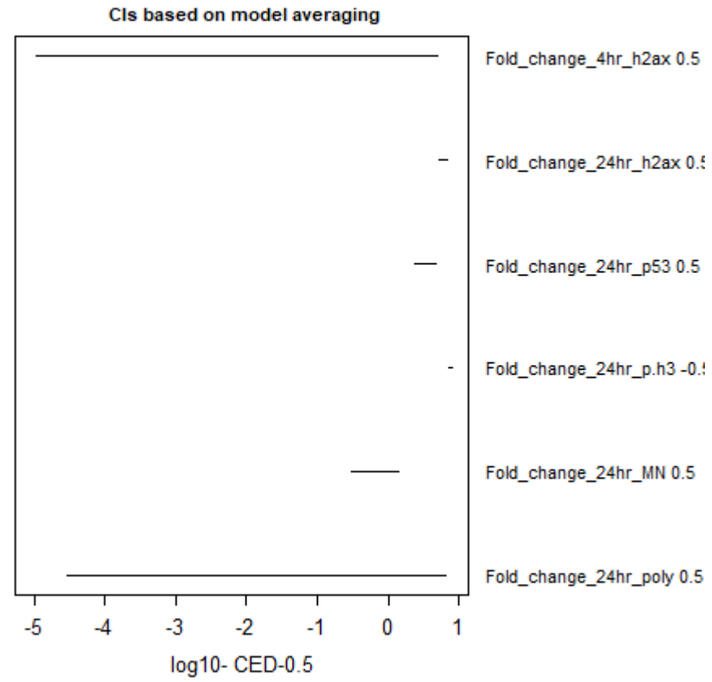


Figure 5.3.1.1 – 90% BMD CIs produced using PROAST for the *in vitro* endpoints that successfully produced BMD CIs for cisplatin, which includes 4hr H2aX, 24hr MN, 24hr H2aX, 24hr poly, 24hr p53, and 24hr p-H3, once a CES of ±0.5 is selected with model averaging of 200 bootstrap runs being applied with the nested exponential models 3/5 approach. The log10(BMD) concentration (in µM) sits on the x-axes, and each compound name on the y-axis where each bar depicts the BMD CI for each compound.

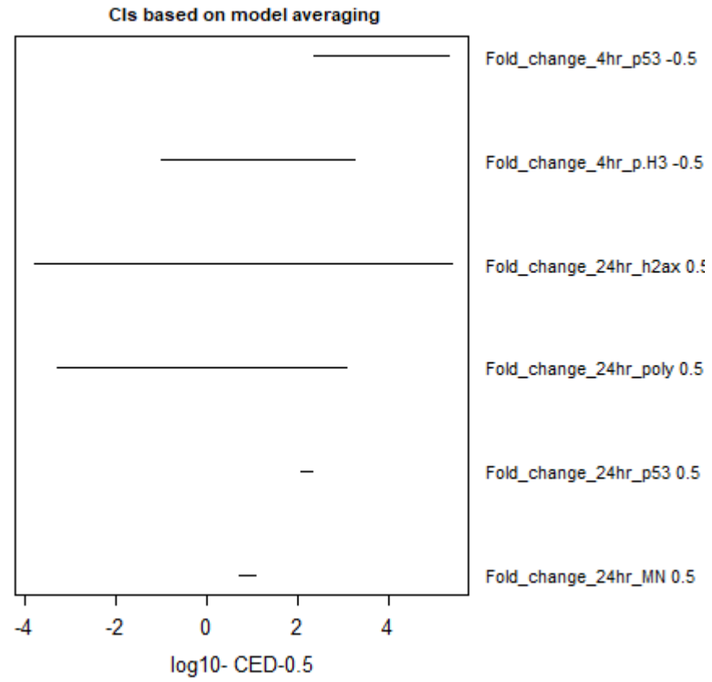


Figure 5.3.1.2 – 90% BMD CIs produced using PROAST for the *in vitro* endpoints that successfully produced BMD CIs for cyclophosphamide monohydrate, which includes 4hr p53, 4hr p-H3, 24hr H2aX, 24hr poly, 24hr MN and 24hr p53, once a CES of ±0.5 is selected with model averaging of 200 bootstrap runs being applied with the nested exponential models 3/5 approach. The log10(BMD) concentration (in µM) sits on the x-axes, and each compound name on the y-axis where each bar depicts the BMD CI for each compound.

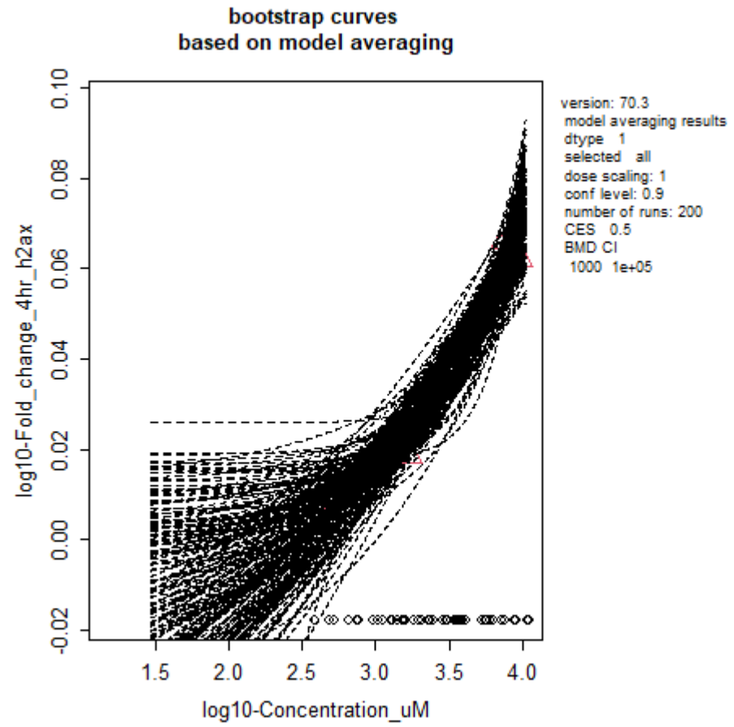


Figure 5.3.1.3 – 90% BMD CIs produced using PROAST for the *in vitro* endpoints that successfully produced BMD CIs in diethylnitrosamine, which includes the 4hr H2aX biomarker, once a CES of ±0.5 is selected with model averaging of 200 bootstrap runs being applied with the nested exponential models 3/5 approach. Each individual model has the concentration on the x-axes, and each fold change value for each endpoint on the y-axis, where each BMDL and BMDU displayed under the BMD CI section. As the 4hr H2aX endpoint is the only endpoint produced, no comparative BMD chart is available as only one endpoint successfully modelled a BMD CI.

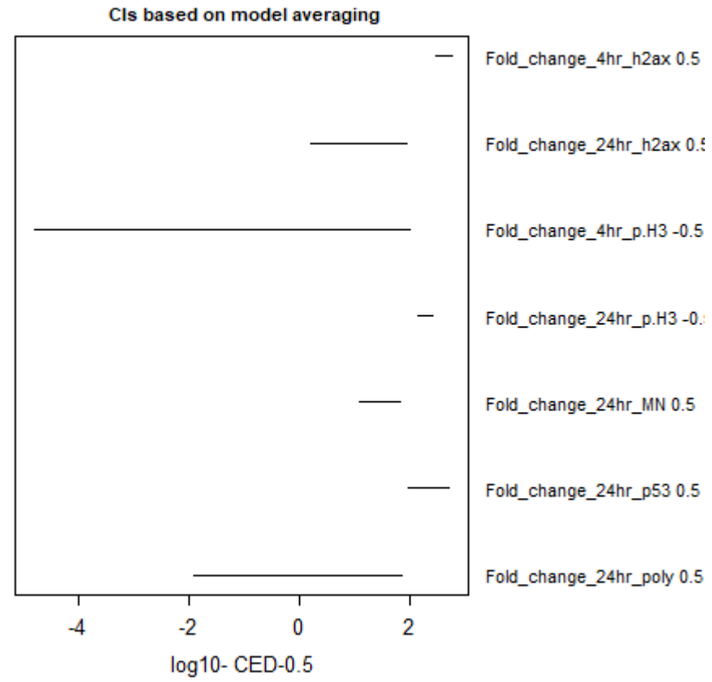


Figure 5.3.1.4 – 90% BMD CIs produced using PROAST for the *in vitro* endpoints that successfully produced BMD CIs for hydroxyurea, which includes 4hr H2aX, 4hr p-H3, 24hr H2aX, 24hr MN, 24hr p53, 24hr p-H3 and 24hr poly, once a CES of ±0.5 is selected with model averaging of 200 bootstrap runs being applied with the nested exponential models 3/5 approach. The log10(BMD) concentration (in µM) sits on the x-axes, and each compound name on the y-axis where each bar depicts the BMD CI for each compound.

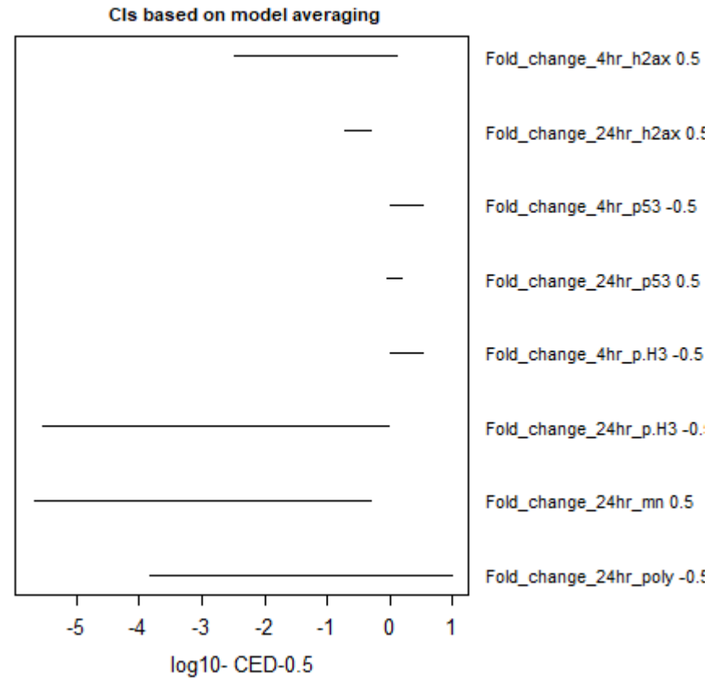


Figure 5.3.1.5 – 90% BMD CIs produced using PROAST for the *in vitro* endpoints that successfully produced BMD CIs for in melphalan, which includes the 4hr H2aX, 4hr p53, 4hr p-H3, 24hr H2aX, 24hr MN, 24hr p53, 24hr p-H3, 24hr poly biomarker, once a CES of ±0.5 is selected with model averaging of 200 bootstrap runs being applied with the nested exponential models 3/5 approach. The log10(BMD) concentration (in μM) sits on the x-axes, and each compound name on the y-axis where each bar depicts the BMD CI for each compound.

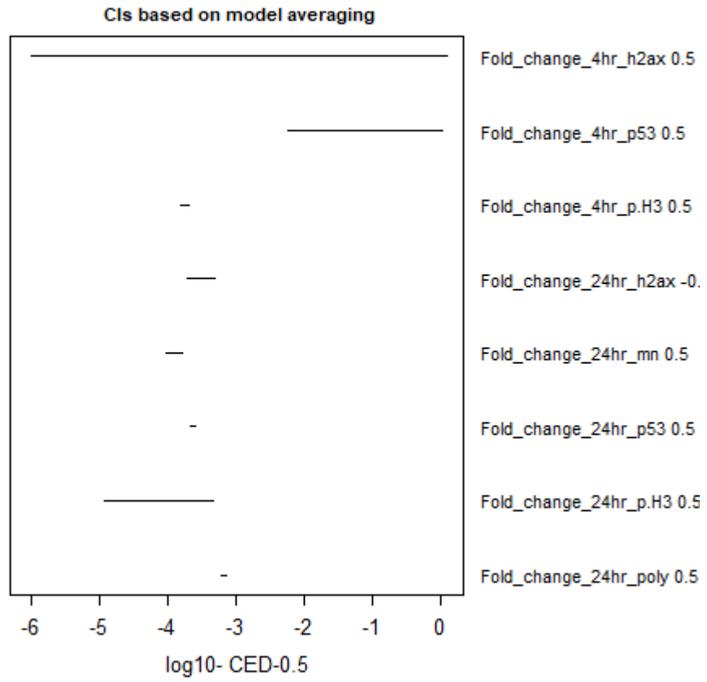


Figure 5.3.1.6 – 90% BMD CIs produced using PROAST for the *in vitro* endpoints that successfully produced BMD CIs for in vinblastine sulphate, which includes the 4hr H2aX, 4hr p53, 4hr p-H3, 24hr H2aX, 24hr MN, 24hr p53, 24hr p-H3, 24hr poly biomarker, once a CES of ±0.5 is selected with model averaging of 200 bootstrap runs being applied with the nested exponential models 3/5 approach. The log10(BMD) concentration (in µM) sits on the x-axes, and each compound name on the y-axis where each bar depicts the BMD CI for each compound.

These successfully produced 90% BMD CIs in sections 5.3.1.1 – 5.3.1.6 reflect the PoD for each compound across the tested *in vitro* endpoints (H2AX, MN, p53). Despite variability in the number of endpoints successfully producing BMD CIs for each compound, the analysis consistently employed the nested exponential models 3/5 with model averaging and a CES of ±0.5, supported by 200 bootstrap iterations. While compounds like melphalan and vinblastine sulphate showed broad endpoint activation, others, such as diethylnitrosamine, were limited. The analysis highlights variability in biomarker sensitivity and underscores the need for additional endpoints and compounds for more robust profiling. These are subsequently used in the following sections where the toxicokinetic parameters are combined with the BMD and HTTK extrapolation to perform IVIVE. The next section highlights a table depicting the *in vivo* BMDs as they were produced and collected in collaboration with Litron Laboratories as highlighted in the methods section. The tables below show the BMDs produced across each set of *in vivo* endpoints, where a CES of 0.5 (50%) was selected and 200 bootstrap runs were taken, which were used to perform AED analysis. The necessary 0.0001 – 10,000 μM transformations to the BMDs have been performed and any unsuccessful BMDs have been highlighted as “NA”.

5.3.2 *In vivo* PROAST outputs

Row	Response	Bootstrap	Compound	3DperWx1W_Ki-67+_BMDL	3DperWx1W_Ki-67+_BMDU	3DperWx3W_Ki-67+_BMDL	3DperWx3W_Ki-67+_BMDU	7DperWx4W_Ki-67+_BMDL	7DperWx4W_Ki-67+_BMDU
1	0.5	200	Cisplatin	NA	NA	0.783	2.77	NA	NA
2	0.5	200	Cyclophosphamide_Monohydrate	NA	NA	NA	NA	NA	NA
3	0.5	200	Diethylnitrosamine	20	37.8	0.279	6.56	0.0649	1.92
4	0.5	200	Hydroxyurea	NA	NA	NA	NA	NA	NA
5	0.5	200	Melphalan	NA	NA	NA	NA	NA	NA
6	0.5	200	Vinblastine Sulphate	0.0369	2.65	NA	NA	NA	NA

Table 5.3.2.1 – All the BMDL and BMDU figures produced for the Ki67+ biomarker across all compounds covering all tested days and weeks, where cisplatin, diethylnitrosamine and vinblastine sulphate produced successful BMD CI results, and all columns highlighted "NA" are instances where a BMD was no determined as dose response curve was able to be modelled.

Row	Response	Bootstrap	Compound	3DperWx1W_MN-RET_BMDL	3DperWx1W_MN-RET_BMDU	3DperWx3W_MN-RET_BMDL	3DperWx3W_MN-RET_BMDU	7DperWx4W_MN-RET_BMDL	7DperWx4W_MN-RET_BMDU
1	0.5	200	Cisplatin	0.12	0.522	0.0106	0.346	0.0864	0.29
2	0.5	200	Cyclophosphamide_Monohydrate	0.729	2.47	0.638	2.64	0.000166	1.04
3	0.5	200	Diethylnitrosamine	NA	NA	NA	NA	NA	NA
4	0.5	200	Hydroxyurea	15.6	40.6	21.5	39.1	25.8	48.2
5	0.5	200	Melphalan	0.142	0.565	0.131	0.833	0.142	0.415
6	0.5	200	Vinblastine Sulphate	NA	NA	0.0581	0.109	NA	NA

Table 5.3.2.2 – All the BMDL and BMDU figures produced for the MN-RET biomarker across all compounds covering all tested days and weeks, where cisplatin, cyclophosphamide monohydrate, hydroxyurea, melphalan and vinblastine sulphate produced successful BMD CI results, and all columns highlighted "NA" are instances where a BMD was no determined as dose response curve was able to be modelled.

Row	Response	Bootstrap	Compound	3DperWx1W_MNHEP_BMDL	3DperWx1W_MNHEP_BMDU	3DperWx3W_MNHEP_BMDL	3DperWx3W_MNHEP_BMDU	7DperWx4W_MNHEP_BMDL	7DperWx4W_MNHEP_BMDU
1	0.5	200	Cisplatin	0.142	0.847	0.152	1.12	NA	NA
2	0.5	200	Cyclophosphamide_Monohydrate	NA	NA	NA	NA	NA	NA
3	0.5	200	Diethylnitrosamine	1.17	7.72	0.0279	1.5	1.9	4.34
4	0.5	200	Hydroxyurea	NA	NA	NA	NA	NA	NA
5	0.5	200	Melphalan	NA	NA	NA	NA	NA	NA
6	0.5	200	Vinblastine Sulphate	0.00777	0.112	0.0842	0.138	NA	NA

Table 5.3.2.3 – All the BMDL and BMDU figures produced for the MNHEP biomarker across all compounds covering all tested days and weeks, where cisplatin, diethylnitrosamine and vinblastine sulphate produced successful BMD CI results, and all columns highlighted "NA" are instances where a BMD was no determined as dose response curve was able to be modelled.

Row	Response	Bootstrap	Compound	3DperWx1W_MUT_RET_BMDL	3DperWx1W_MUT_RET_BMDU	3DperWx3W_MUT_RET_BMDL	3DperWx3W_MUT_RET_BMDU	7DperWx4W_MUT_RET_BMDL	7DperWx4W_MUT_RET_BMDU
1	0.5	200	Cisplatin	NA	NA	0.000016	0.359	0.012	0.0679
2	0.5	200	Cyclophosphamide_Monohydrate	NA	NA	0.000521	7.09	0.117	3.32
3	0.5	200	Diethylnitrosamine	0.0001	31.1	0.0213	15.3	NA	NA
4	0.5	200	Hydroxyurea	NA	NA	NA	NA	NA	NA
5	0.5	200	Melphalan	NA	NA	0.000862	0.705	0.00772	0.183
6	0.5	200	Vinblastine Sulphate	NA	NA	NA	NA	NA	NA

Table 5.3.2.4 – All the BMDL and BMDU figures produced for the MUT RET biomarker across all compounds covering all tested days and weeks, where cisplatin, cyclophosphamide monohydrate, diethylnitrosamine, and melphalan produced successful BMD CI results, and all columns highlighted "NA" are instances where a BMD was no determined as dose response curve was able to be modelled.

Next, section 5.3.3 highlights the toxicokinetic parameters, consisting of the Fub, Clint and LogP figures, which were produced from the ADMET predictor. These figures are used in HTTK to perform IVIVE and produce Css values, which are combined with the BMD responses produced in sections 5.3.1 and 5.3.2 to produce extrapolated AED figures for both *in vitro* and *in vivo* BMD datasets. These findings are then compared and visualised using ToxPi.

5.3.3 – Toxicokinetic parameters

Compound	CASRN	Human.Clint	Human.Fub.Plasma	LogP	MW	TK parameter Source	Css
Cisplatin	15663-27-1	3.334371478*	0.99926254*	0.84455141	300.05	ADMET /Health Canada/HTTK	4.633
Cyclophosphamide Monohydrate	6055-19-2	3.116684558	0.737890755	1.96566818	279.1	Health Canada/HTTK	9.707
Diethylnitrosamine	55-18-5	3.691455389*	0.886659209	0.56239225	102.14	ADMET/HTTK/ Health Canada	11.42
Hydroxyurea	127-07-1	5.229698993*	0.930122831*	-1.45339969	76.05	ADMET/HTTK/ Health Canada	1751
Melphalan	148-82-3	1.607293424	0.511128391	-0.24647734	305.2	ADMET/HTTK/ Health Canada	21.74
Vinblastine Sulphate	143-67-9	289.2961957	0.113989992	1.40558135	909.05	ADMET/HTTK	0.1771

Table 5.3.3.1 – Table depicting the Toxicokinetic parameters for each compound including their human Fub, Clint, logP, and molecular weight using the canonical SMILES and either the ADMET predictor, HTTK or Health Canada's *in vitro* toxicokinetic data produced in-house. Each compound's Css value was extracted using this TK information and any asterisked (*) parameters are considered unstable as their predictions were out of the domain of applicability for the model.

Section 5.3.4 and 5.3.5 highlights the *in vitro* and *in vivo* AED figures respectively, which are produced from obtaining each of these Css values described in Table 5.3.2.1, and dividing them by the BMD (AED = BMD ÷ Css) across both *in vitro* and *in vivo* BMD analyses results. These AEDs are depicted in the vertical min-max graphs below, where the Log10(AED) concentration (μM) are on the y-axis and each successful BMD producing endpoint is on the x-axis including their respective times. This was performed for both sets of *in vitro* (section 5.3.1) *and in vivo* (section5.3.2) results producing two sets of extrapolated AED datasets that are subsequently visualised using ToxPi.

5.3.4 – *In vitro* datasets extrapolated AED figures

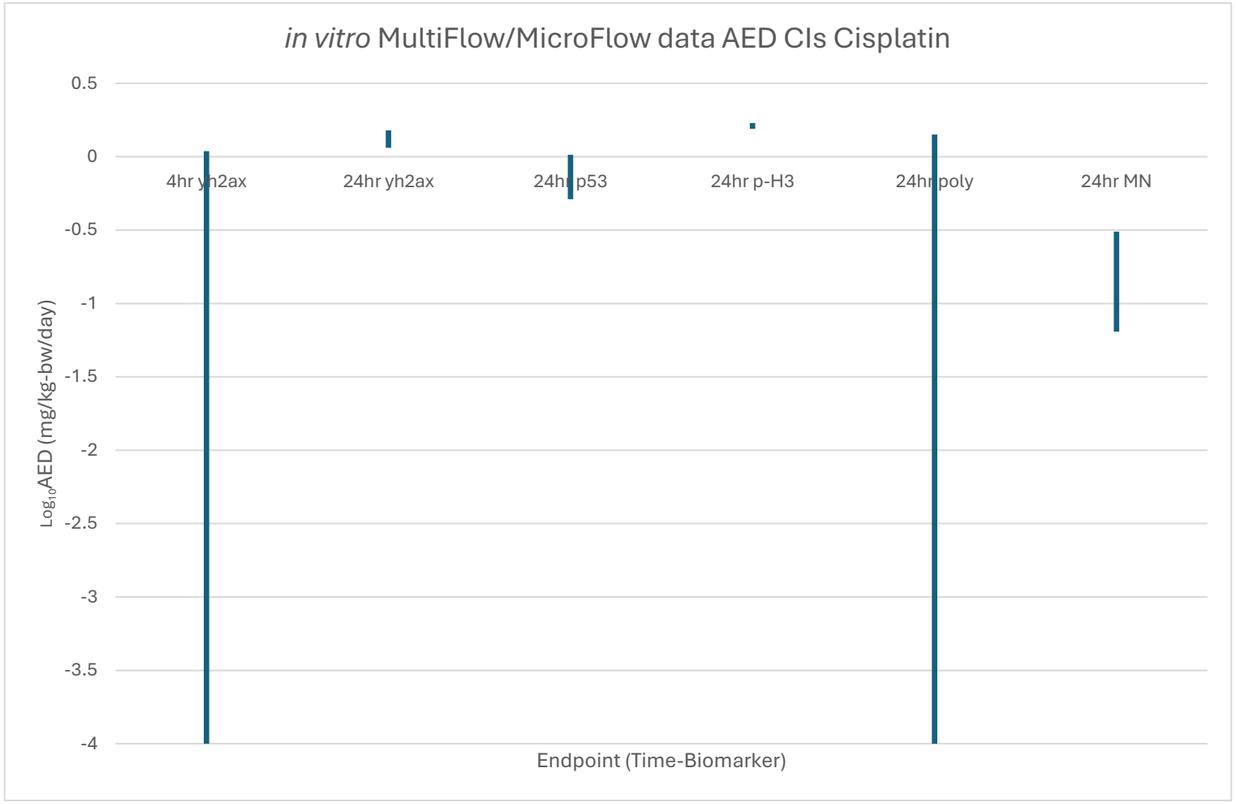


Figure 5.3.4.1 – Predicted human equivalent AED values for cisplatin based on MultiFlow and MicroFlow assay data. 90% BMD confidence intervals were generated for the indicated endpoints using PROAST with a CES of 0.5 (as outlined in sections 5.2 and 5.3.1). These BMD values were then converted to estimated human AEDs using in IVIVE. For each endpoint, the BMDL and BMDU were divided by the Css derived from HTTK modelling (using ADMET predictor-generated toxicokinetic parameters: Fub, Clint, LogP) to produce lower (AEDL) and upper (AEDU) bound values, which are illustrated on the graph, where more uncertainty surrounding the true AED is represented by larger lines. AEDs are presented in mg/kg-bw/day on a Log10 scale across the endpoints shown: 4hr H2AX, 24hr H2AX, 24hr p53, 24hr p-H3, 24hr poly, and 24hr MN.

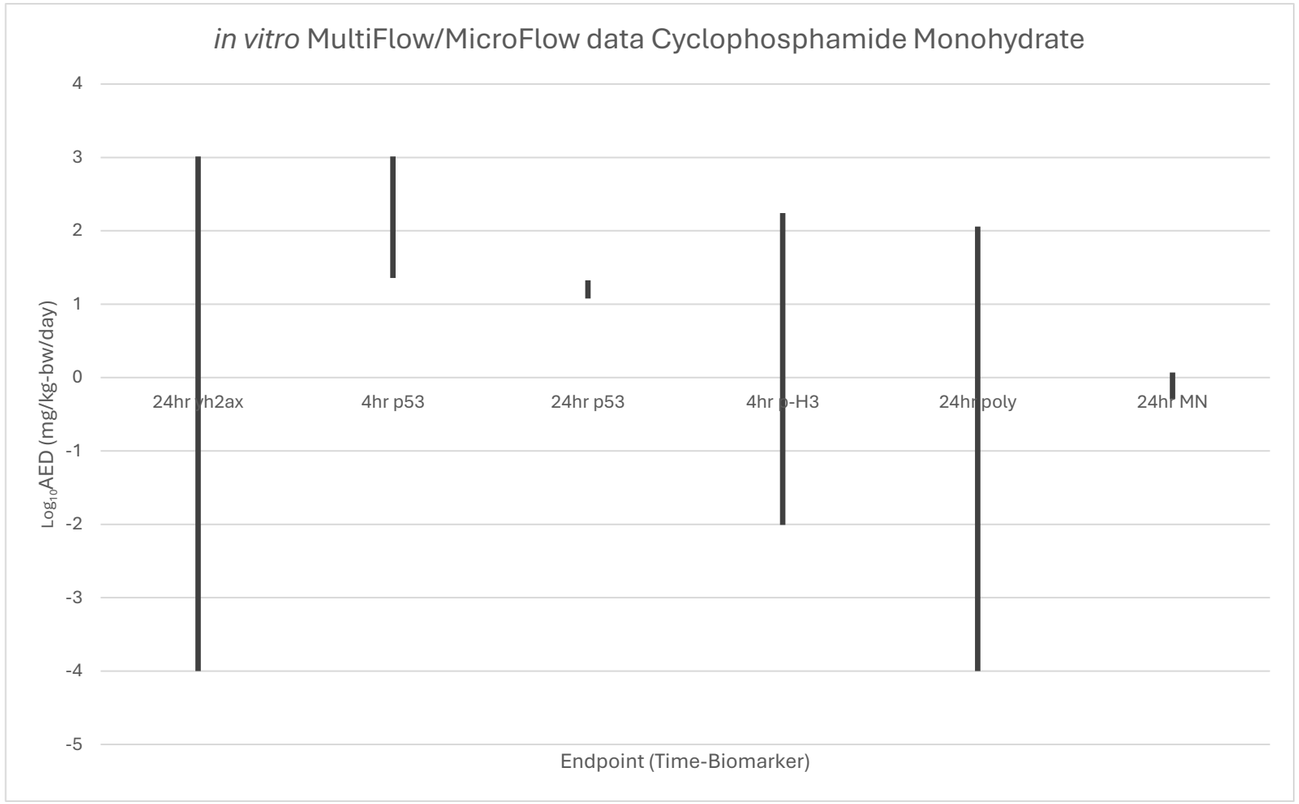


Figure 5.3.4.2 – Predicted human equivalent AED values for cyclophosphamide monohydrate based on MultiFlow and MicroFlow assay data. 90% BMD confidence intervals were generated for the indicated endpoints using PROAST with a CES of 0.5 (as outlined in sections 5.2 and 5.3.1). These BMD values were then converted to estimated human AEDs using IVIVE. For each endpoint, the BMDL and BMDU were divided by the Css derived from HTTK modelling (using ADMET predictor-generated toxicokinetic parameters: Fub, Clint, LogP) to produce lower (AEDL) and upper (AEDU) bound values, which are illustrated on the graph, where more uncertainty surrounding the true AED is represented by larger lines. AEDs are presented in mg/kg-bw/day on a Log10 scale across the endpoints shown: 24hr H2AX, 4hr p53, 24hr p53, 4hr p-H3, 24hr poly, and 24hr MN.

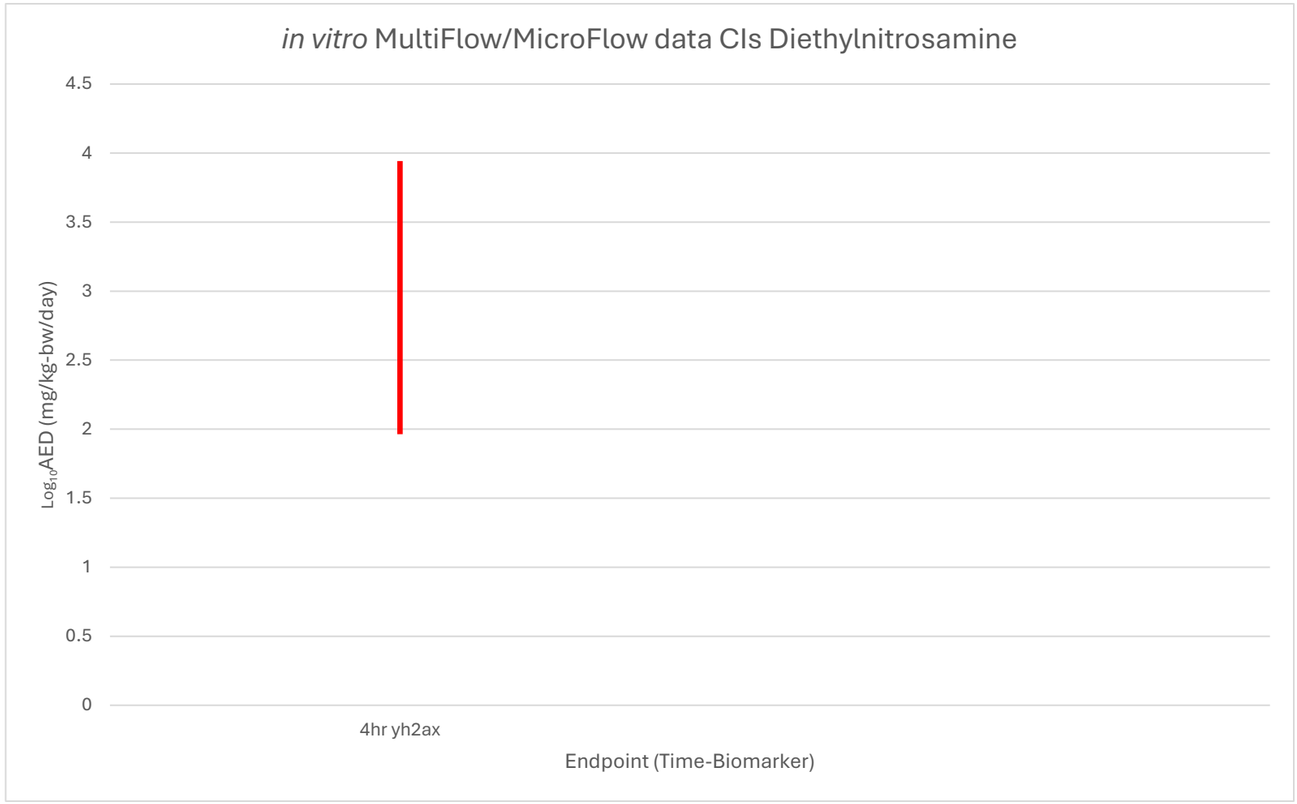


Figure 5.3.4.3 – Predicted human equivalent AED values for diethylnitrosamine based on MultiFlow and MicroFlow assay data. 90% BMD confidence intervals were generated for the 4hr H2AX endpoint using PROAST with a CES of 0.5 (as outlined in sections 5.2 and 5.3.1). These BMD values were then converted to estimated human AEDs using IVIVE. For each endpoint, the BMDL and BMDU were divided by the Css derived from HTTK modelling (using ADMET predictor-generated toxicokinetic parameters: Fub, Clint, LogP) to produce lower (AEDL) and upper (AEDU) bound values, which are illustrated on the graph, where more uncertainty surrounding the true AED is represented by larger lines. AEDs are presented in mg/kg-bw/day on a Log10 scale. Endpoint shown: 4hr H2AX.

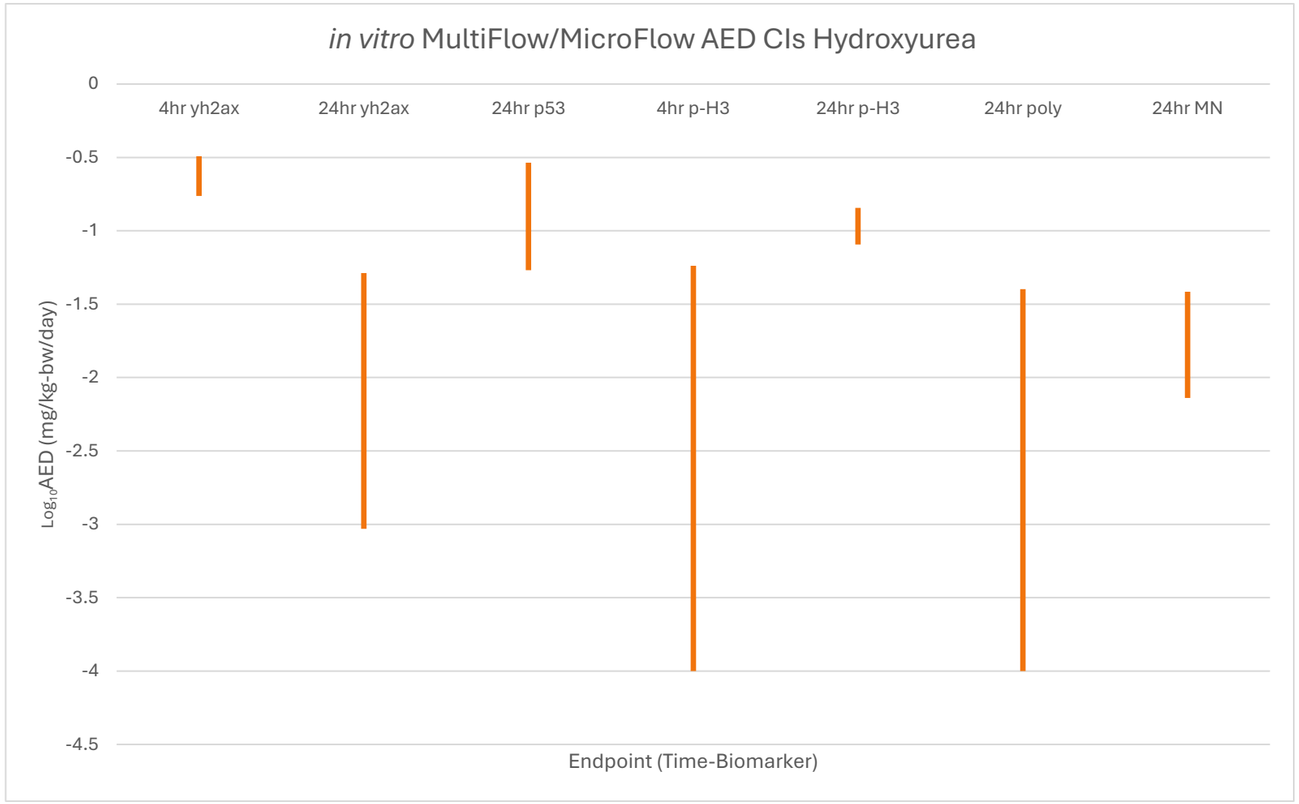


Figure 5.3.4.4 – Predicted human equivalent AED values for hydroxyurea based on MultiFlow and MicroFlow assay data. 90% BMD confidence intervals were generated for the indicated endpoints using PROAST with a CES of 0.5 (as outlined in sections 5.2 and 5.3.1). These BMD values were then converted to estimated human AEDs using IVIVE. For each endpoint, the BMDL and BMDU were divided by the Css derived from HTTK modelling (using ADMET predictor-generated toxicokinetic parameters: Fub, Clint, LogP) to produce lower (AEDL) and upper (AEDU) bound values, which are illustrated on the graph, where more uncertainty surrounding the true AED is represented by larger lines. AEDs are presented in mg/kg-bw/day on a Log10 scale across the endpoints shown: 4hr H2AX, 24hr H2AX, 24hr p53, 4hr p-H3, 24hr poly, and 24hr MN.

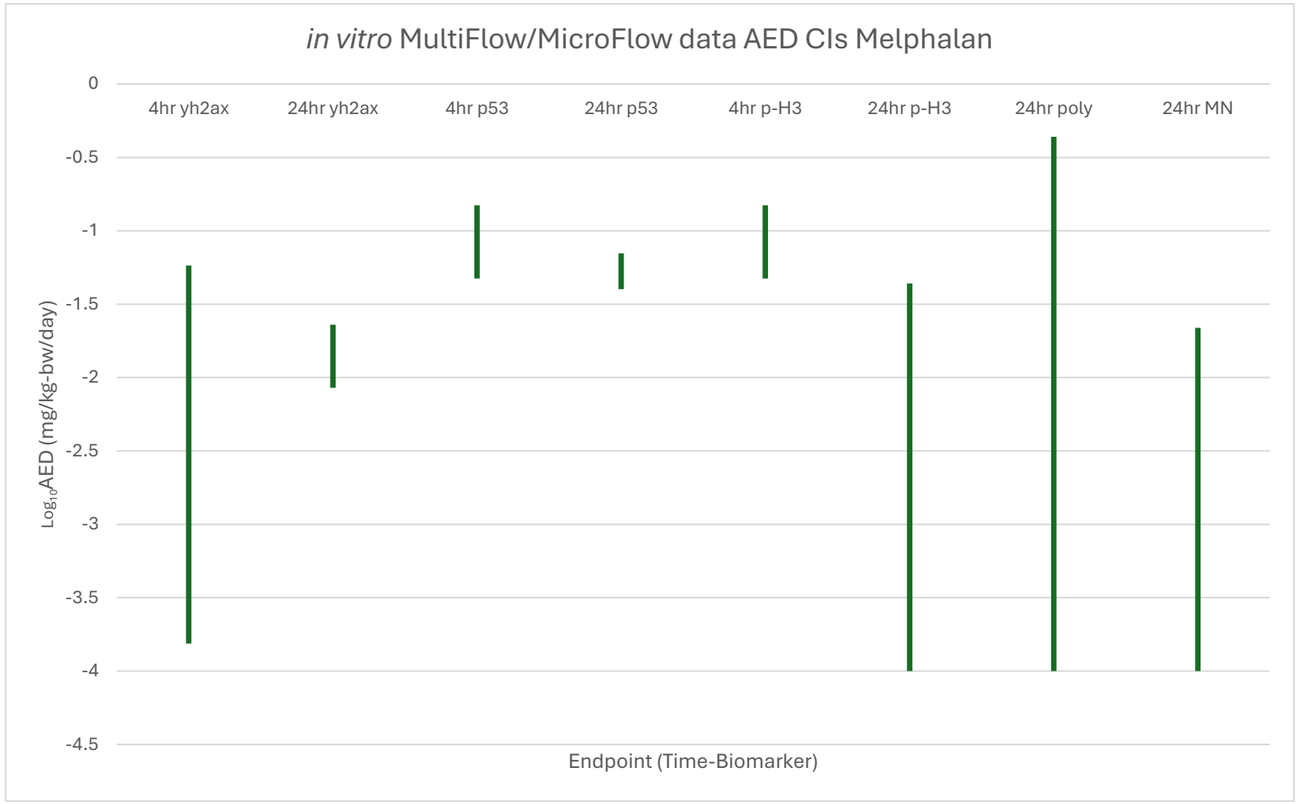


Figure 5.3.4.5 – Predicted human equivalent AED values for melphalan based on MultiFlow and MicroFlow assay data. 90% BMD confidence intervals were generated for the indicated endpoints using PROAST with a CES of 0.5 (as outlined in sections 5.2 and 5.3.1). These BMD values were then converted to estimated human AEDs using IVIVE. For each endpoint, the BMDL and BMDU were divided by the Css derived from HTTK modelling (using ADMET predictor-generated toxicokinetic parameters: Fub, Clint, LogP) to produce lower (AEDL) and upper (AEDU) bound values, which are illustrated on the graph, where more uncertainty surrounding the true AED is represented by larger lines. AEDs are presented in mg/kg-bw/day on a Log10 scale across the endpoints shown: 4hr H2AX, 24hr H2AX, 4hr p53, 24hr p53, 4hr p-H3, 24hr p-H3, 24hr poly, and 24hr MN.

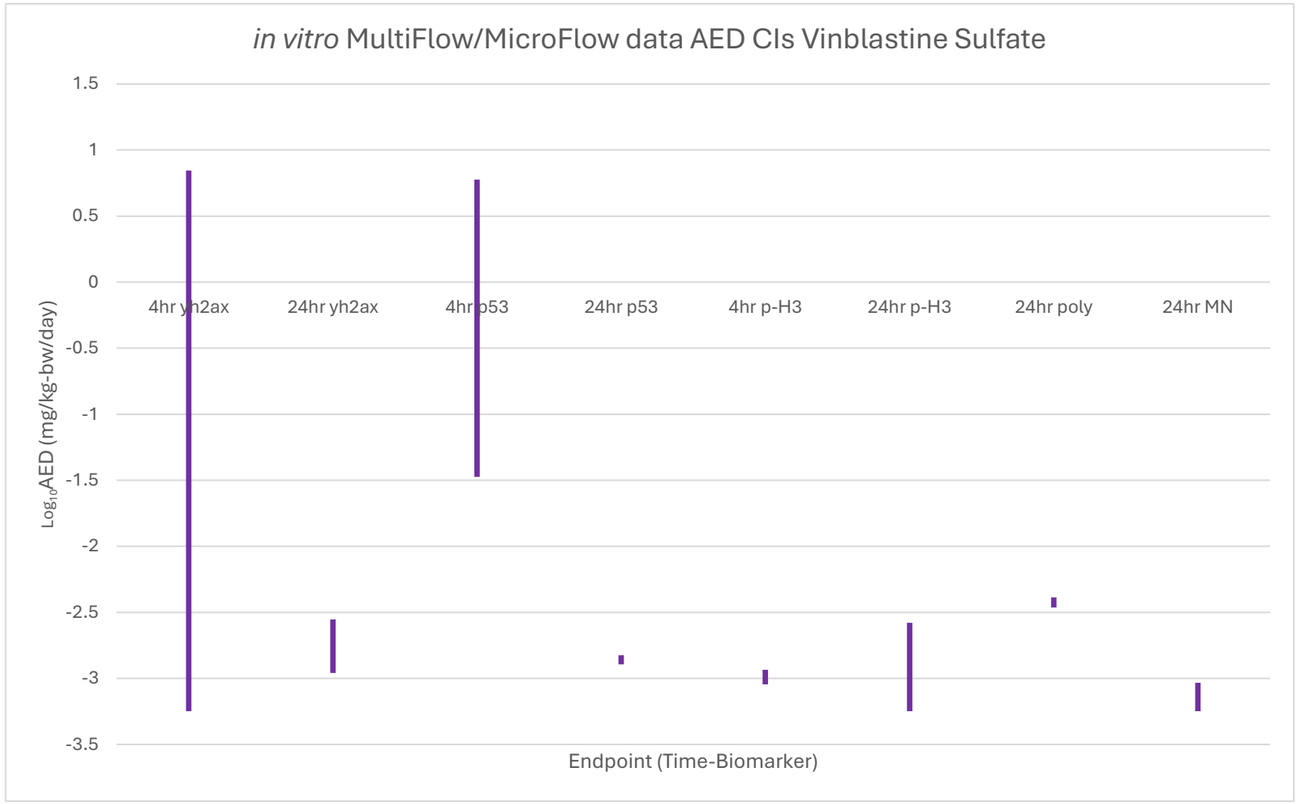


Figure 5.3.4.6 – Predicted human equivalent AED values for vinblastine sulphate based on MultiFlow and MicroFlow assay data. 90% BMD confidence intervals were generated for the indicated endpoints using PROAST with a CES of 0.5 (as outlined in sections 5.2 and 5.3.1). These BMD values were then converted to estimated human AEDs using IVIVE. For each endpoint, the BMDL and BMDU were divided by the Css derived from HTTK modelling (using ADMET predictor-generated toxicokinetic parameters: Fub, Clint, LogP) to produce lower (AEDL) and upper (AEDU) bound values, which are illustrated on the graph, where more uncertainty surrounding the true AED is represented by larger lines. AEDs are presented in mg/kg-bw/day on a Log10 scale across the endpoints shown: 4hr H2AX, 24hr H2AX, 4hr p53, 24hr p53, 4hr p-H3, 24hr p-H3, 24hr poly, and 24hr MN.

The next set of results depicts the AEDs produced using the same toxicokinetic Css values produced in section 5.3.3 alongside the *in vivo* BMDLs and BMDUs produced in section 5.3.3 which are also depicted as vertical min-max graphs to highlight the AED CIs, , where the Log10(AED) concentration (µM) are on the y-axis and each successful BMD producing endpoint is on the x-axis including their respective treatment days/weeks.

5.3.5 – *In vivo* datasets extrapolated AED figures

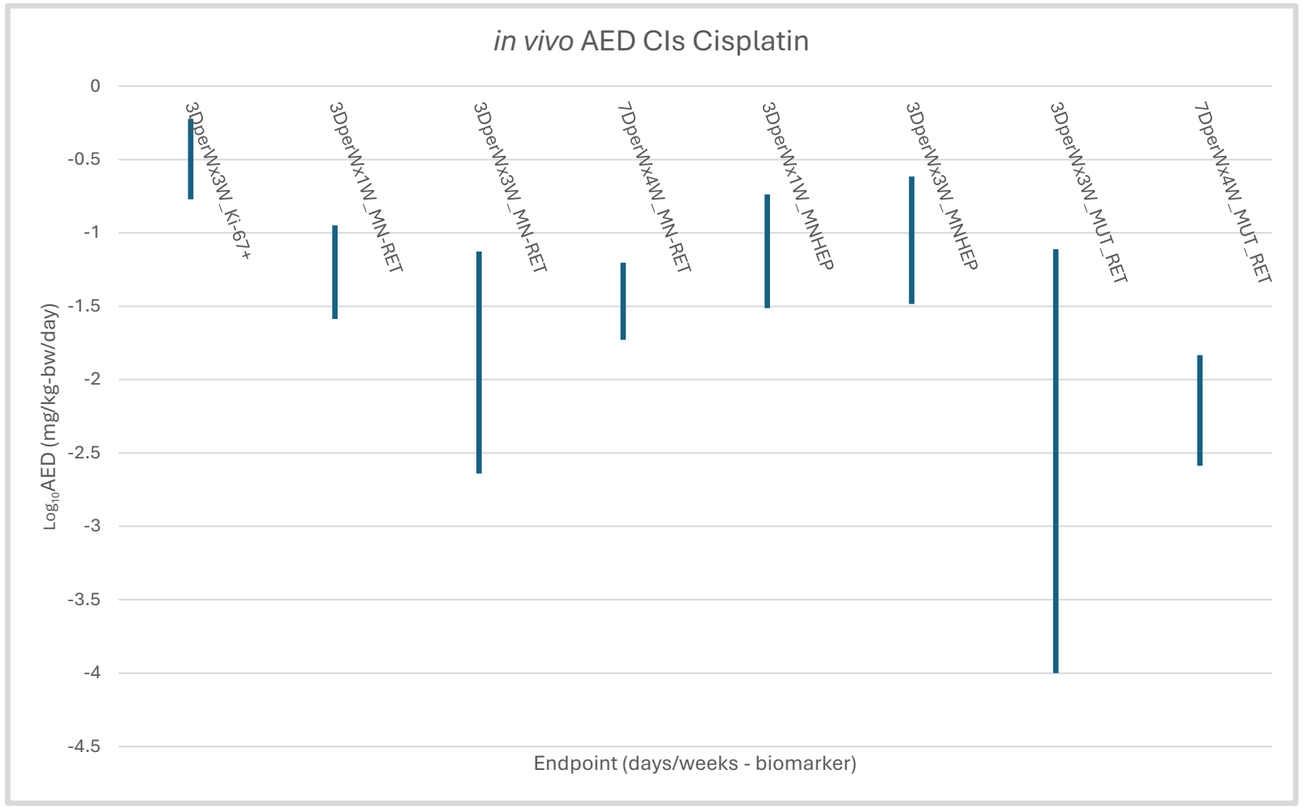


Figure 5.3.5.1 – This graph depicts the *in vivo* AED CIs in mg/kg-bw/day (Log_{10}AED) for cisplatin once 90% BMD CIs are generated using a CES of 0.5 using PROAST (as outlined in section 5.2 and 5.3.1) and each bound (BMDL and BMDU) was divided by the Css figures generated in HTTK using TK parameters to produce AEDL and AEDU values for the endpoints Ki-67+ (3DperWx3W), MN-RET (3DperWx1W, 3DperWx3W and 7DperWx4W), MNHEP (3DperWx1W and 3DperWx3W) and MUT-RET (3DperWx3W and 7DperWx4W).

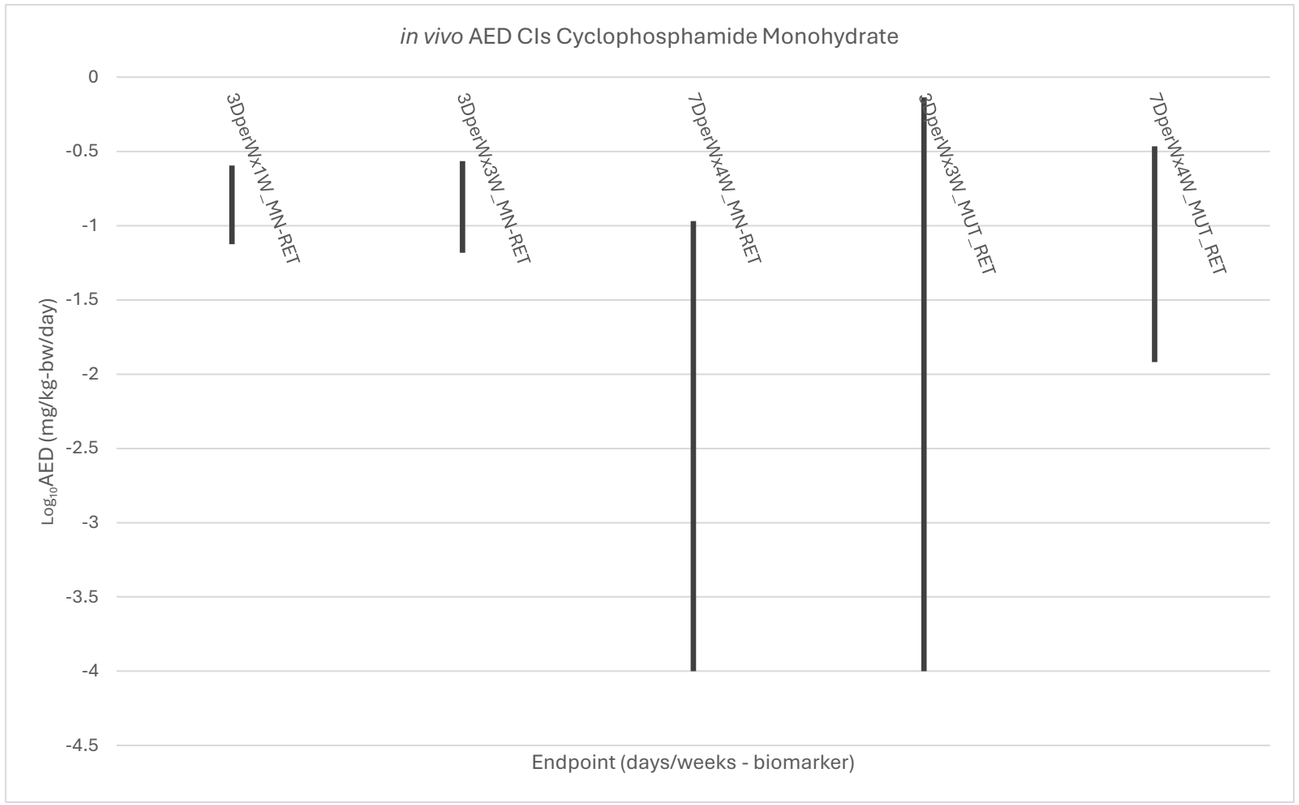


Figure 5.3.5.2 – This graph depicts the *in vivo* AED CIs in mg/kg-bw/day (Log_{10}AED) for cyclophosphamide monohydrate once 90% BMD CIs are generated using a CES of 0.5 using PROAST (as outlined in section 5.2 and 5.3.1) and each bound (BMDL and BMDU) was divided by the Css figures generated in HTTK using TK parameters to produce AEDL and AEDU values for the endpoints MN-RET (3DperWx1W, 3DperWx3W and 7DperWx4W), and MUT-RET (3DperWx3W and 7DperWx4W).

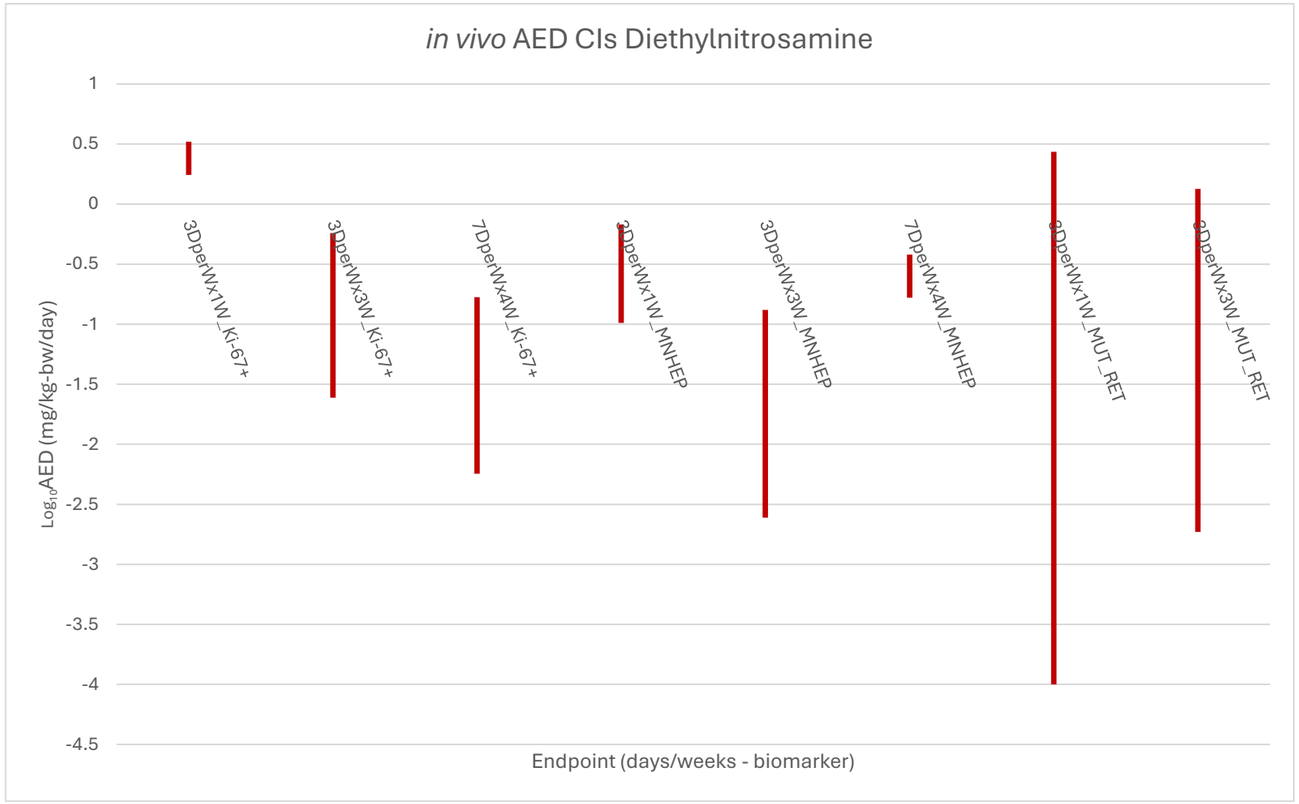


Figure 5.3.5.3 – This graph depicts the *in vivo* AED CIs in mg/kg-bw/day (Log_{10}AED) for diethylnitrosamine once 90% BMD CIs are generated using a CES of 0.5 using PROAST (as outlined in section 5.2 and 5.3.1) and each bound (BMDL and BMDU) was divided by the Css figures generated in HTTK using TK parameters to produce AEDL and AEDU values for the endpoints Ki-67+ (3DperWx1W, 3DperWx3W, 7DperWx4W), MNHEP (3DperWx1W and 3DperWx3W, 7DperWx4W) and MUT-RET (3DperWx1W, 3DperWx3W).

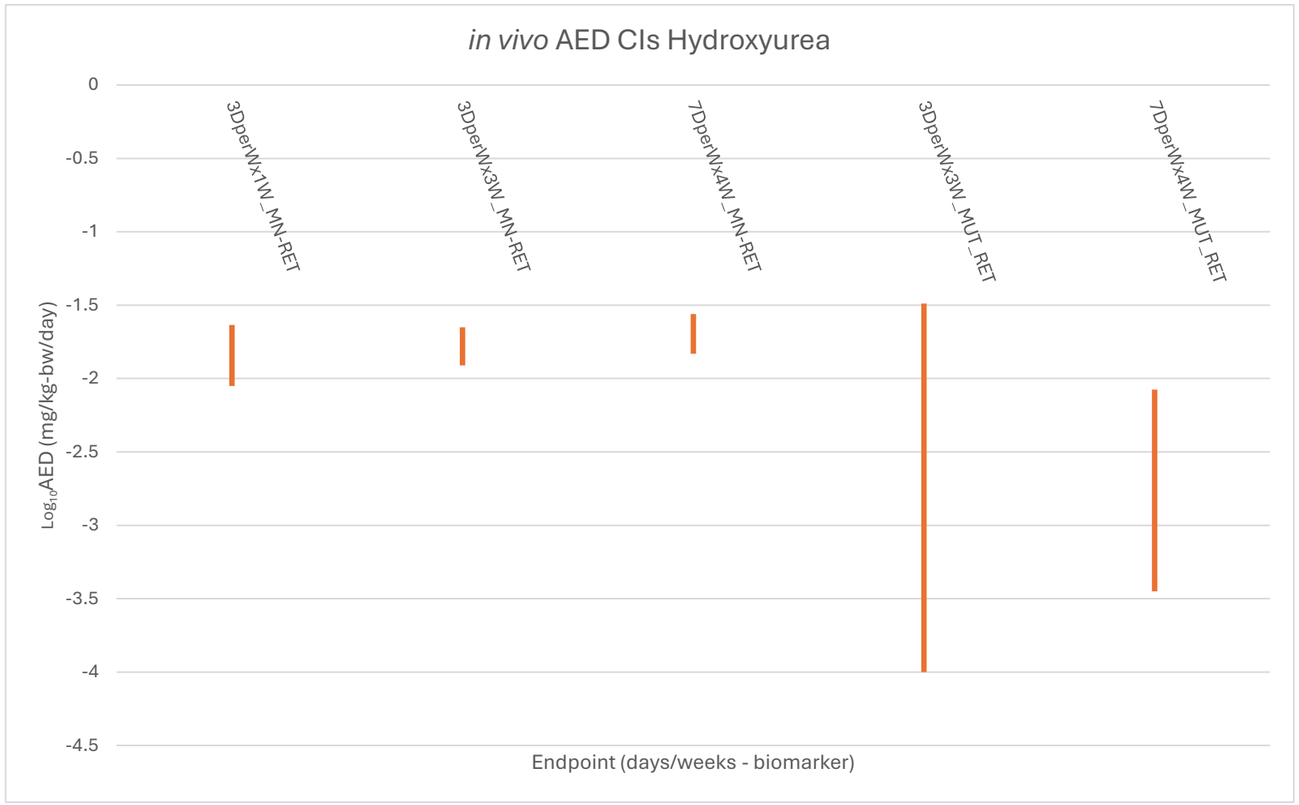


Figure 5.3.5.4 – This graph depicts the *in vivo* AED CIs in mg/kg-bw/day (Log_{10}AED) for hydroxyurea once 90% BMD CIs are generated using a CES of 0.5 using PROAST (as outlined in section 5.2 and 5.3.1) and each bound (BMDL and BMDU) was divided by the Css figures generated in HTTK using TK parameters to produce AEDL and AEDU values for the endpoints MN-RET (3DperWx1W and 3DperWx3W, 7DperWx4W), and MUT-RET (3DperWx3W, 7DperWx4W).

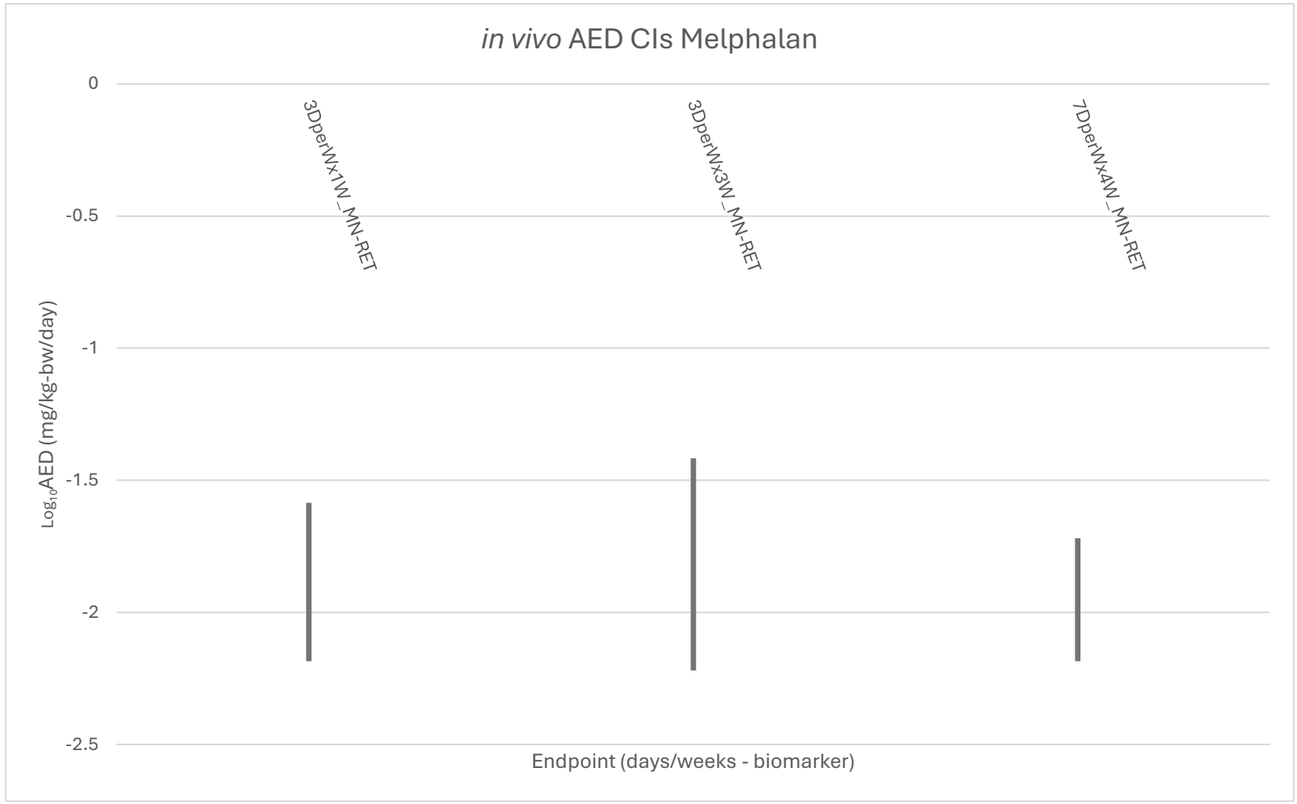


Figure 5.3.5.5 – This graph depicts the *in vivo* AED CIs in mg/kg-bw/day (Log_{10}AED) for melphalan once 90% BMD CIs are generated using a CES of 0.5 using PROAST (as outlined in section 5.2 and 5.3.1) and each bound (BMDL and BMDU) was divided by the Css figures generated in HTTK using TK parameters to produce AEDL and AEDU values for the endpoint MN-RET (3DperWx1W and 3DperWx3W, 7DperWx4W).

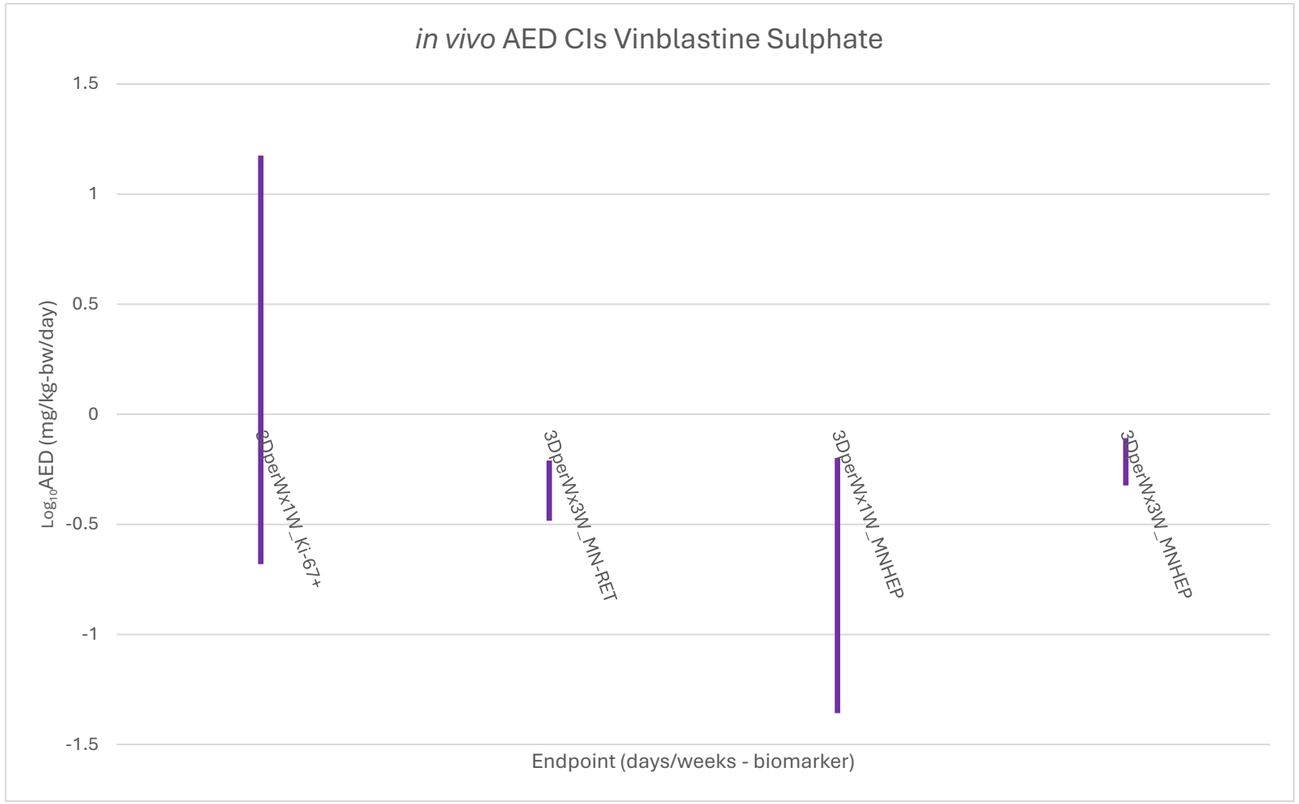


Figure 5.3.5.6 – This graph depicts the *in vivo* AED CIs in mg/kg-bw/day (Log_{10}AED) for vinblastine sulphate once 90% BMD CIs are generated using a CES of 0.5 using PROAST (as outlined in section 5.2 and 5.3.1) and each bound (BMDL and BMDU) was divided by the Css figures generated in HTTK using TK parameters to produce AEDL and AEDU values for the endpoints Ki-67+ (3DperWx1W), MN-RET (3D, 3DperWx3W), and MNHEP (3DperWx1W and 3DperWx3W).

The subsequent section encapsulates both the *in vitro* and *in vivo* visualisations for both sets of AED figures through the ToxPi framework to produce potency rankings, biomarker MoA activation profiles and MoA clustering through both divisive and agglomerative algorithms. These results attempt to contrast both extrapolated *in vitro* and *in vivo* datasets into a conglomerated set of results which can be directly compared in a simplified manner without the need for further excessive dataset transformations.

5.3.6 *In vitro* ToxPi outputs

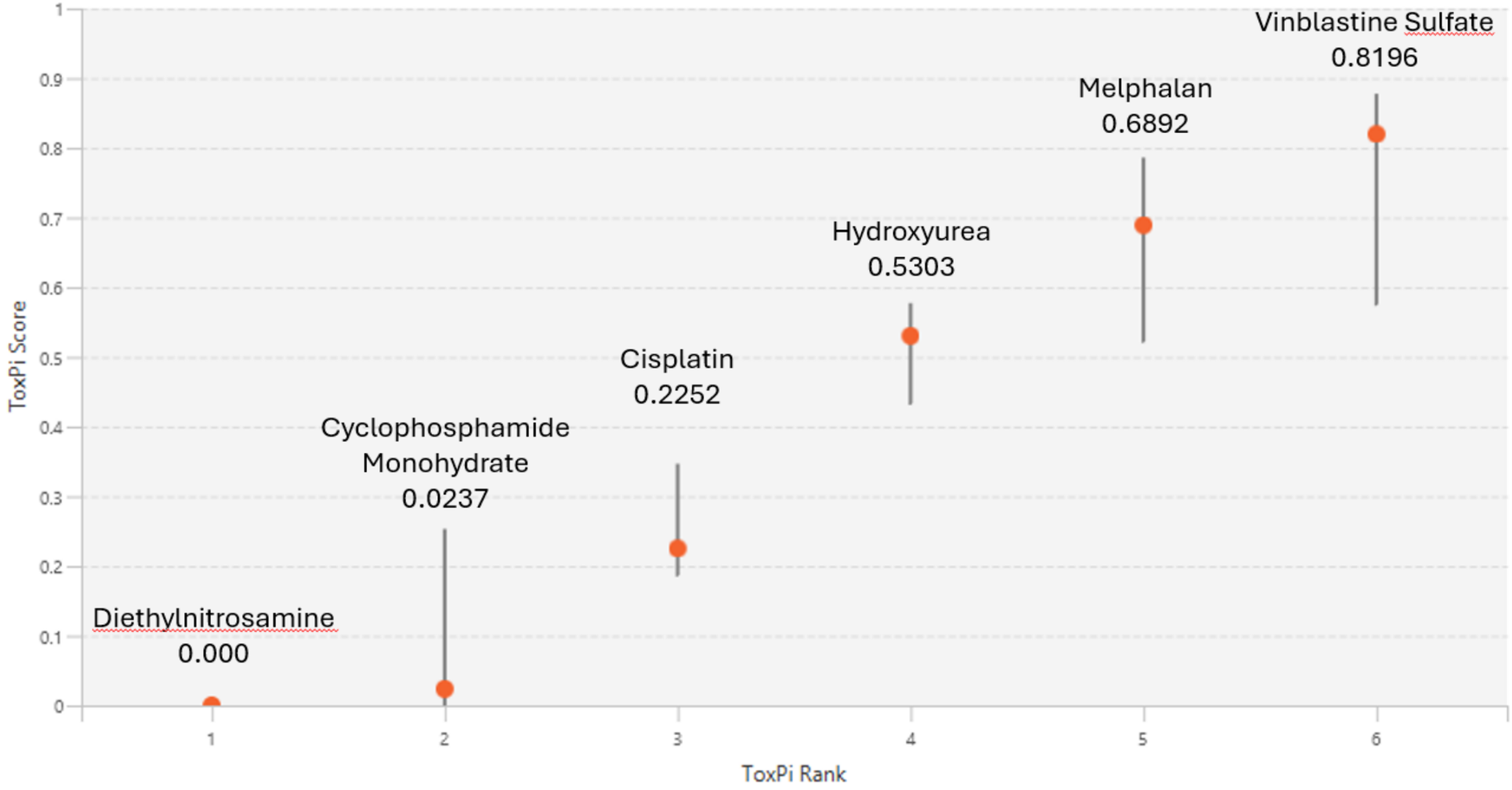


Figure 5.3.6.1 – ToxPi ranking across all 6 compounds tested in this chapter, where the *in vitro* AEDL and AEDU were used by ToxPi across all endpoints (4hr H2aX, 24hr H2aX, 4hr p53, 24hr p53, 4hr p-H3, 24hr p-H3 24hr poly, and 24hr MN) to produce a single ToxPi score utilising the algorithm described in chapter 2.2.7 (all input values were added together, subtracted by the minimum, divided by the rage, multiplied by the weighting value and the component scores were added together and ranged between 0-1), where compounds with a ToxPi score value closer to 1 are considered higher in genotoxic potency and compounds closer to 0 are considered relatively less potent.

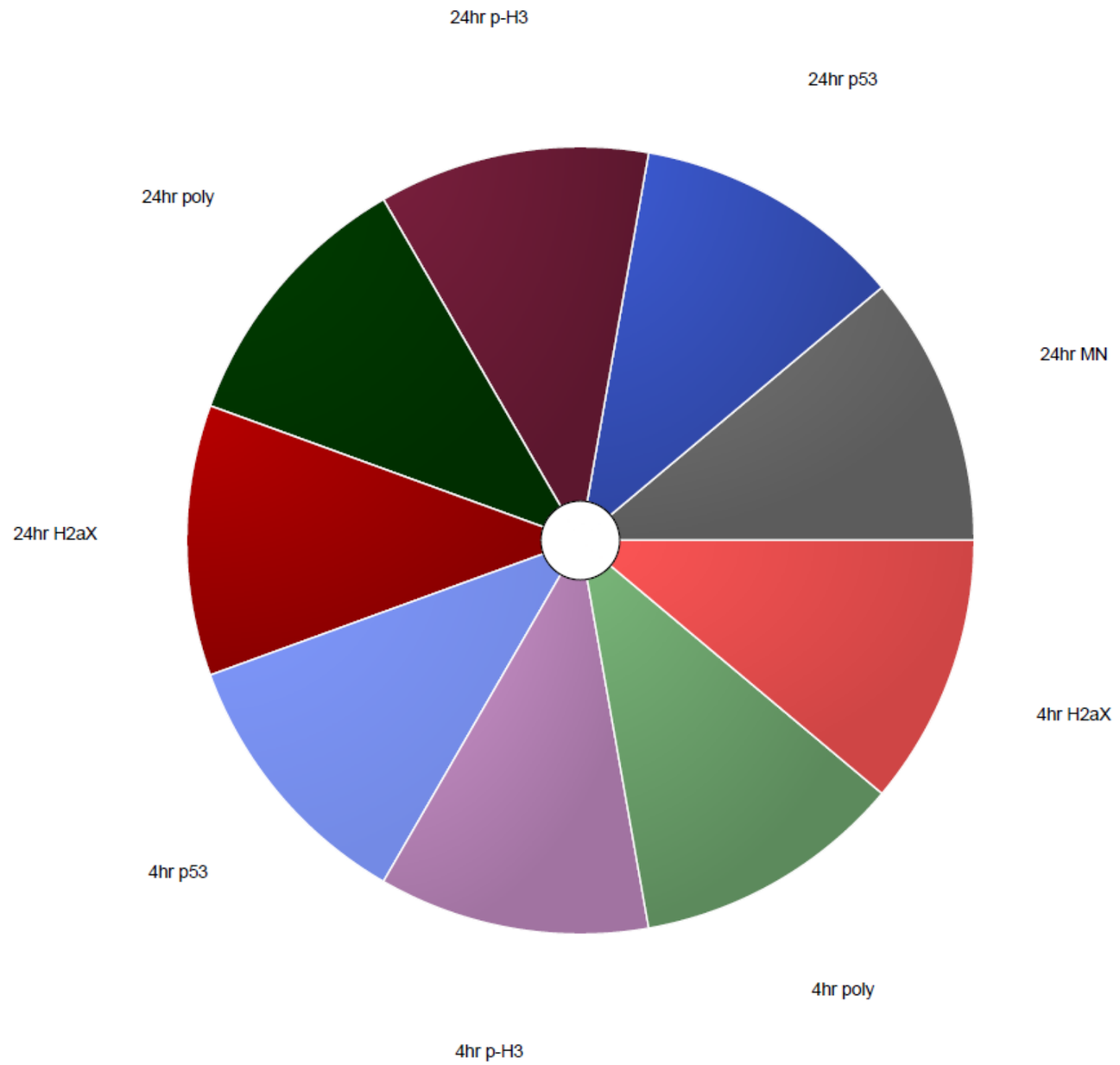


Figure 5.3.6.2 – ToxPi profile key with all slices associated with the specific *in vitro* timepoints and biomarkers, including colour coordination for endpoints H2aX, p53, p-H3, polyploidy and MN at both 4hr and 24hr timepoints.

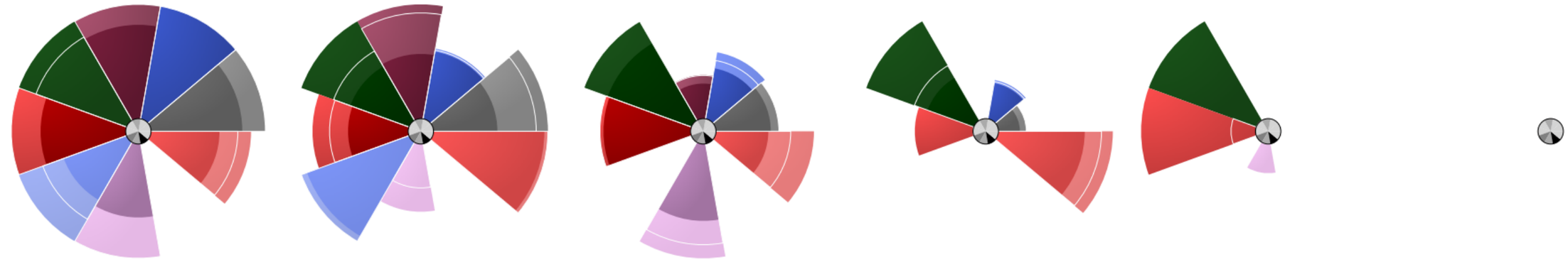


Figure 5.3.6.3 – ToxPi profiles across all tested compounds displaying each overall ToxPi and slice score each using the AEDL and AEDU to produce each endpoint for each all endpoints tested (4hr H2aX, 24hr H2aX, 4hr p53, 24hr p53, 4hr p-H3, 24hr p-H3 24hr poly, and 24hr MN) and -log10(x) transformations are applied to all slices. Additionally, the weighting remained at 1-fold for all endpoints and BMD CIs were calculated using the default 1000 bootstrap runs in ToxPi with the properties of the profile were covered in chapter 2.

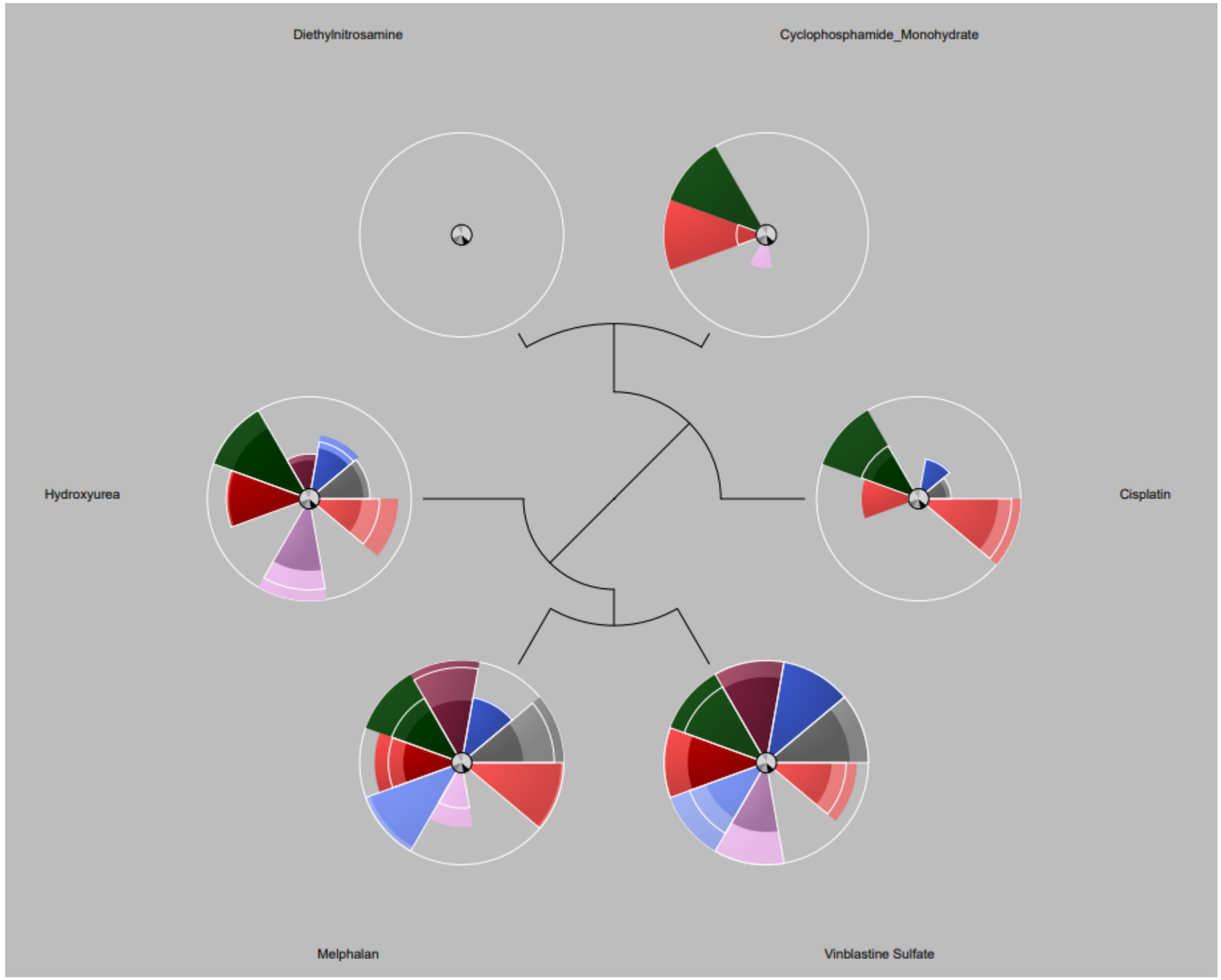


Figure 5.3.6.4 – AGNES hierarchical clustering for the *in vitro* IVIVE endpoints using the complete clustering algorithm based on ToxPi profiles produced from the *in vitro* extrapolated AED figures, which produced and identified distinct chemical groupings based on their respective similarities in MoA. Here, four distinct clusters were produced, where: cluster 1 comprises of vinblastine and melphalan, cluster 2 of hydroxyurea, cluster 3 of cisplatin and cluster 4 of cyclophosphamide monohydrate and diethylnitrosamine.

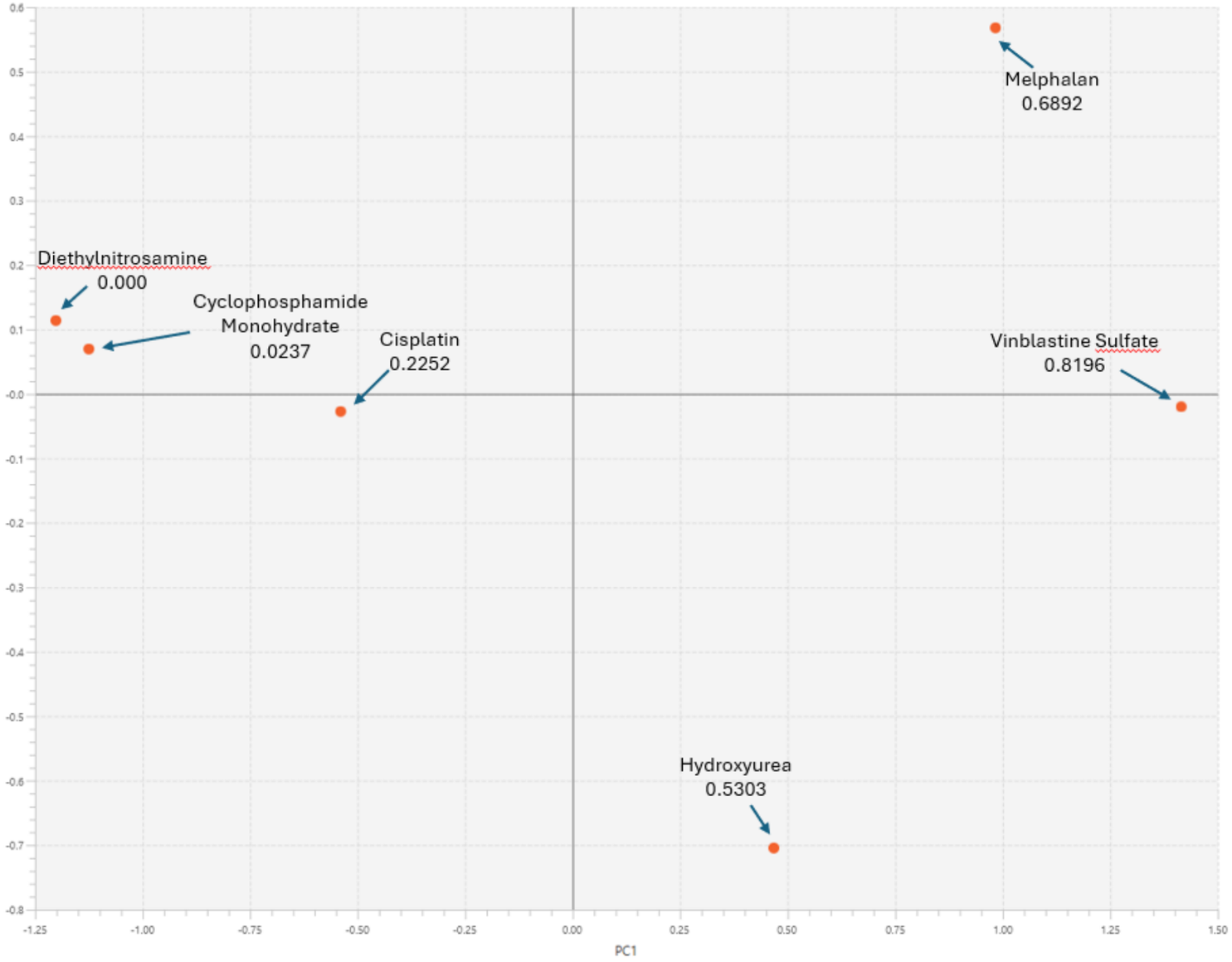


Figure 5.3.6.5 – K-means divisive/PCA clustering produced ToxPi scores, grouping compounds by genotoxicity and biomarker sensitivity. Divisive clustering revealed potency-based groupings: cisplatin, cyclophosphamide monohydrate, and diethylnitrosamine clustered closely, while melphalan grouped with vinblastine sulphate, and hydroxyurea appeared isolated with a lower PC2 value. More potent compounds showed higher PC1 values, while less potent ones had lower PC1 values.

The next section addresses the *in vivo* ToxPi outputs produced from the AEDs made in section 5.3.5.

5.3.7 *In vivo* ToxPi outputs

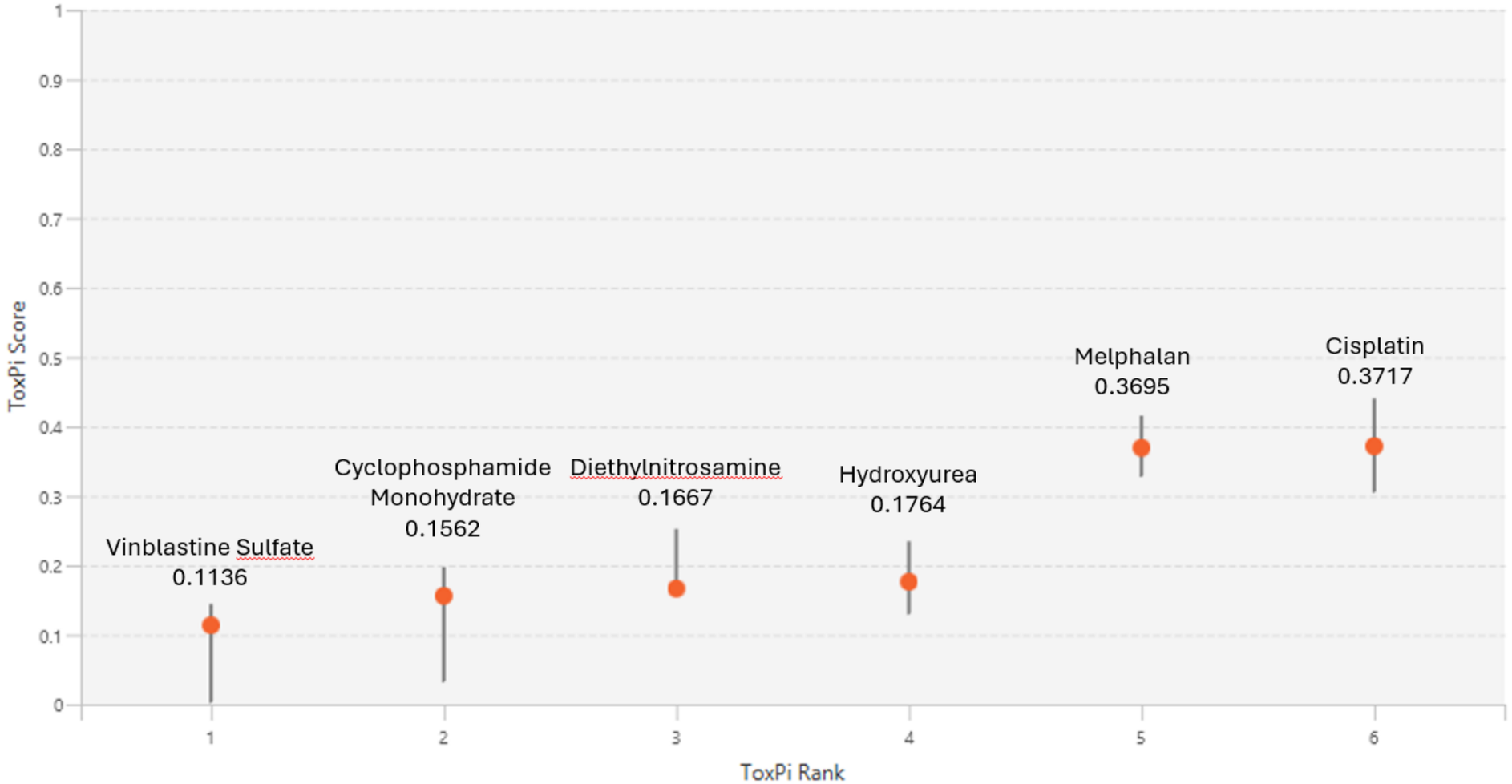


Figure 5.3.7.1 – ToxPi ranking across all 6 compounds tested in this chapter, where the *in vivo* AEDL and AEDU were used by ToxPi across all endpoints Ki67+, MN-RET, MNHEP, and MUT-RET (in conditions 3DperWx1W, 3DperWx3W and 7DperWx4W) to produce a single ToxPi score utilising the algorithm described in chapter 2.2.7 (all input values were added together, subtracted by the minimum, divided by the rage, multiplied by the weighting value and the component scores were added together and ranged between 0-1), where compounds with a ToxPi score value closer to 1 are considered higher in genotoxic potency and compounds closer to 0 are considered relatively less potent.

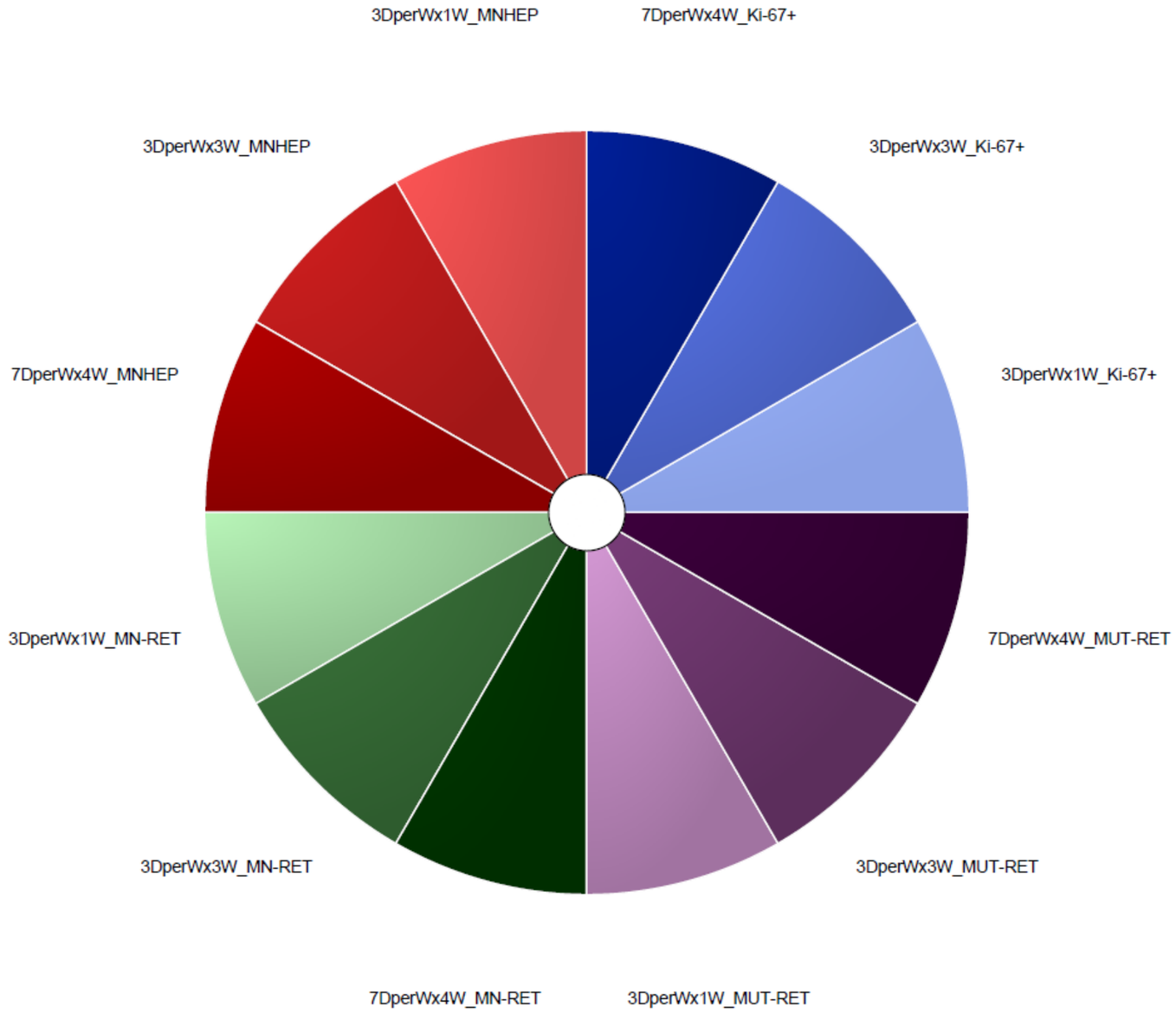


Figure 5.3.7.2 – ToxPi profile key with all slices associated with the specific *in vivo* biomarkers, including colour coordination for endpoints Ki67+, MN-RET, MNHEP, and MUT-RET at 3DperWx1W, 3DperWx3W and 7DperWx4W for each endpoint.

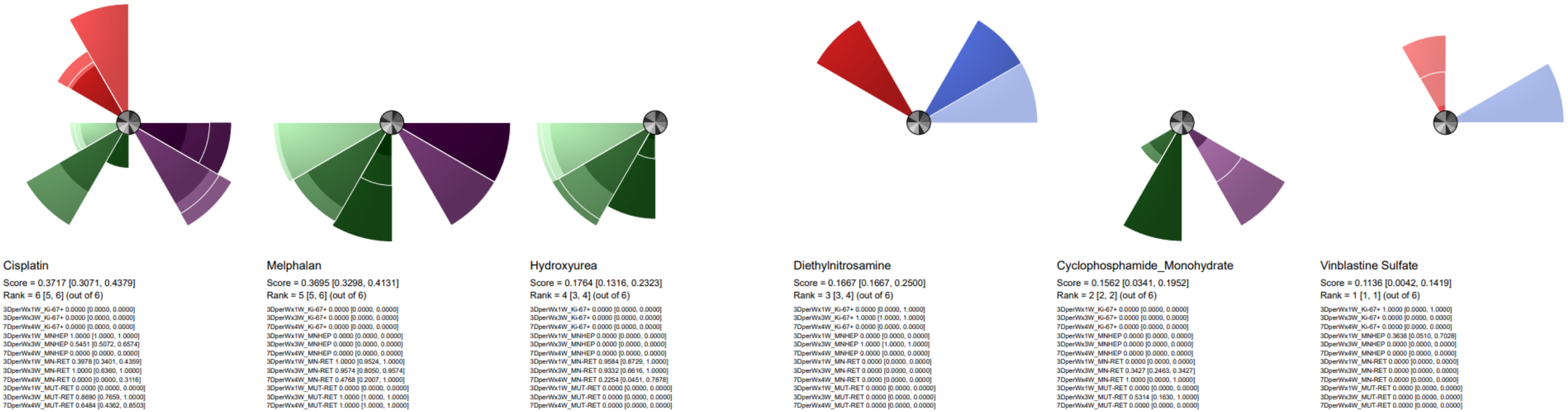


Figure 5.3.7.3 – ToxPi profiles across all tested compounds displaying each overall ToxPi and slice score each using the AEDL and AEDU to produce each endpoint for each all endpoints tested (Ki67+, MN-RET, MNHEP, and MUT-RET in either 3DperWx1W, 3DperWx3W and 7DperWx4W) and -log10(x) transformations are applied to all slices. Additionally, the weighting remained at 1-fold for all endpoints and BMD CIs were calculated using the default 1000 bootstrap runs in ToxPi with the properties of the profile were covered in chapter 2. Furthermore, the degree of greyness in the centre of each profile reflects the degree of missing data, where the darker shades represent more missing data for that ToxPi slice/biomarker.

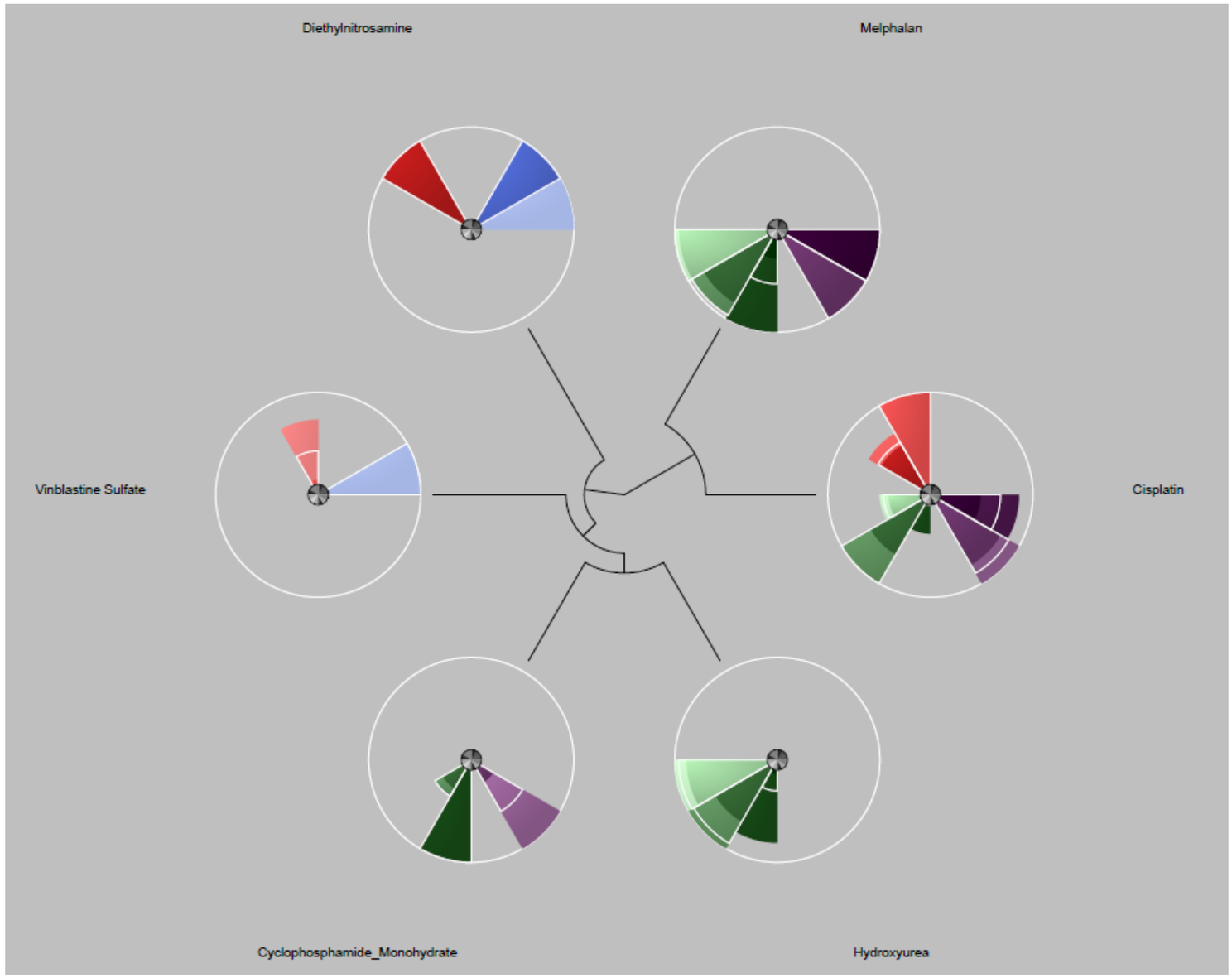


Figure 5.3.7.4 – AGNES hierarchical clustering for the *in vivo* IVIVE endpoints using the complete clustering algorithm based on ToxPi profiles produced from the *in vivo* extrapolated AED figures, which produced and identified distinct chemical groupings based on their respective similarities in MoA. Here, four distinct clusters were produced, where: cluster 1 comprises of cisplatin and melphalan, cluster 2 of hydroxyurea and cyclophosphamide monohydrate, cluster 3 of vinblastine sulphate and cluster 4 of diethylnitrosamine.

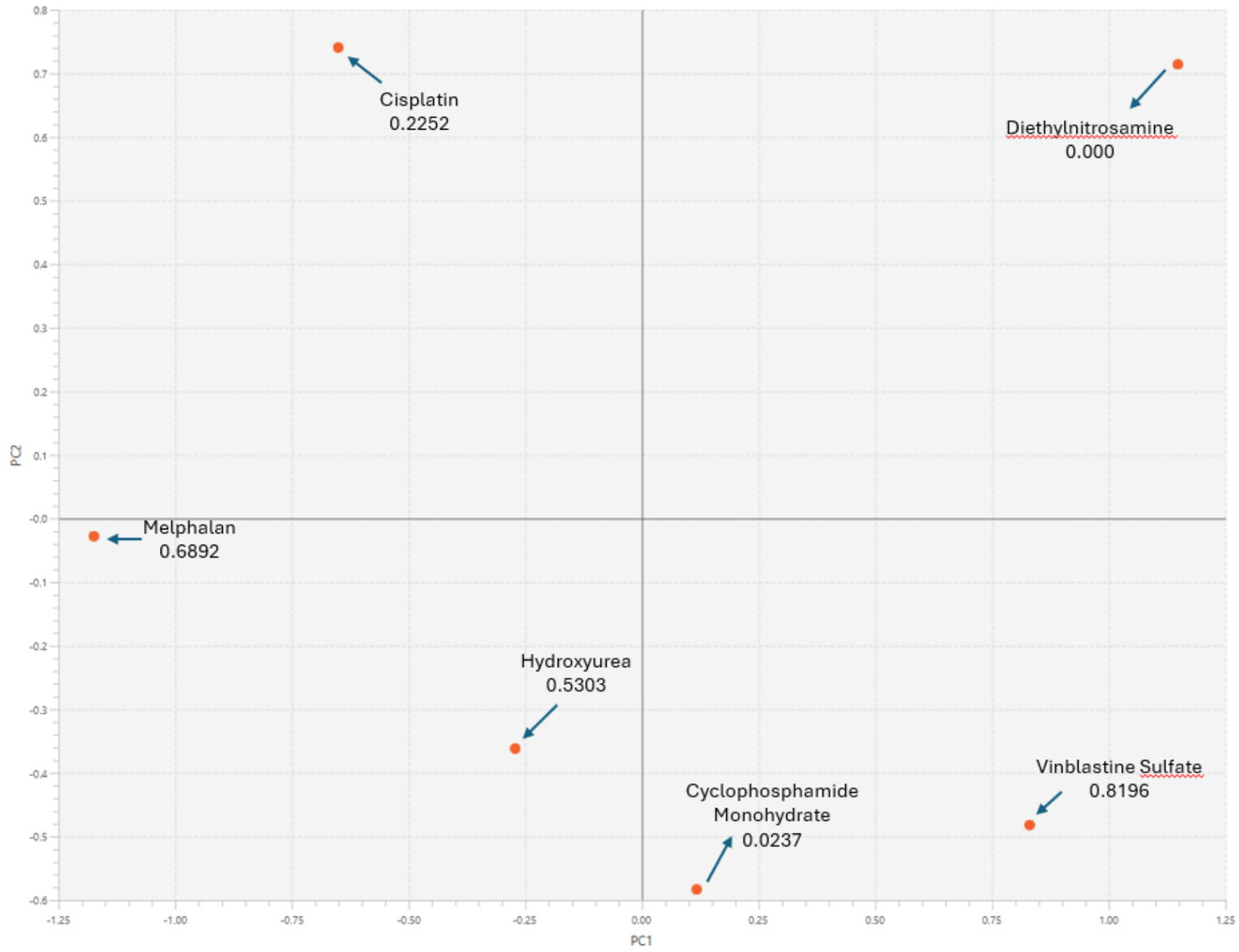


Figure 5.3.7.5 – K-means divisive/PCA clustering produced ToxPi scores, grouping compounds by genotoxicity and biomarker sensitivity. Divisive clustering revealed potency-based groupings: hydroxyurea, cyclophosphamide monohydrate, and vinblastine sulphate clustered closely, while the Euclidian distance of all other remaining compounds appear to be larger, separating out the remaining compounds from each other.

5.4 – Discussion

This section discusses and evaluates the extrapolated MultiFlow and Microflow-derived AEDs alongside the *in vivo* AED findings presented in the results (section 5.3), providing biological and mechanistic context for the observed ToxPi profiles and cross-compound comparisons of genotoxic potency. The critique among the analytical frameworks and pipelines applied across the 6 genotoxicants is scrutinised and the IVIVE outputs generated individually assessed for their ability to somewhat accurately with great consideration surrounding the initial data produced, the methodologies applied to produce each figure and the description/evaluation of any expected/unexpected findings across both *in vitro* and *in vivo* endpoints. This section additionally reviewed the ability of ToxPi to bridge the gap between finalised potency AED results for both extrapolated *in vitro* figures and *in vivo* outputs produced through the combined use of PROAST and HTTK modelling. This framework was designed to involve technology and development systems that draw towards a potential future alternative to current genotoxic studies where *in vivo* methodologies implement some degree of animal testing. This could potentially be phased out for an alternative approach where *in vitro* results could be extrapolated to replace *in vivo* methods with a high degree of accuracy through the extensive utilization of data mining, machine learning, modelling algorithms, and extensively developed databases that incorporates an increased understanding of the toxicokinetic (ADME) and toxicodynamic parameters surrounding compounds and their mechanistic properties. IVIVE continues to see further research and development to revolutionise the industrial landscape, effectively allowing for more accurate and effective analysis of new and existing compounds by introducing new approaches that uses complex and extensive databases to allow for the evaluation of chemicals for human safety while maintaining regulatory compliance and reducing reliance on animal testing systems.

5.4.1 – PROAST BMD outputs

PROAST was used to successfully model the dose-response for most *in vitro* MultiFlow and Microflow endpoints to produce BMD CIs, an excellent normalisation metric for dose response into a singular PoD region as a concentration. The figures in sections 5.3.1 and 5.3.2 display both each a collection of individual dose response curves for each compound and each BMD CIs as bands to be actively compared for each compound. PROAST has successfully used the concentration –fold change from the original *in vitro* MultiFlow and

MicroFlow datasets for each compound to accurately produce 90% BMD CI figures depicting their potencies along a curve with 200 bootstrapping runs. For the *in vivo* dataset, similarly, the Ki67+, MNHEP, MN-RET, and MUT-RET endpoints from the animal study display similar degrees of success however, the application used to generate these BMD results came from a collective effort utilising the EFSA webtool version of PROAST to produce these BMD CIs. Although, it has been demonstrated that there was slight variation in the BMD CIs produced between the R gui and EFSA webtool, they remained minimal. The webtool was used because the *in vivo* datasets were generated using a fixed protocol by Litron Laboratories (Dertinger et al., 2019), and multiple Litron contributors were involved in producing the BMD figures, necessitating their generation outside the R GUI via the EFSA website.

Although this format of analysis is strong in displaying the modelled experimental concentration – fold change data as BMD CIs to mark the PoD, the figures remain limited in their comparative capacity as there is no direct way to compare the results of one endpoint from one compound to another endpoint from a different compound within a multiplexed set of compounds that considers all datasets across all compounds. Only direct comparison of endpoints within a compound or compounds across differing endpoints could be drawn from this format of analysis, meaning that the dimensionality within the dataset is unable to be reduced, additionally complicating the evaluative and analytical component. However, these additional outputs are essential in ensuring that the raw data is correctly normalized, modelled and compared as an additional figure.

For the *in vitro* datasets, BMD CIs were unable to be determined for the 4hr poly endpoint across all compounds likely due to either no data being available or PROAST being unable to map the dose-response curve as the experimental concentration – fold change values for said endpoint was unable to establish a PoD to map the BMD CIs despite the large amount of experimental data available. The concentration limits based on cytotoxicity (80% for MultiFlow endpoints and 60% for MicroFlow endpoints) and only appear to limit very few values relative to the volume of data provided, so as a necessary step to prevent responses due to excessive cell cytotoxicity/death increasing the responses values and lowering cell counts. Additionally, these limits do not produce a significant reduction in data implemented into PROAST to map the BMD so it is highly unlikely that this would be the major contributor to a no dose-response and instead it would be likely due to experimental factors/conditions and that most compounds tested appear to demonstrate polyploidy much

later as the 24hr poly results appear to produce BMD CIs. Unusually, diethylnitrosamine, a known clastogen, produced no BMD CIs as no dose-response was mapped/recorded for most of its MultiFlow/MicroFlow endpoints, except for the 4hr yH2AX biomarker.

The *in vivo* datasets produced significantly fewer successful BMD readings due to a significantly reduced volume of quantitative data analysed and a lot of missing data within the same datasets, due to lacking/non-significant/negative trends observed in the experimental dataset or data being unavailable. This reduction in data assessed would greatly contribute to effects that are displayed in other figures when transformed and visualised, which will be discussed later in this section. However, as no BMD CIs figures are available due to the BMDs being produced, no original PROAST dose-response curves were directly available for analysis as only the BMDLs and BMDUs were recorded. This also explains why hydroxyurea only activates the MN-RET endpoints across all days/weeks and melphalan only activates the MN-RET and MNHEP endpoints at 3dx1w and 3dx3w readings.

5.4.2 – Toxicokinetic parameters and *in vitro*/*in vivo* AED figures

Both sets of BMD CIs were transformed to concentrations between 0.0001-10000 μM for visual scaling purposes and used to successfully produce AED figures which were graphically displayed in section 5.3.3. The length of all bars indicated the 90% CI, where wider bars indicate heightened uncertainty surrounding the true AED value due to larger disparity between the upper and lower bounds of the AED, behaving similarly to the BMD CIs. The Css values derived for each compound used to produce the AED figures were extracted using the Clint, Fub and LogP variables to calculate them. Additionally, the Monte Carlo simulations were performed to take inter-individual variation within the human populations for more accuracy in Css generation by simulating results from the 95th percentile. However, some of the TK parameters were considered unstable by Health Canada when extracted (TK parameter predictions were outside the domain of applicability for the model), resulting in potential variation surrounding the Css produced. For example, cisplatin and hydroxyurea both produced unstable Clint and Fub while diethylnitrosamine produced an unstable Clint value. This would naturally lead to downstream effects on ToxPi outputs as these variables greatly contribute towards the AED for each compound/endpoint. Additionally, from observation, the LogP variables are all approximates/predictions, which produce variation amongst their values produced based on the source used and compound information

available, which for the purposes of this investigation, was the ADMET predictor (unless compound data was unavailable).

Chemical-specific physiology based pharmacokinetic (PBPK) modelling for all compounds were considered for application to alleviate concerns surrounding the unstable TK parameters and support the IVIVE modelling, however after extensive discussion with Health Canada, it was concluded that the use of PBPK modelling was unsuitable for this exercise due to the insufficient data being analysed within the bounds of this investigation, especially surrounding lower doses, and the lack of model validation surrounding the *in vivo* data. However, an increase in datapoints used would facilitate the use of PBPK modelling, however as it is highly complex in its computational capacity due to its resource-intensive nature, already demonstrating its challenging implementation on most IVIVE datasets.

Within the original Beal et al., (2022) study, the *in vitro* genotoxicity data used yielded AEDs that are equal or lower than the established animal model PODs for 65% of the chemicals tested. These chemicals had extensive supporting data from well-established databases. In contrast, this study used two independently generated datasets—one *in vitro* and one *in vivo*—rather than comparing the *in vitro* results to a robust, pre-existing *in vivo* database. Additionally, the *in vivo* dataset in this study had a lower rate of successful BMD modelling. These factors likely account for the variance between this study and Beal's findings, as the limited scope of the *in vivo* data may have reduced the overall success in PoD modelling.

5.4.3 – ToxPi Figures

Upon the extraction of AED CIs bounds, the figures are re-transformed to comply with the visualisation limits within the 0.0001-10000 μM within the bounds of the biologically tested range. Although this step was necessary for visualisation purposes, the implementation of further data transformations, particularly those that scale the original findings could potentially introduce bias and loss of procedural transparency as constant changes to the experimental data could lead to further confusion, even if the original transformations were implemented to provide additional transparency around the experimental data. This would raise concerns surrounding the necessity of such transformations as the additional complexity may hinder the audience from fully understanding the process taken to visualise these datasets, especially when it could potentially be seen as a distortion of dose-response relationships around a complex subject that implements machine learning modelling and simulations, such as IVIVE. However, ToxPi figures are supplementary to the original dose-

response models generated to aid visualisation and provide additional transparency of these complex datasets, not a total substitute, hence the implementation of this transformation remains necessary to preventing excessive skewing of the datasets towards the upper/lower bounds and providing a balanced approach that accurately depicts the endpoint projections accurately describing compound potency across a series of biomarkers using the AED CIs.

Additionally, for both *in vitro* and *in vivo* results, each endpoint slice was built using the AEDL and AEDU for each endpoint and $-\log_{10}(x)$ transformations were applied to all slices to aid in scaling the data and ensuring that more potent compounds with lower AEDs would appear as larger projections, similarly to the BMD approach. However, both *in vitro* and *in vivo* datasets do not share identical endpoints, instead they have endpoints that are similar meaning potency overall will vary due to the differences in MoA, meaning the comparative capacity of endpoints remains limited, placing higher emphasis on comparing compounds based on overall potency and clustering. Furthermore, although these *in vitro*/*in vivo* endpoints share similarities, specific biomarkers *in vivo* will not be identical to *in vitro* outputs despite some level of overlap as their precise MoAs remain different and therefore result in differences in detection within the compounds assessed as their ToxPi profiles would greatly differ as demonstrated in the ToxPi profiles produced in figures 5.3.4.3 and 5.3.6.3. For example, cisplatin demonstrates significant ToxPi slice endpoint projections MNHEP *in vivo* and γH2aX *in vitro* biomarker, however, the endpoints themselves are not identical DNA damage detection endpoints, despite their overlapping MoA (γH2aX being linked to dsDNA damage while MNHEP provides a broader assessment of chromosomal instability, both linked to micronuclei production). Specifically, larger projections are portrayed by the 3dx1w and 3dx3w MNHEP endpoints *in vivo* and similarly to the 4hr γH2aX endpoint (24hr γH2aX had a large bootstrap CI distance, with a much higher upper bound). Additionally, some tissue specific responses *in vivo* may not model similarly *in vitro* as the *in vivo* involves some degree of pharmacokinetics and (ADME) would affect exposures and final dose-responses, whilst the *in vitro* extrapolated model considers these factors through HTTK modelling, hence why further tests with additional compounds are required to establish higher consistency among both testing systems.

Despite ToxPi successfully displaying results among both datasets, another difference between the *in vivo* and *in vitro* outputs is that less data was available for the *in vivo* findings, hence their relatively lower ToxPi scores compared to the *in vitro* results, as the current pipeline means that less data would increase the greying within the centre of each ToxPi

profile and result in lower projections as more endpoints having no results would lead to an overall lowering of ToxPi scores across compounds overall. Successfully generated BMD/AEDs may contribute towards higher ToxPi scores as missing data appears to be the largest contributing factor towards lower ToxPi scores, which becomes very apparent for compounds such as diethylnitrosamine, which only successfully produced a single BMD/AED CIs with 4hr h2ax, hence its significantly low ToxPi score of 0.0000 *in vitro* however, as more endpoints had positive results *in vivo* hence, its higher toxpi score of 0.1667. However, despite this example, this has affected *in vivo* findings more as less available data from more missing information led to overall data reduced ToxPi scoring *in vivo* affecting the endpoint projections and overall scoring of compounds overall, likely increasing the disparity of *in vivo* findings compared to *in vitro* outputs.

Another compound that illustrates this is vinblastine sulphate, which activates all *in vitro* endpoints tested (4hr H2aX, 24hr H2aX, 4hr p53, 24hr p53, 4hr p-H3, 24hr p-H3 24hr poly, and 24hr MN), however, only appears to activate 3dx1w ki67+, 3dx3w mn-ret, 3dx1w/3w MNHEP endpoints *in vivo.* This difference resulted in vinblastine sulphate producing the highest potency scoring *in vitro* and the lowest *in vivo*, a significant and unexpected outcome from each set of conditions, despite the idea that they should display similar levels of aneugenicity across both *in vitro* and *in vivo* assays. It produced low AED figures across both assays, so a higher potency is expected, however, due to a lack of endpoint recordings *in vivo* from irrelevant positives, no significant trend findings for some doses and re-optimization top doses based on 3dx1w results overall reducing the data from the *in vivo* findings as those values are decoded as “no dose-response”, further removing data points from the analysis. So, even though vinblastine sulphate is a potent compound, potential hypertoxicity resulted in less endpoint recordings in the original dataset, thereby correlating with a reduction in ToxPi scoring less outputs were produced *in vivo*. Despite this major limitation in comparison across both assays, it further opens opportunities for additional assessments needed for further develop IVIVE studies.

For the hierarchical cluster modelling, it was clear to see that clustering across both assays were greatly affected by the large volume of *in vivo* missing data has skewed the clusters away from groupings based on shared/related MoA. This is likely due to the reduced variation in MoAs assessed as only 6 compounds were analysed, hence, explaining why vinblastine sulphate would cluster with clastogens *in vitro* and to a lesser extent *in vivo* as demonstrated in figures 5.3.4.4 and 5.3.6.4 respectively. This would also likely explain the

lack of distinct k-means clusters forming within both assays since this lack of compounds and gaps within the data are likely to explain the sporadic segregation among compounds, especially when regarding the *in vivo* datasets. *In vitro*, cisplatin, cyclophosphamide monohydrate and diethylnitrosamine remain within close proximity whilst melphalan, hydroxyurea and vinblastine sulphate remain separate. Meanwhile, *in vivo*, all compounds are sporadically separated (with cyclophosphamide and hydroxyurea displaying some degree of proximity) with no clear patterns/clusters being identified.

Additionally, *in vitro,* a lack of readings in the 4hr poly endpoint *in vitro* would lower the ToxPi score somewhat and prevented vinblastine sulphate from clustering away from the other clastogens, however, the vinblastine profile remained similar to clastogens with relatively higher potency, which likely explains the clustering and PBPK modelling considerations were attempted to alleviate this (difference) to no avail for reasons described earlier (lacking *in vivo* data but PBPK stable and very low *in vitro* compared to *in vivo*). Additional compounds with an increased dataset values for both clastogenic and anuegenic readings may alleviate these concerns, encouraging the further development of IVIVE investigations which explore this aspect.

5.4.4 – Improvements

The results and methodologies presented in this study highlight several areas for improvement to enhance the accuracy and applicability of IVIVE frameworks. Firstly, the limitations in BMD modelling, particularly the inability to generate BMD CIs for certain endpoints (e.g., the 4hr poly endpoint) and the reduced success rate for *in vivo* datasets, suggest a need for improved experimental protocols and data collection. Addressing these gaps, through the inclusion of additional successful endpoints and compounds, could involve optimising assay conditions, increase data points, and ensure consistent data availability across endpoints. Additionally, the use of a CES of 0.5 for *in vivo* BMDs, while necessary for generating IVIVE to produce AED figures, may introduce challenges in PoD determination, as it may not always align with biologically relevant effect sizes. This could lead to potential uncertainties in the *in vivo* findings.

Furthermore, the instability of certain toxicokinetic (TK) parameters, such as Clint and Fub, underscores the importance of refining TK predictions through more robust modelling approaches, such as PBPK modelling, once sufficient data and validation are available. This would improve the reliability of AED calculations and downstream ToxPi analyses. The

discrepancies between *in vitro* and *in vivo* ToxPi scores, as seen with compounds like vinblastine sulphate, emphasise the need for better alignment of endpoints across testing systems. Expanding the dataset to include more compounds and endpoints, particularly those with well-defined MoA, could improve clustering accuracy and reduce the impact of missing data. Finally, enhancing transparency in data transformations and visualisations, such as the -log10(x) scaling applied in ToxPi, would ensure clearer communication of results and reduce potential biases. These improvements would strengthen the IVIVE framework, making it a more reliable alternative to traditional animal testing for genotoxicity assessments.

5.5 – Chapter Conclusions

Overall, IVIVE successfully displayed the capacity of *in vitro* datasets capable of being translated into *in vivo* figures. The *in vitro* analysis being performed using rat cells/liver was combined with Monte Carlo simulations to account for human population variability to model the *in vivo* Css values, which were used to find the AED figures. These AED values were subsequently visualised to enabling the direct comparisons of extrapolated *in vitro* figures and comparing them to empirical *in vivo* readings. Collectively, these methodologies highlight the successes and challenges surrounding the application of pharmacokinetic and pharmacodynamic variables to perform HTTK modelling for AED prediction, which advances IVIVE and PoD determination for genotoxicity assessment across the 6 compounds analysed.

These pipelines require the considerate implementation of both mathematical and biological systems and computational algorithms to successfully perform IVIVE modelling, making them inherently complex due to their reliance on computational modelling, toxicokinetic analysis and visualisation approaches using tools such as ToxPi. However, emphasis is placed on the approach to produce visualised outputs that accurately interprets dose-response values *in vitro* and accurately depicts them *in vivo* to address limitations surrounding traditional toxicological approaches that place heavy reliance on animal testing.

The original Beal et al., (2022) highlights how IVIVE implementation produced accurate results with 65% correctness demonstrating that while the system shows promise, IVIVE remains imperfect and under development. Additionally, high levels of dataset transformations may reduce consistency between *in vitro* and *in vivo* potency rankings, profiling, and clustering, highlighting the need for further refinement to enhance direct extrapolation capacity. The findings demonstrate the utility of PROAST in modelling dose-

response relationships, particularly for *in vitro* endpoints, where robust benchmark dose (BMD) confidence intervals were generated with high reliability. By utilising the BMD (rather than AC50) as the dose-response metric, the modelling leverages a reduced reliance on experimental dose selection, considering the full dose-response curve and flexibility in PoD determination; this approach provided a more robust foundation for IVIVE.

However, additional compounds with increased and *in vivo* endpoints data were an essential requirement to allow for pipeline frameworks and datasets to be accurately analysed as unavailable data appeared to greatly limit the capacity of *in vivo* data to be compared to extrapolated *in vitro* data. The lack of data *in vivo* greatly produced challenges when missing BMD/AED findings for majority of endpoints produced had downstream effects on the final visualised ToxPi profiles/clusters and as the dataset collected was from an exercise with no further scope to acquire the missing data.

The potential variability in toxicokinetic parameters limited the success of BMD transformation into AEDs as a Css figure with high degree of variation is likely to produce compound potency results for diethylnitrosamine, cisplatin and hydroxyurea. This highlighted how Clint, Fub, and LogP are key parameters influencing the production of the Css and subsequent AED outputs. Efforts to generate the AED figures remained successful despite these shortcomings and provided valuable insight into compound potency, assisting in normalising dose responses across both *in vivo* and *in vitro* assays, before the implementation of ToxPi for visualisation purposes.

ToxPi successfully illustrated potency across both *in vitro* and *in vivo* datasets, offering a transparent, visual representation of complex data. Despite this, the necessary transformations applied during visualisation raised concerns about transparency and the potential for introducing bias. Additionally, missing data, particularly from *in vivo* endpoints, resulted in lower ToxPi scores and reduced the comparative power of these visualisations. Hierarchical clustering further underscored the impact of data limitations, with gaps in the dataset disrupting expected groupings and diminishing the clarity of compound-specific MoAs.

Potential steps to further develop IVIVE are currently ongoing by Health Canada and other institutions as its potential as an effective countermeasure to traditional genotoxicity assessments remains paramount. By integrating data from multiple sources and leveraging advanced computational approaches, IVIVE frameworks can enable more accurate and mechanistic evaluations of chemical safety. Improvements such as increased datasets and the

inclusion of factors such as S9 would be highly beneficial to include as variables such as metabolic activation could provide significant advancements and further comply to regulatory EPA OECD guidelines. However, this technology remains quite novel and ongoing/future studies are required to fully realize the potential of this approach, including the expansion of datasets, refinement of modelling algorithms, and improved alignment of *in vitro* and *in vivo* endpoints.

Continued development of IVIVE methodologies would offer a strong foundation for more ethical and effective toxicological assessments through the reduced reliance on animal testing and enhancing the reliability and specificity of predictive models. Research developing visualisation frameworks contributes to a paradigm shift in chemical potency safety evaluation through increased capacity of these *in vitro* models to accurately emulate human equivalents while visualised in a way through the reduction of extensive complex datasets into comparative outputs using dataset dimensionality reduction methods. This would overall balance scientific innovation with regulatory compliance and ethical expectations.

Chapter 6: Thesis Conclusions

6.1 – General Discussion

This thesis examines the quantitative measurement of genotoxic effects across a series of mutagenic compounds using automated *in vitro* multiplexed genotoxicity assays to address the underutilisation of information-dense datasets generated during safety assessments in the pharmaceutical, agricultural, cosmetics and foods industries. This investigation provided an approach to visualise assay findings into efficient figures for improved comparative analysis which accurately depicts compound MoA, metabolic properties, and genotoxic potency. Simultaneously, large complex datasets were condensed into suitable formats, predominantly for potency assessment purposes through the utilisation of modelling programs such as PROAST and visualisation tools such as ToxPi which brought statistical approaches such as agglomerative/divisive hierarchical and k-means (with PCA) clustering to group compounds based on similarities and in MoA, biomarker activation profiles and potency qualities. Datasets were generated and gathered from a variety of HESI GTTC members including Litron US, Procter and Gamble, Pfizer and Health Canada. However, the involvement of other institutions, including the University of Ottawa and Toxys BV further assisted in the development of the pipelines outlined in this investigation for each of their developed assays, including the TGx-DDI analysis and ToxTracker assay respectively.

This report encompasses the effectiveness of utilising the BMD model, as a foundational base, to effectively measuring and determine the PoD across both *in vitro* and *in vivo* studies and evaluate their ability to determine chemical potency and overall genotoxic potential. This project effectively demonstrated the robustness of applying quantitative transformations to experimental dose-response datasets across four distinct testing pipelines to generate BMDs, facilitating the determination of a PoD. This enabled hazard characterisation and a summarised safety assessment across the MultiFlow, MicroFlow, ToxTracker, TGx-DDI, iScreen and IVIVE assays, each employing different quantitative approaches to the raw experimental data based on the assay examined. This took considerable expertise and discussions to develop, with key members that helped develop each testing system, whilst retaining some degree of relatability across the different visualisation processes, particularly with the implementation of ToxPi 0.0001 – 10000 μM concentration boundaries to aid in the translation of BMD datasets into simplified digestible figures for comparative analysis.

6.2 – Thesis Main Findings

The main findings of each chapter in this thesis are subsequently covered in the following sub-sections with a summary of the conclusions for each component laid out here:

1. **Toxpi could be applied to individual assays, such as MultiFlow and MicroFlow, where BMD data generated by PROAST are used to produce compound potency and MoA findings for comparative assessment (Chapter 2).**

The wider scientific community has acknowledged the substantial evidence supporting the utilisation of the BMD approach as a modelling system to perform drug assessment using dose-response modelling to analyse and strongly describe compound behaviour more flexibly than alternatives such as the NOAEL model (Davis, et al., 2011, Rezvanfar, 2014). The wider scientific community has continued to prefer the BMD approach to characterise dose response and use programs such as PROAST produce BMD CIs across a variety of assay endpoints, including the MultiFlow (γH2AX, p53, and p-H3) and MicroFlow (MN and polyploidy) assays (Soeteman-Hernández et al., 2015; Davis, Gift & Zhao, 2011; RIVM, 2021). PROAST was the program used to produce these BMD CIs and ToxPi allowed for these BMD CIs to be effectively visualised into rankings, profiles and clusters based on a pipeline that aimed to effectively determine the properties of each set of compounds (Bryce et al., 2017;). Chapter 2 highlighted this across a series of three investigations each covering a set of: a) 10 clastogens b) eight topoisomerase ii inhibitors c) 30 aneugens/clastogens through figures 2.3.1.1 - 2.3.3.5. It was found that the MultiFlow/MicroFlow pipeline assisted in producing visualisation profiles of each compound, highlighting their predicted potencies and forming clusters that mostly correlated with their known MoAs. This system demonstrated that the MultiFlow and MicroFlow assays can have their potencies and MoAs assessed using BMD information to transform experimental data into legitimate and accurate set of *in vitro* visualisations that encapsulated compound properties that effectively depicted complex data into simplified figures through data dimensionality reduction methods frameworks employed. However, heavy dataset transformations are required for modelling in PROAST and transforming the BMD CI outputs into ToxPi scores as this is necessary for scaling purposes to prevent excessive skewing and this as the clustering remains imperfect, this would pose greater difficulty in analysing unknown compounds as further data on compounds with distinct MoAs are required to aid clustering and discovery of the MoA, properties and AOPs

of newer drugs, which studies such as (Dertinger et al., 2024) aims to address. Despite this, the approach used here provided an effective comparative system for multiplexed assays to displays differences in mechanisms between compounds with identical drug targets, compare a series of compounds and their respective metabolites efficiently and be utilised across a wider array of pathways and biomarkers within both MultiFlow and MicroFlow assays, to further analyse and represent datasets whilst incorporating the entire dataset into the final readings. This reduces information loss and assists in visualising these complex datasets, further providing greater confidence for *in vitro* assays to accurately perform potency assessment for drug development in the pharmaceutical industry.

2. **PROAST and ToxPi computational modelling/visualisation tools can integrate a series of two assays, ToxTracker and TGx-DDI, to enhance mechanistic insight, classify compounds and perform potency assessment on potential hazards while highlighting the need for further refinement in endpoint selection and study design to improve accuracy and clustering outcomes (Chapter 3).**

The 3Rs principle (replacement, reduction and refinement) is constructed and implemented to build frameworks that are less reliant on traditional *in vivo* genotoxicity assays by developing methodologies that provide highly sensitive and mechanistic insight when classifying compounds into their properties (Russel and Burch, 1959; Sewell et al., 2016; Dertinger et al., 2020). The combination of genotoxicity endpoints aligns with the 3Rs mission through the (i) reduction of testing systems required for ethical purposes and the reduction of the amount of tests required to capture chemical potency; (ii) refinement of assessment tools for the improved evaluation new potential biological, physical and environmental hazards and (iii) replacing animal testing where scientifically feasible to approach a more *in vitro* alternatives that accurately captures compound MoA and genotoxicity (Russel and Burch, 1959). Chapter 3 highlighted how ToxTracker and the TGx-DDI assay can be incorporated into a single analytical figure, which aims to reduce the reliance on *in vivo* follow up testing while providing extensive insight into a compound's MoA, potency, and mechanistic groupings, including the reactive properties between parent compounds and their metabolites. The ToxTracker assays attempted to use a series of biomarkers to distinguish between DNA damaging (Bscl2, Rtkn), oxidative stress (Srxn1, Blvrb), p53 cell stress response (Ddit3) and protein damage (Btg2) properties (Hendriks et al., 2011; 2016; 2024). Additionally, the complementary TGx-DDI assay provided information surround the ability for a compound to induce DNA damage through the analysis of 64-gene DDI signature probes into a singular

biomarker, highlighting whether a compound was DDI/non-DDI (Li et al., 2015; 2017; 2019; Buick et al., 2021). The combination of both ToxTracker and TGx-DDI provided an increased understanding surrounding compound properties by utilising a framework that incorporated the introduction of both PROAST and ToxPi to normalise dose-response metrics and visualise the findings respectively and to provide additional information on compound MoA profiles and clusters. Weighting was also further incorporated to ensure that both assays had a degree of applicability and distinction and to avoid diluting the singular TGx-DDI endpoint as it is an assay comprising of 64 genetic probes that was compressed into a singular endpoint, and hence a single ToxPi slice/projection. The results collected in figures 3.3.1.1 – 3.3.2.5 demonstrated that BMD CIs can be transformed for both ToxTracker and TGx-DDI into ToxPi figures which can be congruently ran into a singular analysis set, comprising of all six ToxTracker endpoints and the TGx-DDI as a singular biomarker with increased weighting applied to generate ToxPi potency ranks, profiles, and clusters highlighting which compounds were genotoxic, DDI and their respective AOPs. Despite some highly distinct clusters forming for tubulin binding aneugens and topoisomerase inhibitors, a lack of strong hierarchical/k-means clustering across other compounds implied that a series of compounds lacking strong distinct MoAs were used, leading to many compounds not successfully forming clusters based on MoAs. Regardless, this chapter successfully displayed that two sets of genotoxic assays can be used to distinguish between compounds of varying MoAs and through some improvements in the methodology implemented, will further the analytical capacity that scientists can hold as a highly complex dataset comprising of two unique but complementary assays can be combined into a singular figure to provide more insight into drugs/chemicals, enabling further research and development to be performed on their properties, providing labs with a great tool to maximise information extraction and analytically characterise chemicals into accurate groupings based on their mechanisms.

3. **The iScreen high-content imaging platform, integrated with PROAST and ToxPi, advances *in vitro* genotoxicity testing by enabling sensitive, mechanistic classification of aneugens, clastogens, and non-genotoxic compounds and could enhance its robustness and predictive capacity for early drug development though further refinement in endpoint specificity and biomarker weighting selection (Chapter 4).**

Battery approaches are implemented when performing *in vitro* MN testing, as a necessary step to retain the predictive and specificity/sensitivity capacity of the assay, despite its labour-

intensive manual scoring methods. However, it continued to struggle distinguishing between different MoAs and extract chromosomal mechanistic information, in a highly demanding setting, such as analysing larger compound lists in pharmaceuticals. The detection of these MN formation events relies on highly specific probes, antibodies and microbes to characterise between highly sensitive/specific MoAs and a testing system that could focus on gene mutations that lead to specific chromosomal modifications, such as iScreen, was designed to alleviate this concern by detecting these mutation manifestations events whilst allowing them to be rapidly scored through the implementation of flow cytometric automation (Bryce et al., 2010; 2013; OECD, 2016). The iScreen assay provided a high content imaging platform designed to analyse multiplexed datasets using a series of antibodies and imaging tools to provide a schematic overview of these chromosomal aberration events by incorporating highly specific endpoints that detects micronuclei presence (MN), chromosomal damage/fragmentation in the centromere region (CENPA), dsDNA breaks (γH2AX), mitotic index (↑MPM2 and p-H3) and polyploidy induction/p-H3 dephosphorylation (↓p-H3) events to distinguish between 10 compounds with clastogenic, aneugenic (tubulin binding/ aurora kinase inhibiting), and non-genotoxic mechanisms (Sun et al., 2022). Additionally, a weighting system was implemented where 3-fold, 2-fold and 1-fold weightings were applied to MN, CENPA and H2AX/MPM2/H3 endpoints respectively to reflect the relevance of the biological processes/endpoints detected. The final readings are displayed in figures 4.3.1.1 – 4.3.2.5, which reflect how PROAST/ToxPi complemented and accurately depicted iScreen imagine datasets into digestible potency figures proving both programs as excellent analytical tools that provide insight into multidimensional profiling and grouping based on MoA classifications across the 10 compounds to enhance dose-response analysis. Dose-response figures from the iScreen assay were effectively normalised using PROAST and translated into ToxPi rankings, profiles, and hierarchical/k-means clusters and the weighting applied across the iScreen endpoints allowed for the biological relevance to be accurately reflected in these findings. This chapter also highlighted the potential for further advancements in the iScreen assay by expanding the panel of compounds and biomarkers to encompass a broader range of MoA to improve predictive modelling through enhanced compound potency and the detection of unique genotoxic mechanisms, such as oxidative stress and specific kinase interactions. Additionally, the incorporation of assays like ToxTracker could provide additional mechanistic insights for compounds with atypical MoA and incorporating AI/ additional ML algorithm could reduce endpoint quantification issues, increase throughput accuracy/capacity and optimize data analysis regarding mechanistic

classification. The integration of tools like PROAST and ToxPi further underscores its capability for detailed mechanistic analysis, enabling sophisticated classification of aneugenic and clastogenic compounds by their MoA. By addressing the identified challenges and pursuing methodological innovations, the iScreen platform is poised to become a pioneering cornerstone in next generation genotoxicity testing methodologies, overall supporting safer and more effective drug discovery and development.

4. **IVIVE pipelines, using PROAST and HTTK, support the translation of *in vitro* data into human-relevant *in vivo* predictions through computational modelling algorithms to extrapolate cellular data to human-equivalent doses while visual tools like ToxPi, enable comparative potency and clustering analysis. Further refinement of datasets, algorithms, and endpoint alignment is needed to enhance accuracy, reduce reliance on animal testing, and advance chemical safety evaluation (Chapter 5).**

IVIVE is a reverse dosimetry extrapolatory methodology that adheres to the 3Rs principle to provide an alternative attempting to substitute animal testing method through the direct extrapolation of *in vitro* results into *in vivo* equivalent values to attempt and solve ethical issues surrounding drug testing methodologies (Russel and Burch, 1959; Sewell et al., 2016; Kavlock et al., 2018; Grimm, 2019; Beal et al., 2022). This direct animal testing alternative would encourage the maximisation of information extracted from genotoxicity datasets and whilst simultaneously alleviate the ethical concerns surrounding the animals in testing in industry, which would greatly be beneficial for the wider scientific community as international pressure continues to demand for the complete removal of animals from scientific research (Kavlock et al., 2018; Grimm, 2019). In this chapter, IVIVE employs a series of toxicodynamic and toxicokinetic ADME parameters to extrapolate *in vitro* rodent data into *in vivo* human results , by using *in vitro* blood/target tissue concentrations, cross-species extrapolation algorithms from HTTK (toxicokinetic/toxicodynamic values) and Monte Carlo simulations (to emulate human population variability) to extract blood Css values. This was combined with BMD analysis using PROAST to obtain AED figures, which were visualised using ToxPi to evaluate a set of extrapolated MutliFlow/MicroFlow/pig-a *in vivo* results Dertinger et al., (2019) and compare them to a set of *in vivo* rodent datasets across 6 known aneugenic/clastogenic compounds. This chapter explored the extrapolation of *in vitro* data to *in vivo* exposure estimates by integrating toxicokinetic/toxicodynamic variables with PROAST-based BMD modelling. This approach estimated the Css to calculate

the AED, predicting human exposure risk in mg/kg-bw/day through HTTK modelling (Davis, et al., 2011; Rezvanfar, 2014; EFSA 2017; Paul Friedman et al., 2020; Crump 2011). Subsequently, ToxPi was incorporated to visualise the findings to aid in the summarisation of the extrapolated *in vitro* findings and compare them against the *in vivo* findings. This section found that the application of IVIVE to has a degree of applicability to translate *in vitro* datasets into human-relevant *in vivo* predictions as Css values were successfully modelled. However, as the original evaluation of IVIVE Beal et al., (2022) in somewhat ideal conditions, found a 65% accuracy rate for IVIVE implementation. In this chapter, limitations in dataset consistency (high data transformations) and missing *in vivo* data reduced reliability in the methodologies applied here. High variability in toxicokinetic parameters (e.g., Clint, Fub, LogP) also impacted the accuracy of AED values, however, these figures still provided valuable insights into compound potency. Despite these limitations, the visualisation with ToxPi effectively summarised these complex datasets, despite facing challenges from missing endpoints and potential biases introduced by data transformations. This component also underscored the requirement for expansion to datasets, improved modelling algorithms, reduction in dataset transformations and better endpoint selection for improved alignment between *in vitro* and *in vivo* endpoints to enhance IVIVE's predictive power. These developments are crucial for refining IVIVE frameworks, steering genotoxic research toward reduced reliance on animal models, while advancing ethical, regulatory-compliant toxicological methodologies.

The table below summarises the key insights on the implemented approaches for both BMD analysis and ToxPi visualisation:

Table 6.2.1.1 – Summary table showing the assays used, endpoints assessed, chemicals tested and key findings across all chapters of the thesis.

Chapter	Assay(s) Used	Endpoints	Chemicals Tested	Key Findings
Chapter 2	MultiFlow® MicroFlow®	- p53, γH2aX, p-H3 - MN - Polyploidy	- 8 Topoisomerase II inhibitors: Mitoxantrone, Doxorubicin, Teniposide, Etoposide Genistein, Emodin, Ciprofloxacin, Flumequine. - 10 Clastogens: Mitomycin C, Dibenzol[a,l]pyrene, Resorcinol, 2-Aminoathracene, Benzo[a]pyrene, DMBA, PhIP, Cyclophosphamide Monohydrate, 2-Actylaminofluorene, Diethylnitrosamine - 30 Aneugens/Clastogens: Vinblastine Sulphate, Ixabepilone, Etoposide, Paclitaxel, Colchicine, AMG900, Melphalan, Temozolomide, VX680, PF03814735, Chlorambucil, Cisplatin, Genistein, MMS, Dibenzol[a,l]pyrene Methotrexate, 2-Aminoanthracene, Thiotepa, PhIP, Ethyl Methanesulfonate, Brefeldin A Hydroxyurea, Cyclophosphamide Monohydrate, Thapsigargin, Dexamethasone, Alosetron Hydrochloride Cycloheximide, D-Mannitol, Tolterodine Tartrate , 1, 2-Dimethylhydrazine	-The BMD approach successfully calculated CIs, providing a optimised framework for genotoxic PoD analysis. - While BMD CIs offer insights into genotoxic potency, they fail to integrate all endpoints, time points, and metabolic states into a unified metric, limiting comparative analysis. - ToxPi was used to overcome BMD limitations, enabling simultaneous comparison of multiple biomarkers and providing a more comprehensive evaluation of compound potency. - ToxPi and PROAST were used to analyse compounds with similar mechanisms of action (MoA). Doxorubicin and teniposide showed overlapping BMD CIs but differed in overall potency when all biomarkers were considered. - ToxPi provided a more effective comparative analysis of 10 clastogens, highlighting metabolic activation effects and clustering compounds based on potency and biomarker activation. - ToxPi provided a more effective comparative analysis of 10 clastogens, highlighting metabolic activation effects and clustering compounds based on potency and biomarker activation.
Chapter 3	ToxTracker TGx-DDI	- Blvrb, Bscl2, Btg2, Ddit3, Rtkn, Srxn1 - TGx-DDI (64-gene biomarker)	- 18 Testing Chemicals: 2-Deoxy-D-glucose, Amitrol, Cyclohexanone, D-Mannitol,	- BMD and ToxPi analysis effectively ranked, profiled, and clustered compounds based on genotoxic potency, especially in MultiFlow and MicroFlow assays.

			Sodium diclofenac, 5-Fluorouracil, Camptothecin, Colchicine, ENU, Etoposide, MMS, o-toluidine, Vinblastine, Ethionamide, Eugenol, Isobutyraldehyde, Quercetin, Saccharin sodium	
Chapter 4	iScreen	- MN, CENPA, γH2aX, p-H3, MPM2	- 10 Mixed MoA Compounds: AMG900, Baraserntib, Colchicine, Cisplatin, Dasatinib, ETO, MMC, Nocodazole, Taxol, Tunicamycin	- The iScreen high-content imaging platform demonstrated its ability to distinguish between clastogens, aneugens, and non-genotoxic compounds with high sensitivity and reproducibility. - PROAST successfully generated BMD CIs for multiple endpoints, but the method was limited in its ability to compare compounds and endpoints simultaneously. - ToxPi provided a more comprehensive analysis by weighting biomarkers (e.g., MN, CENPA, γH2AX) based on their biological relevance, allowing for better mechanistic classification of compounds. - Some compounds failed to produce expected dose-response curves due to assay limitations, non-specific toxicity, or lack of metabolic activation. Weighting of biomarkers in ToxPi lacked a standardised system, requiring further refinement. - Expanding the panel of compounds and biomarkers, integrating complementary assays (e.g., ToxTracker), and incorporating AI/ML algorithms could enhance the iScreen platform's predictive capabilities.
Chapter 5	IVIVE	- *In vitro*: p53, γH2aX, p-H3, MN, Polyploidy - *In vivo*: Ki-67, MN-RET, MNHEP, MUT-RET	- 6 Compounds: Cisplatin, Cyclophosphamide Monohydrate, Diethylnitrosamine, Hydroxyurea, Melphalan, Vinblastine Sulphate	- PROAST successfully modelled dose-response curves for most *in vitro* endpoints, but *in vivo* datasets had fewer successful BMD readings due to limited data. - Toxicokinetic (TK) parameters (Clint, Fub, LogP) were used to calculate steady-state concentrations (Css) and extrapolate *in vitro* results to *in vivo* AEDs. However, unstable TK parameters for some compounds (e.g., cisplatin, hydroxyurea) introduced variability.

				- Toxicokinetic (TK) parameters (Clint, Fub, LogP) were used to calculate steady-state concentrations (Css) and extrapolate *in vitro* results to *in vivo* AEDs. However, unstable TK parameters for some compounds (e.g., cisplatin, hydroxyurea) introduced variability. - The lack of *in vivo* data and the complexity of TK modelling limited the accuracy of IVIVE extrapolations. Compounds like vinblastine sulphate showed high potency *in vitro* but low potency *in vivo* due to missing endpoint data. - Clustering was affected by missing data, leading to unexpected groupings and reduced clarity in compound MoA classification. - Expanding datasets, refining modelling algorithms, and improving alignment between *in vitro* and *in vivo* endpoints are essential for advancing IVIVE methodologies and reducing reliance on animal testing.

6.3 – Genetic Toxicology Landscape

6.3.1 – The BMD approach

This research encompassed the utilisation of approaches such as the BMD as a dose-response relationship normalisation tool, formatted to develop new pipelines to analyse the relationship between a series of environmental, physical and chemical compounds that humans regularly encounter, allowing for genotoxic compounds to be identified through the assessment of concentrations and the impact/responses they contribute towards. PROAST was identified as a tool that can be used to map out the BMDs, as the exponential models 3 and 5 from various families of models, accurately attempted to model dose response and extract BMDL and BMDU figures using dose-response datasets (Davis et al., 2011). The BMD provides a strong predetermined change marker for the PoD typically ranging between 1 – 10% in response to adverse effects (but sometimes 50% like in this thesis) and the selection of exponential models used accurately allowed for the comprehensive mapping of dose-response according to human drug response behaviour (Macgregor et al., 2015; Davis, et al, 2011). The selection of the BMD approach as the backbone of this thesis was due to its advantages in reduced dependence on dose selection, considerations in the whole dose-response dataset to map the shape of the BMD curve, increased flexibility in selecting the CES and consistent applicability making it a reliable predictive method to mark the PoD through the production of BMDL and BMDU values (EFSA 2022). However, this left other models, such as the NOAEL, remain unexplored as the BMD is almost always preferred as a dose-response standard used by most researchers to derive a point of reference as a more scientifically advanced approach compared to the NOAEL (Dorato & Engelhardt, 2005; Davis, et al., 2011; Rezvanfar, 2014; Wills et al., 2016; EFSA 2022).

6.3.2 – ToxPi Visualisation

Additionally, ToxPi provides an excellent visualisation tool to exhibit compound BMD CIs in a highly advanced ranking, profiling and clustering system that reduces dataset dimensionality for chemical analysis (Marvel et al., 2018; Reif et al., 2012; Zhu et al., 2007). ToxPi provides visual mappings of compound properties, comparing potencies and grouping compounds based on MoA, however, to perform ToxPi analysis, heavy dataset transformations required depending on the testing implemented and response metrics found as a one-size-fits-all approach was unsuitable for this type of analysis. Despite the ability for ToxPi to effectively process multiplexed dose-response datasets, the ToxPi outputs are highly

dependent on the quality of the original dataset, meaning that if the original data is weak/lower in quality (less information/unexpected endpoint activations/etc), ToxPi would only present outputs for that provided dataset instead of interpolating results to deduce these ToxPi metrics. This means higher quality and information dense datasets to accurately depict chemical compounds as analysis effectiveness depends on dose-response data quality, quantity and testing/assay reliability. Additionally, modifications are a necessary requirement to produce ToxPi metrics as ToxPi outputs are prone to heavy skewing from certain response figures within datasets where response ranges from very high to very low figures, producing highly variable BMD concentrations, hence why 0.0001 – 10000 μM limits are imposed across all BMD datasets. Finally, ToxPi outputs/values operate relatively to each dataset implemented and therefore will effectively compounds within that dataset, but multitudes of dataset may introduce some challenges when attempting to compare them to each other as ToxPi values (i.e. ToxPi scores/90% endpoint projection CIs) operates on a relative basis. However, this does not mean multiple ToxPi outputs cannot be compared, moreover that comparisons drawn across multiple ToxPi findings may introduce some challenges due to the increasing complexity of comparing two sets of outputs that are relative to each set of BMD input values. ToxPi reduced highly complex data into simplified figures, however, dataset integrity varies as some assays, including MultiFlow/MicroFlow, usually retains most of the probity, however, assays such as IVIVE/HTTK may suffer from difficulty as a highly complex and novel methodology with instability across some of its key modelling parameters. However, ToxPi makes for an excellent analysis tool for the comparison of compounds to accurately supplement BMD data and directly compare a series of BMD datapoints into a metrics that considers all inserted parameters, weighting factors and dataset transformations for chemical analysis and drug development research (Marvel et al., 2018; Reif et al., 2012; Zhu et al., 2007).

6.3.3 – Analysis Framework Development

This thesis also encapsulates a multitude of genotoxicity MoA, including mutagenicity, clastogenicity and aneugenicity, covering a wide array of mechanisms and reactive pathways. The assessment of these compound classes did not encapsulate all formats of genotoxicity as the methodologies would assess some but not all unique MoA (EFSA., 2021; Menz et al., 2023; Doah et al., 2023). However, the advancement of new pipelines capable of assessing MoA linked to diverse AOPs remains a key focus in genetic toxicology, driving the creation of tests that identify biological/genetic events leading to adverse human outcomes (Sasaki et

al., 2019; Svingen et al., 2021). However, the design of these pipelines may produce many challenges that researchers must face, including considerations surrounding the *in vitro* assays selected/developed/used to replicate *in vivo* models (abolishing animal testing), visualising complex computational data, degrees of dataset dimensionality reduction, etc (Menz et al., 2023; Dertinger et al., 2024). However, as reality is not always reflected perfectly in research, this produces greater difficult for researchers when developing new frameworks as modelling cannot perfectly replicate/contribute towards consideration of inter-human variation factors, even when using simulation algorithms such as Monte Carlo, which aid in considering this, however, is not perfect. This also fails to address the difficulty in the outright replacement of existing models to detect AOPs (which may be undetected/highly sensitive). These systems also pose risks of incorrectly addressing/highlighting obscure MoAs exhibited by compounds, meaning mechanisms may not be clearly distinguished/classified. Additionally, with the incorporating artificial intelligence into genotoxic research, this could push for the combination of multiple assays into a singular analytical platform that gives in-depth highly detailed, yet directly comparative analyses that provide a full set of findings that may increase our understanding of novel/existing compounds, despite the additional complexity that such a system could add. Overall, with continued strides by the global scientific community, these matters would continue to evolve in a manner where these pipelines can eventually substitute existing methods to provide a more comprehensive, holistic and realistic depiction of the interactive relationship between humans and existing/novel drug and how these impacts physical, biological and environmental factors.

6.4 – Final Conclusion

This thesis demonstrates the transformative potential of integrating advanced assays such as MultiFlow, MicroFlow, ToxTracker, TGx-DDI, and IVIVE with computational tools like PROAST and ToxPi to enhance genotoxicity testing. These methodologies collectively enable robust compound classification, dose-response analysis, and mechanistic insights, significantly improving the depth and precision of toxicity assessments. By combining high-throughput imaging, multiplexed biomarker analysis, and *in vitro*-to-*in vivo* extrapolation, these approaches address key challenges in predictive toxicology, reduce reliance on animal testing, and advance the understanding of adverse outcome pathways (AOPs) for genotoxic compounds. Despite challenges such as refining endpoint selection and standardising data

integration, this work lays a strong foundation for safer compound evaluation and innovation in early-phase drug discovery.

6.5 – Future Research and Development

Current genotoxicity research is developing ideas surrounding both the genetic toxicology and data science landscape, using the BMD approach as a dose-response base for future research. Each protocol required thorough evaluation and encompassed strong statistical, biological and genetic considerations to comparatively analyse chemicals, while strictly adhering to regulatory OECD frameworks and guidelines. The approaches adopted in this thesis highlighted both the benefits and drawbacks of a variety of pipelines implemented for each component, however, it would be greatly beneficial if the findings of this thesis were improved by considering the following aspects that require further investigation:

1. No strong hierarchical or k-means clusters were observed when analysing the combined ToxTracker and TGx-DDI assays using PROAST and ToxPi. Adding a series of positive/negative control compounds strongly associated with specific MoAs could improve clustering by grouping similar ToxPi profiles and MoAs. Including the 24-hour MN (±S9) endpoint may further distinguish genotoxicants from non-genotoxicants based on the presence or absence of micronuclei, indicating chromosomal damage or mis segregation events.
2. Incorporating additional complementary assays into the iScreen assay could provide deeper insights into a compound's MoA by expanding the biomarkers assessed. Refining the weighting system would further enhance the accuracy of endpoint significance for genotoxicity, improving compound classification and potency assessment.
3. Using biochemically related complementary endpoints between *in vitro* and *in vivo* systems enables more direct comparisons between profiles. Adding more *in vivo* data points, compounds, and endpoints addresses data gaps, improving potency analysis and clustering. Developing an effective pipeline reduces data transformations, incorporates metabolic activation, and considers new concentration limits for *in vivo* data, distinct from the *in vitro* range of 0.0001–10,000 μM.

Appendices

List of Appendices

Appendix 1. Published Literature (Dertinger et al., 2024)

Appendix 2. Published Literature (Hussien, Dertinger & Johnson, 2025)

Appendix 3. Genetic Toxicology Association (GTA) 2024 Annual Meeting Abstract Award

Appendix 4. Genetic Toxicology Association (GTA) 2024 Annual Meeting Presentation

Appendix 5. Health and Environmental Science Institute Genetic Toxicology Technical Committee (HESI GTTC) 2024 Annual Meeting Presentation

Appendix 6. Beal 2022 IVIVE presentation

Appendix 7. Genetic Toxicology Association (GTA) 2025 Annual Meeting Abstract

Appendix 8. BEMS 2024 Networking Lunch Invitation

Appendix 9. Litron Laboratories Collaboration Invitation/Certification of Attendance

Appendix 1. Published Literature (Dertinger et al., 2024)

Check for updates

Received: 12 March 2024 | Revised: 22 April 2024 | Accepted: 23 April 2024
DOI: 10.1002/em.22604

Environmental and Molecular Mutagenesis | Environmental Mutagenesis and Genomics Society | WILEY

RESEARCH ARTICLE

Visualization strategies to aid interpretation of high-dimensional genotoxicity data

Stephen D. Dertinger[1] | **Erica Briggs**[1] | **Yusuf Hussien**[2] | **Steven M. Bryce**[1] | **Svetlana L. Avlasevich**[1] | **Adam Conrad**[1] | **George E. Johnson**[2] | **Andrew Williams**[3] | **Jeffrey C. Bemis**[1]

[1]Litron Laboratories, Rochester, New York, USA
[2]Institute of Life Sciences, Swansea University, Swansea, UK
[3]Environmental Health Science and Research Bureau, Health Canada, Ottawa, Canada

Correspondence
Stephen D. Dertinger, Litron Laboratories, 3500 Winton Place, Rochester, NY, USA.
Email: sdertinger@litronlabs.com

Funding information
National Institute of Environmental Health Sciences, Grant/Award Number: R44 ES033138

Accepted by: P. White

Abstract

This article describes a range of high-dimensional data visualization strategies that we have explored for their ability to complement machine learning algorithm predictions derived from MultiFlow® assay results. For this exercise, we focused on seven biomarker responses resulting from the exposure of TK6 cells to each of 126 diverse chemicals over a range of concentrations. Obviously, challenges associated with visualizing seven biomarker responses were further complicated whenever there was a desire to represent the entire 126 chemical data set as opposed to results from a single chemical. Scatter plots, spider plots, parallel coordinate plots, hierarchical clustering, principal component analysis, toxicological prioritization index, multidimensional scaling, t-distributed stochastic neighbor embedding, and uniform manifold approximation and projection are each considered in turn. Our report provides a comparative analysis of these techniques. In an era where multiplexed assays and machine learning algorithms are becoming the norm, stakeholders should find some of these visualization strategies useful for efficiently and effectively interpreting their high-dimensional data.

KEYWORDS
dimensionality reduction, hierarchical clustering, multidimensional scaling, parallel coordinate plots, principal component analysis, spider plots, ToxPi, t-distributed stochastic neighbor embedding, uniform manifold approximation

1 | INTRODUCTION

The field of toxicology has embraced multiplexed biomarker assays and the integration of numerous data streams in order to more holistically study deleterious consequences of chemical exposures. Interested readers are directed to recent reviews by Krewski et al. (2022) and Pognan et al. (2023). With the discipline of genetic toxicology working toward providing more detailed information, for example about test substances' predominant genotoxic mode of action or mechanism, high-information content genotoxicity assays have proliferated (Bryce et al., 2016, 2017, 2018; Chapman et al., 2021; Corton et al., 2018; Hall et al., 2022; Hendriks et al., 2012; Salk & Kennedy, 2020; Sun et al., 2022; Wilson et al., 2021).

Challenges associated with the interpretation of increasingly complex data sets have naturally led to the use of machine learning (ML) models, at times even model ensembles (Bryce et al., 2016; Buick

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.
© 2024 His Majesty the King in Right of Canada and The Authors. *Environmental and Molecular Mutagenesis* published by Wiley Periodicals LLC on behalf of Environmental Mutagenesis and Genomics Society. Reproduced with the permission of the Minister of Health Canada.

et al., 2020, 2021, 2022; Dertinger et al., 2019; Fortin et al., 2023; Thienpont et al., 2023; Wilson et al., 2021). The generation of predictive models, especially those that prove highly applicable to previously untested chemicals, certainly represents an efficient way to process large volumes of data. Yet to varying degrees, some stakeholders view ML-based predictions with suspicion. One common refrain is that the underlying algorithms are challenging to understand, and this resemblance to a "black box" does not advance their trust. For this and other reasons, we have been actively exploring data visualization techniques that complement ML model(s) by helping explain and contextualize underlying data structure.

As noted by Tufte (2001) "…of all methods for analyzing and communicating statistical information, well-designed data graphics are usually the simplest and at the same time the most powerful." When done well, graphics enhance interpretability, and thereby the depth and breadth of our understanding of toxicological response profiles. While the benefits of informative visualizations are considerable, non-trivial obstacles can exist, for instance, datasets that exhibit high dimensionality.

This report describes our experiences with a variety of data visualization strategies that were evaluated for their ability to aid in the interpretation of a modestly high-dimensional dataset. Specifically, we describe our work with seven biomarker/time point combinations resulting from a flow cytometry-based assay known as MultiFlow® (Bryce et al., 2016, 2017, 2018). Each of nine tools were considered in turn: scatter plots, spider plots, parallel coordinate plots, hierarchical clustering, principal component analysis (PCA), toxicological prioritization index (ToxPi), multidimensional scaling, t-distributed stochastic neighbor embedding (t-SNE), and uniform manifold approximation and projection (UMAP). Beyond portraying results for each of seven biomarker/time point combinations, we increased the complexity of the visualizations by simultaneously evaluating response profiles for 126 diverse chemicals (scatter and spider plots excepted). By examining and contrasting these nine tools, we hope to provide researchers and other stakeholders with a better understanding of the strengths and limitations of each approach.

While strengths and limitations are discussed, it will become apparent from the analyses that follow that there is no "best" or "worst" visualization tool. Rather, each has merit, depending on the intention. We anticipate that colleagues who may be unfamiliar with some of these approaches will gain a better understanding of how and when to apply these visualization tools to their datasets of interest.

2 | MATERIALS AND METHODS

2.1 | Chemicals, cells, culture conditions

The identity and source of 126 test chemicals can be found in Table 1. Of these 126 chemicals, 123 were previously tested, and citations to the original reports are provided. Regardless of the source of the data, we have created two Supplemental Microsoft Excel files (no. 1 and 2) that includes all the data used for the analyzes that follow, with biomarker results provided both as fold-change values on an individual concentration basis, as well as area-under-the-curve (AUC) values. As described in more detail below, these AUC values do not include overly-cytotoxic concentrations. Table 1 and Supplemental Files 1 and 2 also include our *a priori* classifications regarding predominant in vitro mammalian cell cytogenetic damage activity: clastogen, non-genotoxicant, or aneugen. In the case of aneugens, Table 1 also includes information about presumptive mechanism, i.e., tubulin binder (TB), or aurora kinase inhibitor phenotype (AKIP).

All experiments were performed with TK6 cells from ATCC (cat. no. CRL-8015). Details about growth medium and other general culture conditions can be found within the original reports (Avlasevich et al., 2021; Bryce et al., 2016, 2017, 2018; Dertinger et al., 2019; Wheeldon et al., 2020).

2.2 | Cell treatments

Details about exposure of TK6 cells to test chemicals can be found with the original reports (Avlasevich et al., 2021; Bryce et al., 2016, 2017, 2018; Dertinger et al., 2019; Wheeldon et al., 2020). Briefly, treatments occurred in 96 well plates, most often using DMSO as the solvent, delivered from 100x stock solutions (exceptions to DMSO are indicated in Table 1) Unless limited by solubility or precipitate, the highest concentration tested was 1 mM, with up to 19 additional/lower concentrations. In most cases, a square root 2 dilution scheme was used—that is, each concentration differed from the next higher by a factor of 70.71%. For the vast majority of these experiments, each concentration was tested in a single well, whereas solvent was evaluated in at least four replicate wells. Upon addition of test chemical, the plates were incubated at 37°C for 4 and 24 h.

2.3 | MultiFlow DNA damage assay: cell processing and flow cytometric analysis

Details about executing MultiFlow assays can be found with the original reports (Avlasevich et al., 2021; Bryce et al., 2016, 2017, 2018; Dertinger et al., 2019; Wheeldon et al., 2020). Briefly, cells were prepared for analysis using reagents and instructions included in the MultiFlow® DNA Damage Kit—p53, γH2AX, Phospho-Histone H3 (Litron Laboratories, Rochester, NY). At 4 and 24 h sampling times, cells were combined with the MultiFlow reagent solution in wells of 96 well plates. Flow cytometric analysis was carried out using either a FACSCanto™ II flow cytometer equipped with a BD™ High Throughput Sampler or a Miltenyi Biotec MACSQuant® Analyzer 10 flow cytometer with integrated 96-well MiniSampler device.

2.4 | MultiFlow DNA damage assay: Data pre-processing

Data analyses and visualizations described herein were restricted to those concentrations that did not exceed the MultiFlow assay's

10982280, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/em.22604 by Test, Wiley Online Library on [17/05/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 1 Chemicals and *a priori* classifications.

Chemical, Abbreviation	CAS No., Source if not Sigma-Aldrich	*a priori* Mammalian Cell Genotoxicity & Predominant MoA Classification	MultiFlow Publication	Misc. Notes; References
Alisertib	1028486-01-2; Selleckchem	Genotoxic; Aneugen—AKIP	Bernacki et al., 2019	Aurora kinase A > B inhibition; Sehdev et al., 2012
AMG-900	945595-80-2	Genotoxic; Aneugen—AKIP	Avlasevich et al., 2021	Pan-Aurora kinase inhibitor (A/B/C); Payton et al., 2010
Barasertib	722544-51-6; Selleckchem	Genotoxic; Aneugen—AKIP	Bernacki et al., 2019	Highly specific for Aurora kinase B; Yang et al., 2007
Carbendazim	10605-21-7	Genotoxic; Aneugen—TB	Bryce et al., 2016	Mitotic spindle poison; Van Hummelen et al., 1995
Colchicine	64-86-8	Genotoxic; Aneugen—TB	Bryce et al., 2016	Mitotic spindle poison; Kirkland et al., 2016
Crizotinib	877399-52-5	Genotoxic; Aneugen—AKIP	Dertinger et al., 2019	Tyrosine kinase inhibitor, potent activity against c-Met and ALK, with evidence of off-target Aurora kinase inhibition [Zou et al., 2007; Kong et al., 2018]
Danusertib	827318-97-8; Selleckchem	Genotoxic; Aneugen—AKIP	Bernacki et al., 2019	Potent activity against Aurora kinases with cross-reactivity against some cancer-relevant tyrosine kinases; Carpinelli et al., 2007
Diethylstilbestrol	56-53-1	Genotoxic; Aneugen—TB	Bryce et al., 2016	Synthetic estrogen; Parry et al., 2002
Epothilone A	152044-53-6; Selleckchem	Genotoxic; Aneugen—TB	Bernacki et al., 2019	Bollag et al., 1995
17beta-Estradiol	50-28-2	Genotoxic; Aneugen	Bryce et al., 2016	Steroid hormone; Hernández et al., 2013
Flubendazole	31430-15-6	Genotoxic; Aneugen—TB	Bryce et al., 2016	Mitotic spindle poison; Tweats et al., 2016
Griseofulvin	126-07-8	Genotoxic; Aneugen—TB	Bryce et al., 2016	Mitotic spindle poison; Oliver et al., 2006
Hesperadin	422513-13-1; Selleckchem	Genotoxic; Aneugen—AKIP	Dertinger et al., 2019	Aurora kinase inhibitor (B); in vitro MN pos., aberrant metaphases; Hauf et al., 2003; Kurihara et al., 2006
Ixabepilone	219989-84-1	Genotoxic; Aneugen—TB	Bernacki et al., 2019	Lee et al., 2001
Mebendazole	31431-39-7	Genotoxic; Aneugen—TB	Bryce et al., 2016	Mitotic spindle poison; Van Hummelen et al., 1995
Nocodazole	31430-18-9	Genotoxic; Aneugen—TB	Bryce et al., 2016	Mitotic spindle poison; Verdoodt et al., 1999
Noscapine	128-62-1	Genotoxic; Aneugen—TB	Dertinger et al., 2019	Mitotic spindle poison; Schuler et al., 1999
Paclitaxel	33069-62-4	Genotoxic; Aneugen—TB	Bryce et al., 2016	Mitotic spindle poison; Kirkland et al., 2016
PF-03814735	942487-16-3	Genotoxic; Aneugen—AKIP	Bernacki et al., 2019	Jani et al., 2010
Rigosertib sodium	1225497-78-8; Selleckchem	Genotoxic; Aneugen—TB	Bernacki et al., 2019	Tubulin destabilization attributed to an impurity found in commercial products, not the clinical-grade drug; Baker et al., 2020
Vinblastine sulfate	143-67-9	Genotoxic; Aneugen—TB	Bryce et al., 2016	Mitotic spindle poison; Kirkland et al., 2016
Vincristine sulfate	2068-78-2	Genotoxic; Aneugen—TB	Bryce et al., 2016	Mitotic spindle poison; Kondo et al., 1992

(Continues)

10982280, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/em.22604 by Test, Wiley Online Library on [17/05/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 1 (Continued)

Chemical, Abbreviation	CAS No., Source if not Sigma-Aldrich	*a priori* Mammalian Cell Genotoxicity & Predominant MoA Classification	MultiFlow Publication	Misc. Notes; References
VX-680 (Tozaserib)	639089-54-6; Selleckchem	Genotoxic; Aneugen—AKIP	Dertinger et al., 2019	Pan-Aurora kinase inhibitor (A/B/C); Gollapudi et al., 2014
ZM-447439	331771-20-1; Selleckchem	Genotoxic; Aneugen—AKIP	Dertinger et al., 2019	Aurora kinase inhibitor (A/B); Gollapudi et al., 2014
Aphidicolin	38966-21-1	Genotoxic; Clastogen	Bryce et al., 2016	DNA polymerase inhibitor; Glover et al., 1984
Azathioprine	446-86-6	Genotoxic; Clastogen	Bryce et al., 2016	Prodrug of mercaptopurine, purine analog; Henderson et al., 1993
AZD2858	486424-20-8; Selleckchem	Genotoxic; Clastogen	Bryce et al., 2018	Glycogen synthase-3 inhibitor; in vitro MN pos., in vitro CA pos., Ann Doherty, personal communication
Bleomycin sulfate	9041-93-4	Genotoxic; Clastogen	Bryce et al., 2016	Radiomimetic; Rosefort et al., 2004
Cadmium chloride	10108-64-2	Genotoxic; Clastogen	Bryce et al., 2018	in vitro MN, CA, and Hprt pos., in vivo CA and MN pos.; Kirkland et al., 2016
Camptothecin	7689-03-4	Genotoxic; Clastogen	Bryce et al., 2016	Topoisomerase I inhibitor; Attia et al., 2009
Carmofur	61422-45-5	Genotoxic; Clastogen	Unpublished	Prodrug of 5-fluorouracil; Islam & Mirza, 2022
Chlorambucil	305-03-3	Genotoxic; Clastogen	Bryce et al., 2016	Nitrogen mustard-type alkylator; Dertinger et al., 2012
Cisplatin	15663-27-1	Genotoxic; Clastogen	Bryce et al., 2016	Atypical alkylator; Kirkland et al., 2016
Cytosine arabinoside	147-94-4	Genotoxic; Clastogen	Bryce et al., 2016	Anti-metabolite; Kirkland et al., 2016
Dasatinib	302962-49-8; Selleckchem	Genotoxic; Clastogen	Bryce et al., 2018	Tyrosine kinase inhibitor, especially Ber-Abl, Scr, c-Kit; Ames neg., clastogenic in CHO, in vivo MN neg.; Sprycel® (Dasatinib) package insert, 2010
N,N'-Diphenyl-p-phenylenediamine	74-31-7	Genotoxic; Clastogen	Bryce et al., 2018	Ames pos., in vitro SCE pos., in vitro CA neg.; CEBS database
Doxorubicin HCl	25316-40-9	Genotoxic; Clastogen	Bryce et al., 2018	Anthracycline, likely several modes of action that includes inhibition of topoisomerase II; Gewirtz, 1999
Emodin	518-82-1	Genotoxic; Clastogen	Dertinger et al., 2019	Anthraquinone, topoisomerase II inhibitor; Li et al., 2010
Ethyl methanesulfonate	62-50-0	Genotoxic; Clastogen	Bryce et al., 2016	Alkylator; Gocke et al., 2009
N-Ethyl-N-nitrosourea	759-73-9	Genotoxic; Clastogen	Bryce et al., 2016	Alkylator; Kirkland et al., 2016
Etoposide	33419-42-0	Genotoxic; Clastogen	Bryce et al., 2016	Topoisomerase II inhibitor; Kirkland et al., 2016
Flumequine	42835-25-6	Genotoxic; Clastogen	Wheeldon et al., 2020	Topoisomerase II inhibitor; Kashida et al., 2002
5-Fluorouracil	51-21-8	Genotoxic; Clastogen	Bryce et al., 2016	Anti-metabolite, thymidylate synthase inhibitor; Kirkland et al., 2016
Genistein	446-72-0	Genotoxic; Clastogen	Bryce et al., 2018	Topoisomerase II inhibitor; in vitro MN pos.; Klein & King, 2007

10982280, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/em.22604 by Test, Wiley Online Library on [17/05/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 1 (Continued)

Chemical, Abbreviation	CAS No., Source if not Sigma-Aldrich	*a priori* Mammalian Cell Genotoxicity & Predominant MoA Classification	MultiFlow Publication	Misc. Notes; References
Glutaraldehyde	111-30-8	Genotoxic; Clastogen	Bryce et al., 2018	Ames pos. and weakly pos. and equivocal, in vitro CA weakly pos. and negative, in vitro SCE pos. and weakly pos., MLA pos., in vivo MN neg. and equivocal, in vivo CA pos.; CEBS database
Glycidamide	5694-00-8	Genotoxic; Clastogen	Dertinger et al., 2019	Major in vivo metabolite of acrylamide; Paulsson et al., 2003
Hydralazine HCl	304-20-1	Genotoxic; Clastogen	Bryce et al., 2016	Prepared in RPMI medium; Martelli et al., 1995
Hydrogen peroxide	7722-84-1	Genotoxic; Clastogen	Bryce et al., 2016	ROS, prepared in RPMI medium; Kimura et al., 2013
Hydroquinone	123-31-9	Genotoxic; Clastogen	Bryce et al., 2018	Clastogenic, but likely more than one mode of action; Kirkland et al., 2016
8-Hydroxyquinoline	148-24-3	Genotoxic; Clastogen	Bryce et al., 2018	Ames pos., in vitro CA weakly pos., SCE pos., MLA pos., in vivo MN neg.; CEBS database
Hydroxyurea (Hydroxycarbamide)	127-07-1	Genotoxic; Clastogen	Bryce et al., 2016	Anti-metabolite, ribonucleotide reductase inhibitor; Dertinger et al., 2012
Irinotecan	97682-44-5; Selleckchem	Genotoxic; Clastogen	Bryce et al., 2018	Topoisomerase I inhibitor; Ames neg., in vitro ChromAb pos., in vivo MN pos.; Camptosar® FDA approved label, 2014
beta-Lapachone	4707-32-8; Selleckchem	Genotoxic; Clastogen	Bryce et al., 2018	Topoisomerase I inhibitor; in vitro ChromAb and comet pos.; Degrassi et al., 1993
Melphalan	148-82-3	Genotoxic; Clastogen	Bryce et al., 2016	Nitrogen mustard-type alkylator; Dertinger et al., 2012
Menadione	58-27-5	Genotoxic; Clastogen	Bryce et al., 2016	ROS implicated; Cojocel et al., 2006
Methotrexate	59-05-2	Genotoxic; Clastogen	Bryce et al., 2016	Anti-metabolite; Keshava et al., 1998
Methyl methanesulfonate	66-27-3	Genotoxic; Clastogen	Bryce et al., 2016	Alkylator; Kirkland et al., 2016
N-Methyl-N′-nitro-N-nitrosoguanidine	70-25-7	Genotoxic; Clastogen	Dertinger et al., 2019	Alkylator; Nikolova et al., 2014
Mercuric chloride	7487-94-7	Genotoxic; Clastogen	Bryce et al., 2018	Ames neg., in vitro CA pos., MLA pos.; SCE weakly pos., CEBS database
Mitomycin C	50-07-7	Genotoxic; Clastogen	Bryce et al., 2016	DNA cross-linker; Kirkland et al., 2016
Mitoxantrone 2HCl	70476-82-3; Selleckchem	Genotoxic; Clastogen	Dertinger et al., 2019	Topoisomerase II inhibitor; in vitro MN pos., γH2AX pos.; Smart et al., 2008
4-Nitroquinoline 1-oxide	56-57-5	Genotoxic; Clastogen	Bryce et al., 2016	Likely several modes of clastogenic action that may include ROS; Kirkland et al., 2016
Olaparib	763113-22-0	Genotoxic; Clastogen	Bryce et al., 2016	PARP inhibitor; Lynparza™, 2014
1,3-Propane sultone	1120-71-4	Genotoxic; Clastogen	Bryce et al., 2016	Alkylator; Dertinger et al., 2011
Propyl gallate	121-79-9	Genotoxic; Clastogen	Dertinger et al., 2019	ROS likely; Tayama & Nakagawa, 2001

(Continues)

10982280, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/em.22604 by Test, Wiley Online Library on [17/05/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 1 (Continued)

Chemical, Abbreviation	CAS No., Source if not Sigma-Aldrich	*a priori* Mammalian Cell Genotoxicity & Predominant MoA Classification	MultiFlow Publication	Misc. Notes; References
Pyrimethamine	58-14-0	Genotoxic; Clastogen	Bryce et al., 2018	Ames neg., in vitro SCE weakly pos., in vitro CA pos.; CEBS database
Resorcinol diglycidyl ether	101-90-6	Genotoxic; Clastogen	Dertinger et al., 2019	in vitro SCE pos., in vitro CA pos.; Gulati et al., 1989
Resveratrol	501-36-0; ChromaDex, Inc. via NTP	Genotoxic; Clastogen	Bryce et al., 2018	Topo 2 inhibitor; Basso et al. (2013)
SN-38	86639-52-3	Genotoxic; Clastogen	Unpublished	Topoisomerase I inhibitor; Voigt et al., 1998
Stavudine	3056-17-5	Genotoxic; Clastogen	Dertinger et al., 2019	Nucleoside analog; Zerit® FDA approved label, 2002
Temozolomide	85622-93-1	Genotoxic; Clastogen	Bryce et al., 2016	Alkylator; Chinnasamy et al., 1997
Teniposide	29767-20-2; Selleckchem	Genotoxic; Clastogen	Dertinger et al., 2019	Topoisomerase II inhibitor; in vitro CA pos., in vitro MLA pos.; DeMarini et al., 1987
6-Thioguanine	154-42-7	Genotoxic; Clastogen	Bryce et al., 2018	Antimetabolite, purine analog; Ames pos., in vitro CA pos., in vivo MN pos.; CEBS database
Thiotepa	52-24-4	Genotoxic; Clastogen	Bryce et al., 2016	Alkylator; Dertinger et al., 2012
Topotecan	123948-87-8	Genotoxic; Clastogen	Bryce et al., 2016	Topoisomerase I inhibitor; Aydemir & Bilaloğlu, 2003
Zinc dimethyldithiocarbamate (Ziram)	137-30-4	Genotoxic; Clastogen	Bryce et al., 2018	Ames pos., in vitro MLA pos., in vitro CA pos.; in vitro SCE neg., in vivo CA neg., CEBS database
Alosetron HCl	122852-69-1	Non-genotoxic	Bryce et al., 2016	5-HT_3 antagonist; Kirkland et al., 2016
3-Amino-1,2,4-triazole (Amitrole)	61-82-5	Non-genotoxic	Bryce et al., 2016	Kirkland et al., 2016
Ampicillin trihydrate	7177-48-2	Non-genotoxic	Bryce et al., 2018	Ames neg., in vitro ChromAb neg., in vivo MN neg.; Kirkland et al., 2016
Anisomycin	22862-76-6	Non-genotoxic	Bryce et al., 2018	Protein biosynthesis inhibitor; in vitro MN neg. with high levels of apoptosis; Sun et al., 2022
Anthranilic acid	118-92-3	Non-genotoxic	Bryce et al., 2016	Kirkland et al., 2016
Brefeldin A	20350-15-6	Non-genotoxic	Bryce et al., 2016	ER-golgi transporter inhibitor, ER stress-induced apoptosis; Moon et al., 2012
Butyl chloride (1-Chlorobutane)	109-69-3	Non-genotoxic	Bryce et al., 2016	Fumigant; Kirkland et al., 2016
Caffeine	58-08-2	Non-genotoxic	Bryce et al., 2016	Mitochondria-dependent apoptosis; Lu et al., 2008
Carbonyl cyanide 3-chlorophenylhydrazone	555-60-2	Non-genotoxic	Bryce et al., 2016	Uncoupler of oxidative phosphorylation; de Graaf et al., 2004
Chlorocholine chloride	999-81-5	Non-genotoxic	Bryce et al., 2018	Ames neg., in vitro CA neg., in vivo CA neg.; Kirkland et al., 2016
Clofibrate	637-07-0	Non-genotoxic	Bryce et al., 2016	Antilipidemic agent; IARC monograph, Clofibrate, 1996
Cyclohexanone	108-94-1	Non-genotoxic	Bryce et al., 2016	Industrial chemical; Kirkland et al., 2008
Cycloheximide	66-81-9	Non-genotoxic	Bryce et al., 2016	Protein synthesis inhibitor; Youngblom et al., 1989

10982280, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/em.22604 by Test, Wiley Online Library on [17/05/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 1 (Continued)

Chemical, Abbreviation	CAS No., Source if not Sigma-Aldrich	*a priori* Mammalian Cell Genotoxicity & Predominant MoA Classification	MultiFlow Publication	Misc. Notes; References
Daidzein	486-66-8	Non-genotoxic	Bryce et al., 2018	Ames neg.; CEBS database
Dexamethasone	50-02-2	Non-genotoxic	Bryce et al., 2016	Glucocorticoid receptor agonist; Krishna et al., 1995
Dextrose	50-99-7	Non-genotoxic	Bryce et al., 2016	Sugar; Lotz et al., 2009
1,3-Dicyclohexyl-2-thiourea	1212-29-9	Non-genotoxic	Unpublished	Ames neg., in vitro MLA uninterpretable, in vitro CA neg.; Kirkland et al., 2016
Diethanolamine	111-42-2	Non-genotoxic	Bryce et al., 2016	Secondary amine; Kirkland et al., 2016
Di-(2-ethylhexyl)phthalate	117-81-7	Non-genotoxic	Bryce et al., 2016	Organic plasticizer; Kirkland et al., 2016
Erythromycin	114-07-8	Non-genotoxic	Bryce et al., 2016	Antibiotic; Kirkland et al., 2016
Famotidine	76824-35-6	Non-genotoxic	Bryce et al., 2016	Histamine H_2 receptor antagonist; Pepcid® FDA approved label, 2011
Gefitinib	184475-35-2	Non-genotoxic	Bryce et al., 2018	in vitro and in vivo genetic toxicity neg.; EGFR tyrosine kinase inhibitor; Iressa® FDA approved label, 2003
Hexachloroethane	67-72-1	Non-genotoxic	Bryce et al., 2016	Industrial chemical; Ames neg., in vitro MN neg., in vitro CA neg., in vivo MN neg.; Kirkland et al., 2016
Imatinib mesylate	152459-95-5	Non-genotoxic	Bryce et al., 2016	Protein-tyrosine kinase inhibitor; Gleevec® FDA approved label, 2001
Lidocaine	137-58-6	Non-genotoxic	Bryce et al., 2016	Amide, local anesthetic; Lidoderm® FDA approved label, 2004
D-Limonene	5989-27-5	Non-genotoxic	Bryce et al., 2016	Male rat kidney tumors due to α2μ-globulin nephropathy; Kirkland et al., 2016
Lovastatin	75330-75-5	Non-genotoxic	Bryce et al., 2016	HMG-CoA reductase inhibitor; Mevacor® FDA approved label, 2012
D-Mannitol	69-65-8	Non-genotoxic	Bryce et al., 2016	Polyol; Kirkland et al., 2016
Melamine	108-78-1	Non-genotoxic	Bryce et al., 2016	Industrial organic base; Kirkland et al., 2016
Menthol	89-78-1	Non-genotoxic	Bryce et al., 2018	Ames neg., in vitro MN neg. in p53 competent cell lines, in vivo MN and comet neg.; Kirkland et al., 2016
Methyl carbamate	598-55-0	Non-genotoxic	Bryce et al., 2016	Industrial intermediate; Kirkland et al., 2016
Ofloxacin	82419-36-1	Non-genotoxic	Bryce et al., 2016	Fluoroquinoline antibiotic; Floxin® FDA approved label, 2008
Osimertinib	1421373-65-0	Non-genotoxic	Bryce et al., 2018	EGFR kinase inhibitor; in vitro and in vivo genetox neg.; Tagrisso® FDA approved label, 2012
Paroxetine	61869-08-7	Non-genotoxic	Bryce et al., 2016	SSRI antidepressant; Paxil® FDA approved label, 2011
Phenanthrene	85-01-8	Non-genotoxic	Bryce et al., 2016	Polycyclic aromatic hydrocarbon; Kirkland et al., 2008
Phenformin HCl	834-28-6	Non-genotoxic	Bryce et al., 2016	Biguanide antidiabetic; Kirkland et al., 2016
Progesterone	57-83-0	Non-genotoxic	Bryce et al., 2016	Steroid hormone; Kirkland et al., 2008

(Continues)

10982280, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/em.22604 by Test, Wiley Online Library on [17/05/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 1 (Continued)

Chemical, Abbreviation	CAS No., Source if not Sigma-Aldrich	*a priori* Mammalian Cell Genotoxicity & Predominant MoA Classification	MultiFlow Publication	Misc. Notes; References
Pyridine	110-86-1	Non-genotoxic	Bryce et al., 2016	Heterocyclic organic compound; Kirkland et al., 2016
Rosuvastatin calcium	147098-20-2	Non-genotoxic	Bryce et al., 2018	Ames neg., in vitro MLA and CA neg.; Kirkland et al., 2016
Sodium chloride	7647-14-5	Non-genotoxic	Bryce et al., 2016	Prepared in RPMI medium; Matsushima et al., 1999
Sodium dodecyl sulfate	151-21-3	Non-genotoxic	Bryce et al., 2016	Ionic detergent; Ames neg., in vitro CA neg., in vitro MLA neg.; CEBS database
Sucrose	57-50-1	Non-genotoxic	Bryce et al., 2016	Diaz et al., 2007
Tert-butyl alcohol	75-65-0	Non-genotoxic	Bryce et al., 2016	Kirkland et al., 2016
Thapsigargin	67526-95-8	Non-genotoxic	Bryce et al., 2016	ER stress-induced apoptosis; Futami et al., 2005
Tolterodine L-tartrate	124937-52-6	Non-genotoxic	Bryce et al., 2016	Muscarinic receptor antagonist; Kirkland et al., 2016
Topiramate	97240-79-4	Non-genotoxic	Bryce et al., 2018	Ames, in vitro CA and MLA neg., in vivo CA Neg.; Kirkland et al., 2016
Tris (2-ethylhexyl) phosphate	78-42-2	Non-genotoxic	Bryce et al., 2018	Ames neg., in vitro CA neg., in vivo CA and MN neg.; Kirkland et al., 2016
Tunicamycin	11089-65-9	Non-genotoxic	Bryce et al., 2016	Glycosylation inhibitor, ER stress-mediated apoptosis; Han et al., 2008
Zafirlukast	107753-78-6	Non-genotoxic	Bryce et al., 2018	Ames, in vitro CA, in vitro MLA neg., in vitro Hprt neg.; Kirkland et al., 2016
Zonisamide	68291-97-4	Non-genotoxic	Dertinger et al., 2019	Sulfonamide anticonvulsant; Kirkland et al., 2016

Abbreviations: AKIP, aurora kinase inhibitor phenotype; CA, chromosome aberration; CEBS, chemical effects in biological systems; Hprt, hypoxanthine phosphorybosyl transferase; MLA, mouse lymphoma assay; MN, micronucleus; SCE, sister chromatid exchange; TB, tubulin binder.

10982280, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/em.22364 by Test Wiley Online Library on [17/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

established cytotoxicity limit, that is, the top concentration of each chemical had to exhibit ≤80% reduction to relative nuclei count at the 24 h time point. Additionally, only two concentrations within the cytotoxicity range 70%–80% were acceptable (Dertinger et al., 2019).

Details about measuring biomarker responses at 4 and 24 h can be found with the original reports (Avlasevich et al., 2021; Bryce et al., 2016, 2017, 2018; Dertinger et al., 2019; Wheeldon et al., 2020). Briefly, γH2AX and p53 results were derived from median fluorescence intensity measurements, whereas p-H3 and polyploidy results were based on their frequency among all nuclei. The various biomarker data were normalized by taking individual well measurements and deriving fold-change values relative to the plate-specific solvent control mean value. Note that given the large fold-change polyploidy values induced by some aneugens at the 24 h time point, polyploidy fold-change values were square root transformed.

For the largest experiments, one that considered 20 concentrations, 4 solvent control wells, and 7 biomarker/time point combinations per chemical, the resulting data set comprised 168 fold-change values. Thus, for the majority of the visualization strategies described below, we utilized a previously described data reduction strategy (Dertinger et al., 2019). Briefly, one AUC value was calculated for each biomarker/time point combination, thereby reducing each chemical's response profile to 7 data points. As explained in greater detail in the original report, the AUC calculations employed unity-based normalization, that is, each test article's concentration range was scaled to 0–1. Also, a baseline adjustment was made by subtracting 1 from every fold-change value before AUC calculations were made in order to set the no effect value to zero.

2.5 | Scatter plots

Scatter plots were created in JMP® Pro for Mac (v17.2.0). These graphs plot biomarker responses against test article concentration for each of three exemplar genotoxicants. Each test article generated a series of seven plots, one for each biomarker/time point combination. The X axes show μM test article concentrations that have been log10 transformed. The Y axes corresponds to biomarker fold-change values relative to concurrent mean solvent control. Average values for each concentration are represented by individual points, and a smooth line was fit to these points using the settings: Method = spline; Lambda = default.

2.6 | Spider plots

Spider plots were created in Excel for Mac (v16.77.1). These graphs plot biomarker responses against test article concentration for each of three exemplar genotoxicants. Each spider plot portrays the responses of 7 biomarker/time point combinations as fold-change values. For these graphs, each test article concentration is illustrated by a distinct continuous line that encircles the plot. Each spoke of the spider plot corresponds to a specific biomarker, and the degree to which the lines project away from the origin signifies the magnitude of the response. As noted by the figure legends, each colored line represents a different test article concentration.

2.7 | Parallel coordinate plots

Parallel coordinates plots are used to represent multidimensional data on a two-dimensional plane. Parallel plots were originally described by Inselberg (1985), and later by Wegman (1990). Parallel plots draw connected line segments that represent each row in a data table. We generated parallel plots in JMP for Mac (v17.2.0) to visualize AUC values for seven biomarker responses and each of 126 test articles. The test articles were grouped according to *a priori* predominant modes of action as described in Table 1. In this manner, parallel plots were used to provide a concise representation of the biomarker response profiles for each of the predefined categorical groups.

2.8 | Hierarchical clustering

Hierarchical clustering is a method used to group similar data points into nested clusters based on their pairwise similarities (Everitt, 1974; Hartigan, 1975). The process begins by treating each data point as a separate cluster and iteratively merging clusters (using a linkage method) to form a hierarchical tree-like structure, known as a dendrogram. The resulting dendrogram visually represents the relationships and hierarchy among clusters, allowing for the identification of both fine-grained and coarse-grained structures within the data. For the current exercise, JMP Pro for Mac (v17.2.0) hierarchical clustering platform was used to visualize AUC values of seven biomarker responses for each of 126 test articles. The dendrograms presented herein position AUC data horizontally, chemical names vertically. Software clustering settings were as follows: Clustering Method = hierarchical; Method for Calculating Distances Between Clusters = complete (linkage method); Data as Usual = standardize data; Data Visualization = dendrogram, with two-way clustering. The two-way clustering option was chosen in order to supplement the dendrogram with a color-coded heat map. Note that we color-coded the three *a priori* MoA categories as follows: clastogens = red; non-genotoxicants = green; aneugens = blue.

2.9 | PCA

Principal Component Analysis (PCA), originally described by Pearson at the turn of the 20th century, is a linear dimensionality reduction method employed to transform high-dimensional data into a lower-dimensional representation (Pearson, 1901). The primary objective is to capture the most significant variations in the data by generating new variables, known as "principal components" (PCs), through weighted combinations of the original features. PCA utilizes orthogonal axes, representing the directions of maximum variance in the

10982280, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/em.22604 by Test, Wiley Online Library on [17/05/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

original data. The first PC captures the most variance, followed by subsequent components arranged in descending order.

PCA was executed with JMP Pro for Mac (v17.2.0), and was based on 126 test articles that each provided 7 AUC values, one per biomarker. Default JMP PCA settings were used: Standardized = standardized; Method Family = default; Eigenvalues = shown. The PCs were graphed in 2D space, both PC1 vs. PC2 and PC1 vs. PC3. Thus, each point on a plot represents an individual chemical's PCs scores, and the distance between points indicates the degree of similarity, in this case, between biomarker responses. Note that we color-coded the three *a priori* MoA categories as follows: clastogens = red; non-genotoxicants = green; aneugens = blue.

2.10 | ToxPi analyses

The Toxicological Prioritization Index (ToxPi™) has been described in detail by Marvel et al. (2018). ToxPi analyses presented herein were conducted with a JavaScript available at www.ToxPi.org. For this exercise, we formatted each ToxPi profile to represent a unique chemical, where each profile consists of 7 slices—one per biomarker. The projection length of each slice represents a particular biomarker's AUC value, thereby conveying the responsiveness of the biomarker to the chemical treatment. (Note that as described above, all AUC values were baseline adjusted, −1. Since ToxPi does not allow for negative values, +1 was added back.) Besides presenting ToxPi profiles for each of three exemplar genotoxicants, we also utilized a ToxPi hierarchical clustering algorithm to generate a circular dendrogram that includes all 126 test chemicals. This was accomplished with the following settings: Clustering Method = ward.D2; Layout Options = circle; Background Color = white; Number of Clusters = 1. Also note that each of the seven ToxPi slices used the default arc width setting, that is, equal width, meaning no adjustments were made to biomarker weightings.

2.11 | Multidimensional scaling

Multidimensional scaling (MDS) is a widely used technique that transforms high-dimensional data into a lower-dimensional representation (Borg & Groenen, 2005). Through MDS algorithms, each chemical is positioned within this lower-dimensional representation, with the objective to preserve inter-chemical distances to the greatest extent possible. When considering classical scaling and Euclidean distances, MDS and PCA are equivalent. MDS's popularity is driven in part by high-dimensionality biomedical applications such analysis of single-cell RNA sequencing data (Chen et al., 2019). For our exercise, MDS was conducted with a data set comprised of 126 chemicals, with 7 biomarker response values (AUC) each. Each resulting point in the MDS plot represents a unique chemical, and the proximity of points indicates the similarity of their biomarker response profiles. We performed these analyses with JMP Pro for Mac (v17.2.0), with default MDS settings: Data Format = auto; Transformation = none; Set Dimensions = 2. The MDS plot is accompanied by a Shepard diagram. The Shepard diagram plots actual proximities versus the predicted proximities. The plot indicates how well the MDS plot reflects the actual proximities. Ideally, the points fall on the Y = X line, which is shown in red. Note that for these analyses we color-coded the three *a priori* MoA categories as follows: clastogens = red; non-genotoxicants = green; aneugens = blue.

2.12 | t-SNE

t-SNE is a non-linear dimensionality reduction method used to transform high-dimensional data into a two-dimensional map. It focuses on local pairwise similarities between data points, making it easier to optimize compared to its predecessor, Stochastic Neighbor Embedding (van der Maaten & Hinton, 2008). t-SNE generates superior visualizations by reducing the clustering of points in the center of the map. It excels in revealing structures at various scales, making it especially valuable for high-dimensional data with multiple related low-dimensional representations. Although computationally intensive, t-SNE is widely used for visualizing complex datasets, making it valuable in applications where preserving local relationships is important. For this exercise, t-SNE was performed with 126 chemicals that each contributed 7 biomarker response (AUC) values. Each resulting point in the t-SNE plot represents a unique chemical, and the proximity of points indicates the similarity of their biomarker response profiles. The following JMP Pro for Mac (v17.2.0) t-SNE settings were used: Output Dimensions = 2; Random Seed = 123 (Standardize; Missing Value Imputation); Perplexity = 30; Maximum Iterations = 1000, Initial Scale = 0.0001; Convergence Criterion = 0.0001; Eta = 200; Inflate Iterations = 250; Initial Principal Component Dimensions = 50 (Sparse). Note that for these analyses we color-coded the three *a priori* MoA categories as follows: clastogens = red; non-genotoxicants = green; aneugens = blue.

2.13 | Uniform manifold approximation and projection

Uniform Manifold Approximation and Projection (UMAP) is a non-linear dimensionality reduction technique that efficiently captures both local and global structures in high-dimensional data (McInnes & Healy, 2018). UMAP aims to create a low-dimensional representation where similar points are modeled as nearby and others are pushed apart. UMAP's ability to preserve both local and global relationships contributes to its popularity in tasks such as clustering, visualization, and exploration of complex data sets while maintaining computational efficiency. For this exercise, UMAP was conducted with 126 chemicals that each contributed 7 AUC values, one for each biomarker. Each resulting point in the UMAP plot represents a unique chemical, and the proximity of points indicates the similarity of their biomarker response profiles. We performed these analyses with JMP Pro for Mac (v17.2.0), using the following settings: Output Dimensions = 2;

10982280, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/em.22604 by Test, Wiley Online Library on [17/05/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Random Seed = 123 (Standardize; Missing Value Imputation); Number of Neighbors = 15; Number of Epochs = 500; Learning Rate = 1; Minimum Distance = 0.01; Local Connectivity = 1; a, b = 0, 0; Negative Sample Rate = 5; Batch Mode if N Greater Than = 4096; Nearest Neighbor Method = Default; Distance Matrix = Euclidean; Gradient Descent Method = SGD. Note that for these analyses we color-coded the three *a priori* MoA categories as follows: clastogens = red; non-genotoxicants = green; aneugens = blue.

3 | RESULTS AND DISCUSSION

3.1 | Scatter plots

Scatter plots for mitomycin C, vinblastine, and VX-680 (i.e., clastogen, TB-type aneugen, and AKIP-type aneugen, respectively) are provided in Figure 1. This widely encountered graph format excels at displaying biomarker response data against test article concentration. Given

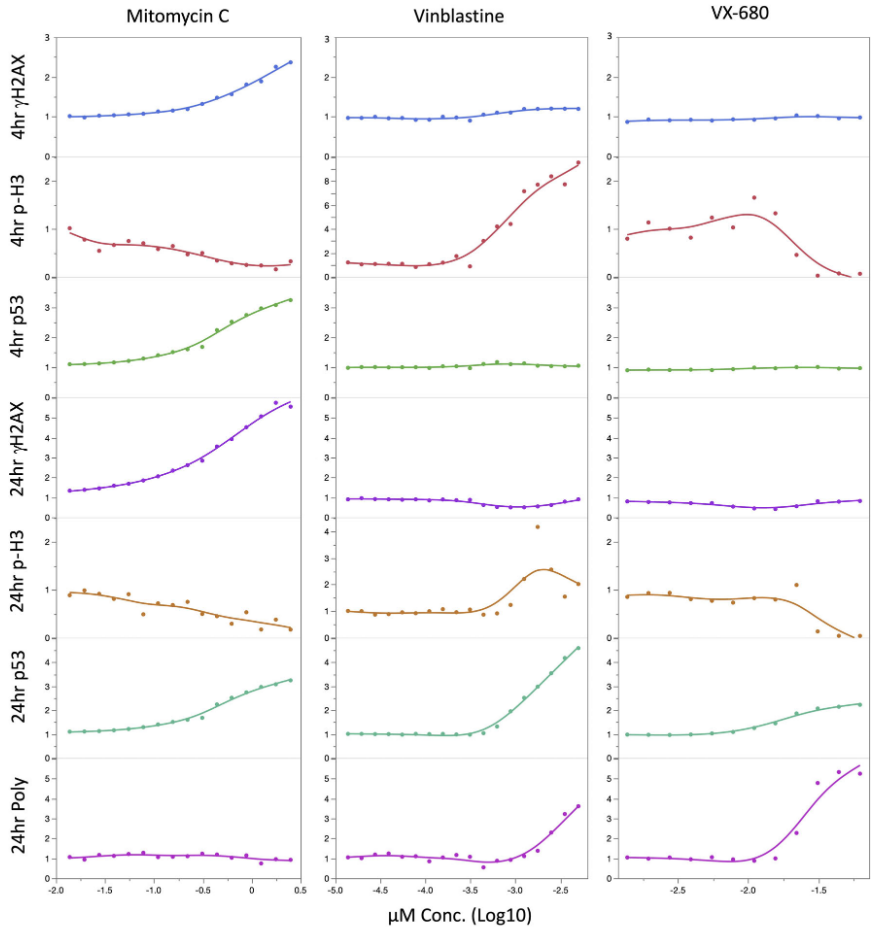


FIGURE 1 Scatter plots of TK6 cell-based MultiFlow biomarker data for each of 7 biomarker/time point combinations. The Y-axis shows biomarker responses as fold-change values versus solvent control, and the X-axis represents test article concentration (μM). The left panels correspond to mitomycin C treatment, and show a prototypical clastogenic response pattern whereby γH2AX and p53 are elevated at both time points, phopho-histone-H3 (p-H3) positive cells are reduced, and there is little to no polyploidization (Poly). The middle panels correspond to vinblastine treatment, and show a prototypical tubulin binder-type aneugenic response pattern whereby γH2AX shows little to no effect, p-H3 is elevated at both time points, and p53 and polyploidy are elevated at the later time point only. The right panels correspond to VX-680 treatment, and show a prototypical aneugenic response pattern associated with the aurora kinase inhibitor phenotype whereby γH2AX shows little to no effect, p-H3 is strongly reduced, and p53 and polyploidy are elevated at the later time point only. Concentrations have been log10 transformed, and Poly fold-change values have been square root 2 transformed.

10982280, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/em.22604 by Test, Wiley Online Library on [17/05/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

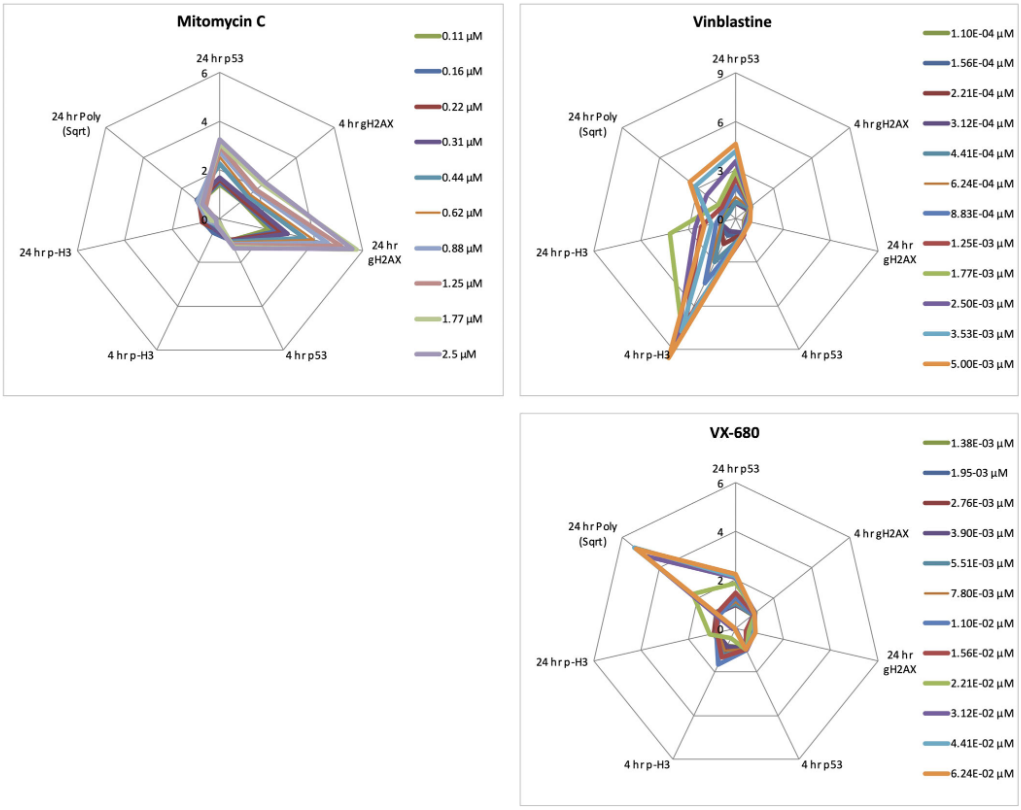


FIGURE 2 Spider plots show seven biomarker response profiles for three genotoxicants, the clastogen mitomycin C, the tubulin binder-type aneugen vinblastine, and aurora kinase inhibitor phenotype-type aneugen VX-680. The biomarker response data are fold-increase values, and as specified at the right of these plots, each chemical concentration appears as a different colored line. Regarding the position of individual biomarkers around the perimeter of the plot: 24 h p53, a pan-genotoxic biomarker, is positioned at 12 o'clock, whereas the biomarkers arranged on the right side of the graph are responsive to clastogens and those arranged on the left are responsive to aneugens.

people's familiarity with scatter plots, the data points and fitted lines are readily interpretable to a wide range of audiences. Since test article concentration appears on these X-axes, at least to a certain extent, the graphs convey information regarding genotoxic potency. Also, the side-by-side presentation used here helps highlight some important biomarker response differences observed for these chemicals. That being said, this depiction does not, in and of itself, convey the degree to which these biomarker data represent prototypical clastogen, TB-type aneugen, and AKIP-type responses, or whether they are exceptional in some respect(s). Thus, scatter plots as utilized here are quite effective at providing a granular view of one or even several particular chemicals' dose response profiles, but they do not represent an economical approach for contextualizing results across greater chemical space. The inefficiencies of this graphing choice are obvious when one considers the volume of space needed to show 7 biomarker responses for a mere three chemicals.

3.2 | Spider plots

Figure 2 shows MultiFlow fold-change data associated for the same three genotoxicants that were presented in Figure 1, this time as spider plots. These depictions are more efficient than scatter plots, in the sense that a collection of 21 plots have been condensed to 3. By plotting several colored lines that each correspond to a different test article concentration, dose response relationships are conveyed. That being said, the dose-response relationships are less intuitively presented relative to scatter plots. Arguably the greatest strength of

10982280, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/em.22604 by Test, Wiley Online Library on [17/05/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

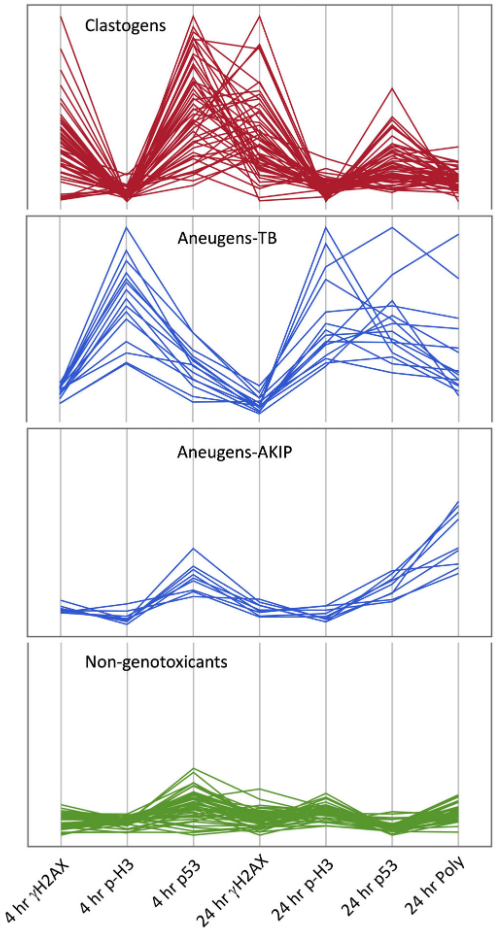


FIGURE 3 Parallel coordinates plots display seven biomarker responses for 126 chemicals. Each line represents a unique chemical's response profile in the form of an AUC value. Notice that the chemicals were subdivided into four panels that correspond to presumed mode of action categories: clastogens (red), tubulin (TB) type aneugens (blue), aurora kinase inhibitor phenotype (AKIP) aneugens (blue), and non-genotoxicants (green).

these spider plots is the ease by which comparative analyses can be performed across graphs that have been positioned side-by-side. Indeed, we reinforced this feature through the careful attention given to the placement of biomarkers around the graphs' perimeters. More specifically, the pan-genotoxicant biomarker 24 h p53 is positioned at 12 o'clock, clastogen-responsive biomarkers appear on the right-hand side, and aneugen-responsive biomarkers are positioned on the left. In this way, the mitomycin C spider plot is observed to be right leaning, whereas the aneugens' spider plots lean left.

Thus, for this use case, we find that spider plots are more efficient and effective at conveying MultiFlow's MoA information relative to analogous scatter plots. However, even given a 7-fold reduction to the number of graphs, whenever there is a desire to efficiently compare and contrast very large numbers of chemical response profiles, spider plots will usually not be the most practical visualization approach.

3.3 | Parallel coordinate plots

Parallel coordinate plots (Figure 3) display seven biomarker responses for all the chemicals listed in Table 1. In this depiction, each line represents a unique chemical's response profile, with the four panels grouping chemicals into *a priori* genotoxic MoA categories: clastogens, TB-type aneugens, AKIP-type aneugens, and non-genotoxicants. The economy by which the information is conveyed stems in part from the data reduction step whereby each of the seven biomarker responses were converted to an AUC value. This transformation simplified the dose response relationships, with a consequential loss of information. For instance, since these are concentration-normalized AUC values, the profiles really only convey information about biomarker responsiveness. A desire for insights regarding genotoxic potency would require different analyses that have not eliminated chemical concentration information. Perhaps the chief positive attribute of normalized AUC values in conjunction with parallel plots is that they represent a highly efficient approach for conveying overall biomarker response patterns across different groups. In this particular case, the plots effectively highlight biomarker signature differences for four modes of action, and the number of chemicals in each group provide insights as to the degree to which generalities can be made. Parallel plots therefore represent another useful aid to those faced with ML model predictions and/or other data analysis outputs that may be challenging to understand without visual cues.

3.4 | Hierarchical clustering

A two-way hierarchical clustering map (with heatmap) is presented in Figure 4. As with parallel plots, this representation considers seven biomarker responses for all 126 chemicals. The economy by which this many results are provided also benefitted from the data reduction step whereby individual biomarker dose responses were converted into AUC values. With this simple transformation, colors and their intensities intuitively convey information about biomarker responsiveness. Of course, details around the shape of dose response relationships are lost, as well as information regarding genotoxic potency. On the other hand, important benefits are realized, perhaps most importantly, clusters corresponding to genotoxic MoA become evident. This is apparent from the dendrogram, the hierarchical tree-like structure shown to the right side of Figure 4. Among the interesting findings is that in addition to distinguishing between clastogenic and aneugenic chemicals, the dendrogram also makes it clear that two subclasses of aneugens are discernable—TB versus AKPI. Thus, as with parallel

10982280, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/em.22604 by Test, Wiley Online Library on [17/05/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

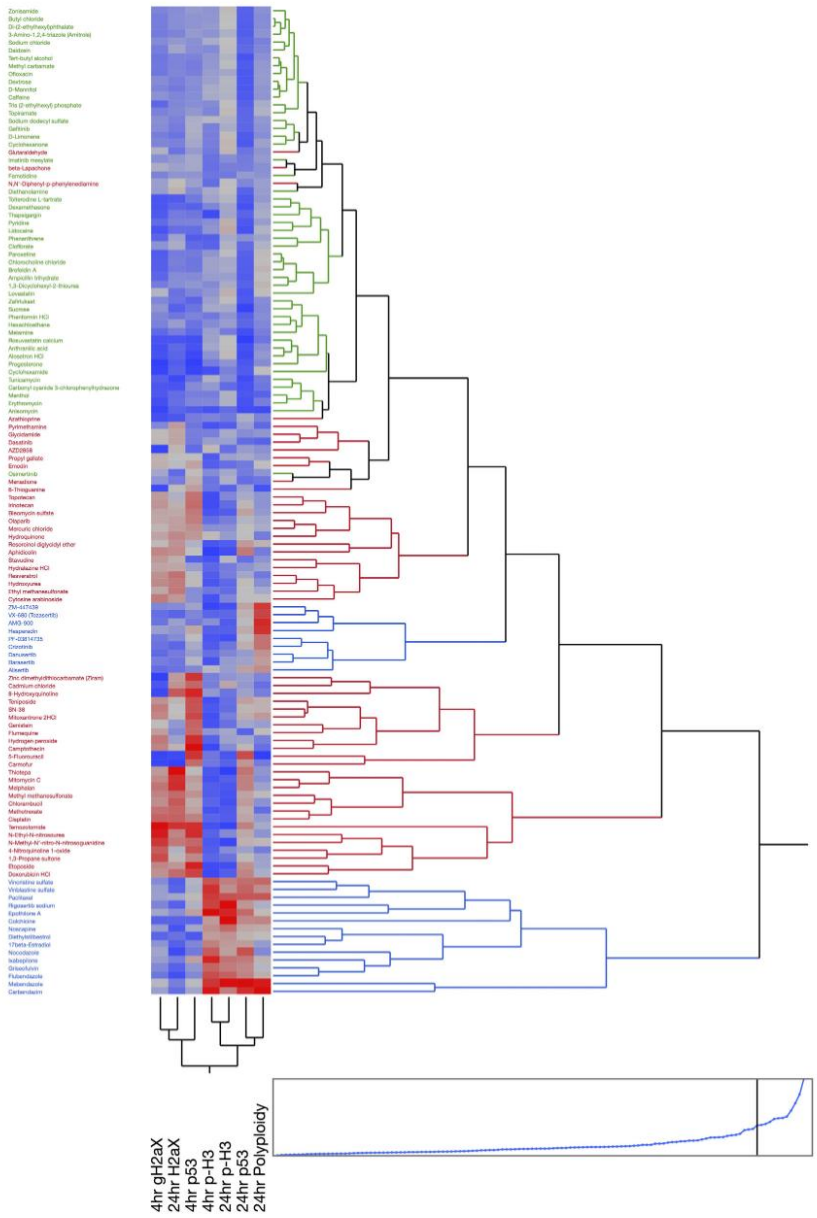


FIGURE 4 Hierarchical clustering results are shown for 7 biomarker responses and 126 chemicals. Note that these analyses are based on biomarker AUC values. Whereas chemicals are arranged along the vertical axis, the horizontal dendrogram illustrates the hierarchical clustering of chemicals based on similarities in their biomarker response profiles. Warm colors indicate higher AUC values, while cooler colors indicate lower AUC values. The bottom-most graph shows the horizontal distance between joint points. Note that the chemicals were color-coded according to their *a priori* mode of action category: clastogens (red), aneugens (blue), and non-genotoxicants (green).

10982280, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/em.22604 by Test, Wiley Online Library on [17/05/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

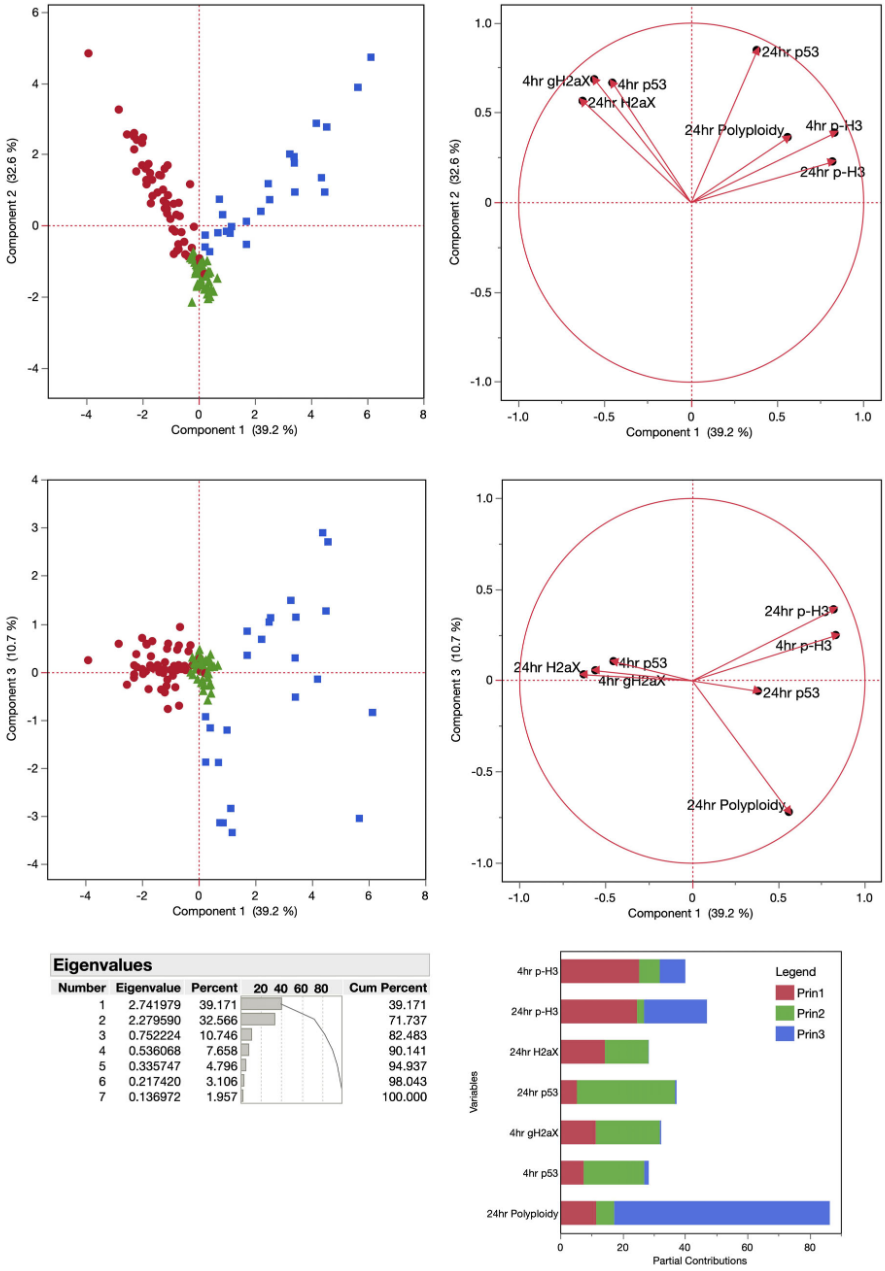


FIGURE 5 Principal components analysis plots are shown for 7 biomarker responses and 126 chemicals. The second principal component is graphed against the first principal component, and the third principal component is graphed against the first principal component. The percentage of the variance captured by each component are displayed in brackets on the x and y axis. The scatter plots are also accompanied by the biplots to visualize the coefficients or loadings of the eigenvectors for each of the biomarkers. The eigenvalues for the seven eigenvectors are displayed with the associated percentage of variance at the bottom of the figure. Note that the chemicals were color-coded according to their *a priori* mode of action category: clastogens (red), aneugens (blue), and non-genotoxicants (green).

10982280, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/em.22604 by Test, Wiley Online Library on [17/05/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

plots, hierarchical clustering represents a useful tool for contextualizing high-dimensionality genotoxicity data. That is, for any one chemical, they efficiently convey how multiple biomarkers responded, and furthermore how closely the overall signature resembles other tested chemicals (reference or otherwise).

3.5 | PCA

Figure 5 shows results from PCA and includes two PCs plots: PC2 versus PC1, and PC3 versus PC1. By definition, these are the two most important plots, since the first PC captures the most variance, followed by subsequent components arranged in descending order. Eigenvalues associated with each PC are provided at the bottom left of Figure 5. Eigenvalues indicate the amount of variability explained by each PC, where higher values correspond to more important components. Variance (%) is also provided as a small horizontal bar chart to the right of the eigenvalues, and conveys how much of the total variability in the dataset is explained by each PC. Cumulative variation is also shown.

The position of the points on these PCs plots reflects their degree of similarity in AUC-based biomarker responses. Recall that for this analysis, chemicals were color-coded, corresponding to an *a priori* predominant MoA. Thus, it is encouraging that the PC2 versus PC1 plot shows three major groupings—non-genotoxicants near the origin, clastogens stretching toward the 10 o'clock position, and aneugens extending toward 2 o'clock. Interestingly, although the two aneugen subtypes are not clearly differentiated in this first plot, they are when PC3 is graphed versus PC1. In this case, the 9 AKPIs are observed to form a cluster that extends toward 6 o'clock. The reason for this repositioning/differentiation provides an opportunity to discuss some important aspects of PCA, below.

In addition to PCs plots, it can be extremely useful to study the information provided by PCA's loading plots. In Figure 5, two loading plots are shown to the right of corresponding PC plots. Loading plots can be interpreted as follows: (i) vectors represent the original variables in the dataset, (ii) the direction of an arrow indicates the direction in which the corresponding variable changes the most, (iii) the length of the arrow represents the importance of the variable in that direction, and (iv) angles between vectors reflect correlations between variables, where small angles indicate a high positive correlation, and large angles indicate a low or negative correlation.

Thus, from the loading plots, we can see that three correlated biomarkers are responsive to clastogenicity (4 and 24 h γH2AX, and 4 h p53), three correlated biomarkers are responsive to aneugenicity (4 and 24 h p-H3, and 24 h polyploidy), and one is responsive to both modes of action (24 h p53). Also, regarding the differentiation of aneugen subtypes, it is not until a PC that is heavily weighted on polyploidy is plotted against a PC that is heavily weighted on the p-H3 biomarker that TB and AKPI are clearly differentiated. This clarifies why PC3 versus PC1, rather than PC2 versus PC1, differentiates these subtypes. Additionally, regarding weighting, besides using the length of the PCs plots' arrows to assess weighting of factors, one can also see this from a contribution of variables plot, bottom right of Figure 5. One final comment about the PC3 versus PC1 plot: one somewhat intermediate blue square can be found midway between the group of TB and AKPI agents (at approximately the 6,- 3 position). This is carbendazim, a TB whose response profile is dominated by robust polyploidization, a result that is shared by the AKPIs more so than other TBs. This is an example of PCA analysis representing a useful device for helping highlight a somewhat unusual/outlier biomarker response profile.

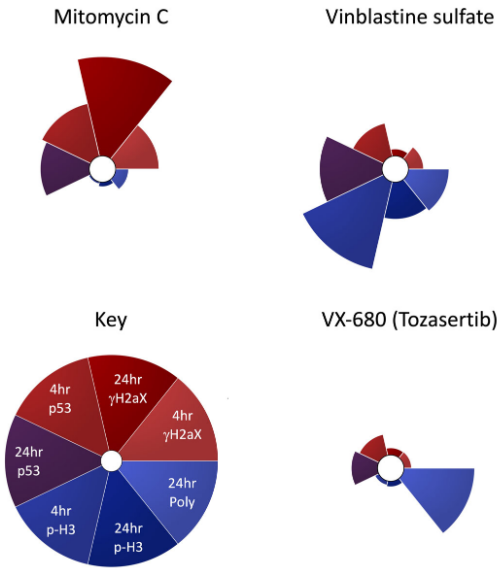


FIGURE 6 ToxPi profiles are shown for a representative clastogen (mitomycin C), tubulin binder-type aneugen (vinblastine), and aurora kinase inhibitor-type aneugen (VX-680). For each slice, the distance from the origin represents a particular biomarker's AUC value (larger values = longer protrusions). As shown in the Key, the clastogen-responsive biomarkers have been colored red and are arranged near the top; the aneugen-responsive biomarkers have been colored blue and are arranged near the bottom; and the pan-genotoxic biomarker 24 h p53 has been colored purple. Given this arrangement, clastogens tend to exhibit top-heavy profiles, whereas aneugens are bottom-heavy.

PCA is a mature, widely used, highly informative tool for evaluating high-dimensional datasets. It assumes linear relationships among factors, and tends to emphasize local data structure at the expense of global structure. While PCA is a computationally efficient method, its scalability is limited relative to certain alternative approaches such as MDS, t-SNE, and UMAP, which are described below.

3.6 | ToxPi

Figure 6 displays three ToxPi profiles, mitomycin C, vinblastine, and VX-680. As seen in the associated key, clastogen responsive

10982280, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/em.22604 by Test, Wiley Online Library on [17/05/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

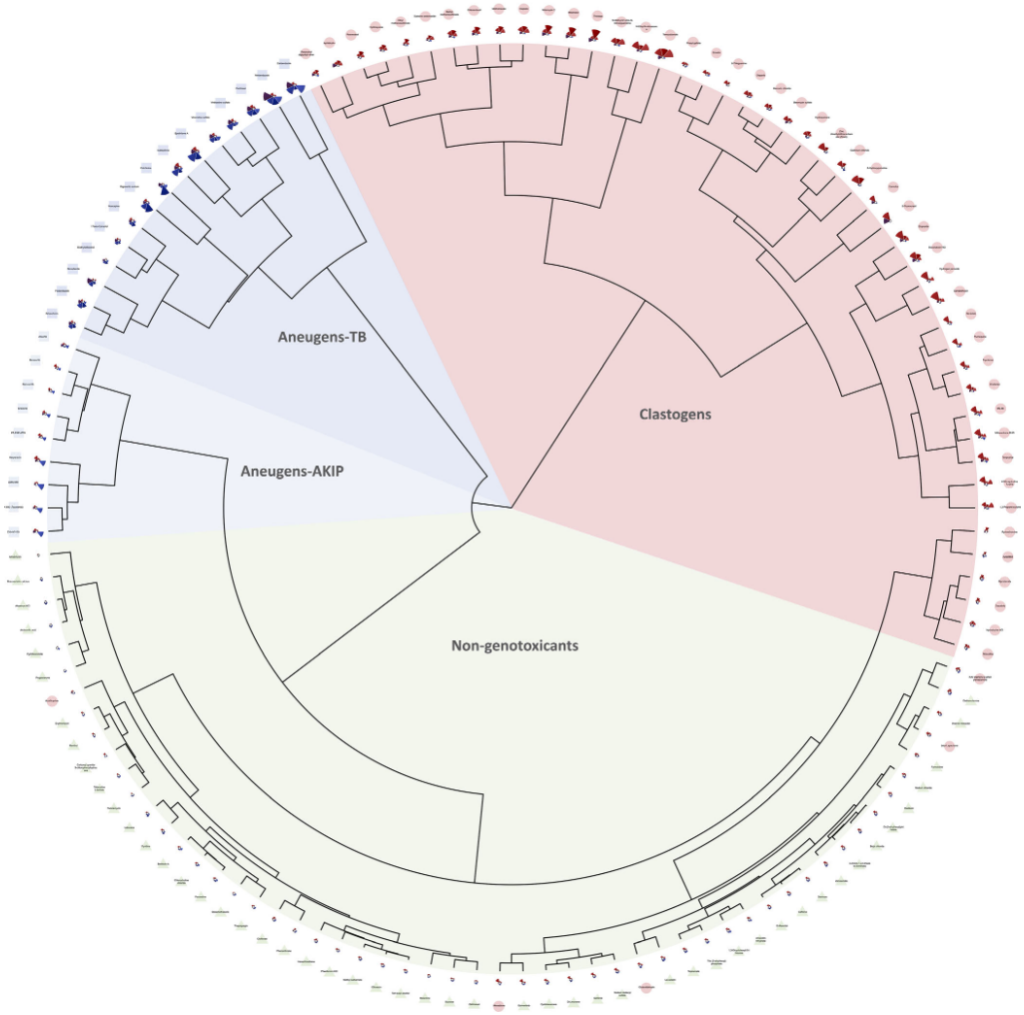


FIGURE 7 ToxPi profiles are shown for 7 biomarker responses and 126 chemicals. The same biomarker arrangement described in Figure 6 is used here. This image is the result of ToxPi software's hierarchical clustering algorithm, which places like-responding chemicals near each other. The clustering function shows good overall discrimination of clastogens, two sub-types of aneugens, and non-genotoxicants. Note that the outer perimeter of image color-codes the chemicals according to their *a priori* mode of action: clastogens (red circles), aneugens (blue squares), and non-genotoxicants (green triangles).

biomarkers (4 h p53, 4 h γH2AX, and 24 γH2AX) are shaded red, aneugen responsive biomarkers (4 h p-H3, 24 h p-H3, and 24 h polyploidy) are shaded blue, and the pan-genotoxic biomarker (24 h p53) is purple. These ToxPi depictions were generated as part of the 126 compound ToxPi analysis, utilizing AUC for each biomarker response. ToxPi slice projections illustrate biomarker responses relative to responses in the entire ($n = 126$) compound set, facilitating comparative analysis. As seen in Figure 6, mitomycin C's ToxPi profile shows a clear clastogenic biomarker signature with notable projections in p-53 and γH2AX. The vinblastine ToxPi profile demonstrates an aneugenic biomarker profile reflective of TBs (increased p-H3 and polyploidy). VX-680 also depicts an aneugenic response, but in this case, one that is reflective of an AKIP profile (decreased p-H3 and increased polyploidy). These clear depictions provide a

10982280, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/em.22604 by Test, Wiley Online Library on [17/05/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

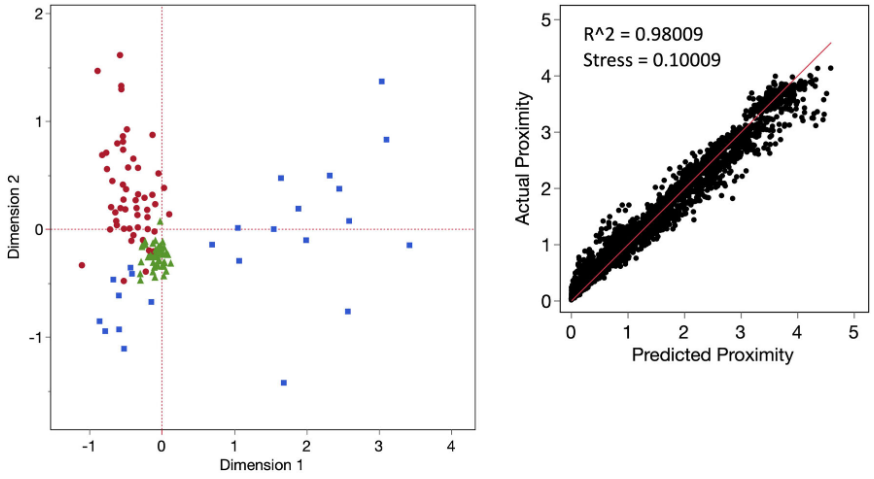


FIGURE 8 Multidimensional scaling plot displays the two-dimensional representation of the MultiFlow biomarker data for 7 biomarker responses and 126 chemicals. The horizontal and vertical red lines through the origin helps to identify contrasts with the intension to aid in the interpretation of each component. The scatterplot is accompanied with the Shepard plot to visualize the goodness of fit of the MDS solution. Note that the chemicals were color-coded according to their *a priori* mode of action category: clastogens (red), aneugens (blue), and non-genotoxicants (green).

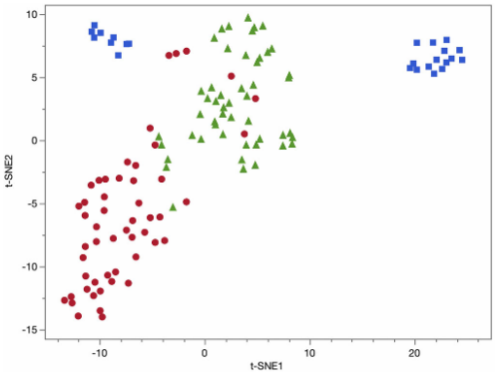


FIGURE 9 Scatter plot of the t-SNE solution for 7 biomarker responses and 126 chemicals. Four clusters are evident, and correspond to chemicals' *a priori* mode of action: clastogens (red), two aneugen sub-types (blue), and non-genotoxicants (green).

straightforward representation of multiplexed biomarker responses, while providing visual insight into genotoxic MoA. While biomarker responses can easily be compared using ToxPi, the use of AUC removes insights into dose response relationship and therefore genotoxic potency.

One advantageous feature of ToxPi software is that it provides algorithms for hierarchical clustering. In Figure 7, a circular hierarchical clustering map was generated using ToxPi's ward.D2 algorithm ($n = 126$ chemicals). This analysis shows clusters that reflect several different genotoxic modes of action—clastogens, TB-type aneugens, AKIP-type aneugens, and non-genotoxicants. Altogether, ToxPi provides insightful visuals and clustering analysis that can help contextualize multidimensional data sets comprised of large numbers of chemicals.

3.7 | MDS

Figure 8 provides results from JMP software's MDS platform. For assessing the goodness of fit of the MDS analysis, the stress and R-square value were calculated and a Shepard plot was generated. Stress quantifies the discrepancy between the original distances or dissimilarities and those derived from the MDS solution. Lower stress values indicate better fit. Generally, stress values below 0.1 are excellent, indicating a strong fit between the observed dissimilarities and the distances in the MDS solution. Stress values between 0.1 and 0.2 are considered good, suggesting a reasonable fit but with some discrepancies. Stress values exceeding 0.2 may indicate a poor fit, suggesting that the MDS solution does not adequately capture the original dissimilarities (Dexter et al., 2018; Kruskal, 1964). R-square measures the proportion of variance explained by the MDS solution and is visually represented using a Shepard plot. Taken together, the low stress (0.10009) and high R^2 values (0.98009) observed here are indicative of a good fit.

The image on the left is the actual MDS plot, with Dimension 2 graphed against Dimension 1. As with the PCA plots, each dot is a

10982280, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/em.22604 by Test, Wiley Online Library on [17/05/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

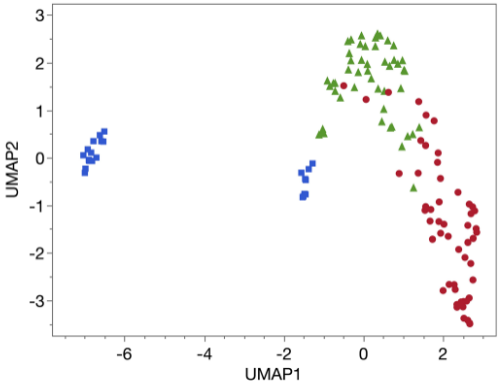


FIGURE 10 Scatter plot of the UMAP embeddings for 7 biomarker responses and 126 chemicals. The embeddings are color-coded, and correspond to chemicals' *a priori* mode of action: clastogens (red), two aneugen sub-types (blue), and non-genotoxicants (green).

different chemical, with red, green and blue corresponding to an *a priori* mode of action. Whereas non-genotoxicants form a fairly tight cluster near the origin (0,0), clastogens are observed to fall somewhat above the non-genotoxicants and extend toward 11 o'clock. Since the majority of the aneugens appear furthest away relative to the other MoA groups, one can say that these MDS results suggest that aneugens' data structure (i.e., biomarker response profiles) exhibit the greatest dissimilarities. As was observed with PCA, it is apparent that MDS is also able to resolve differences between aneugen subtypes, as the AKIP-type agents are the group that extents toward 7 o'clock while TBs appear to the right.

In summary, MDS is a useful method that constructs a low-dimensional representation of data by preserving the pairwise distance between data points as much as possible. While MDS can be computationally expensive when data sets are large, it is useful to consider this approach whenever one is interested in striking a balance between the preservation of local and global data structure.

3.8 | t-SNE

Unlike PCA, which yields a deterministic final solution, t-SNE exhibits stochastic behavior when multiple minima exist, potentially leading to divergent solutions. Consequently, multiple runs may be needed to investigate the reproducibility of the final result. While t-SNE adeptly captures local or neighboring observations, its ability to represent global trends accurately can be limited. Figure 9 displays the 2-dimensional representation generated by t-SNE. From this analysis, four main clusters emerge: two in the upper left and upper right of the plot, corresponding to aneugens, one in the lower left, representing clastogens, and a fourth cluster in the center of the plot representing non-genotoxicants.

When employing t-SNE, the embedding is directly learned from the data, lacking an eigenvector or analogous construct for accommodating new data. In contrast, PCA's eigenvectors offer a straightforward means of projecting new data. Like MDS, t-SNE is also computationally expensive when data set are large and the original implementation of t-SNE had quadratic memory and runtime O(N^2) thus suffering from the curse of dimensionality. Using the Barnes-Hut t-SNE algorithm improves O(N log N). Using the R environment, the original or exact t-SNE can be conducted by setting theta to 0 (speed/accuracy trade-off, default = 0.5).

3.9 | UMAP

The UMAP algorithm is competitive with t-SNE for visualization quality, and arguably preserves more of the global structure with superior run time performance. It is viable as a general-purpose dimension reduction technique for machine learning as it has no computational restrictions on embedding dimension. UMAP is not designed for very small datasets as the optimization choices that are made assumes reasonable dataset size. In practice it is probably best not to use UMAP on small data sets.

Figure 10 shows the UMAP representation of these data. There appears to be a continuum of points with a slight slope with clastogens with negative component 2 values on the right moving towards non-genotoxicants with positive component 1 values. Aneugens appear in two locations in this plot: one as a cluster of chemicals at the far-left side of the plot, and a second group adjacent to the non-genotoxicants with lower component 1 and 2 values. In the literature, UMAP has shown significant advantages in data reduction and visualization. For these data the linear approaches such as PCA and MDS are performing equally as well and that may be due to the relatively small sample size. Arguably the sample size of these data is on the lower end of what is recommended in the literature for the application of this method.

4 | CONCLUSIONS

For the use case considered here—describing the MultiFlow assay's biomarker response profiles for reference clastogens, aneugens, and non-genotoxicants—each of the nine visualization techniques had merit. Some appeared most suited for describing the biomarker signature of one to several chemicals, especially scatter and spider plots. On the other hand, the dimensionality reduction techniques (parallel coordinate plots, hierarchical clustering, PCA, ToxPi, MDS, t-SNE, UMAP) excelled at highlighting generalizable response patterns for many test chemicals. For the current data set, rather similar clustering was observed across these methods for non-genotoxicants, clastogens, and the two aneugen subtypes. These similarities imply that the

10982280, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/em.22604 by Test, Wiley Online Library on [17/05/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

various reduced dimensionality representations adequately summarized genuine features of the data, and were not artifacts of the specific techniques used.

Given the increased use of ML approaches in (geno)toxicology data analysis pipelines (Bryce et al., 2016; Buick et al., 2020, 2021, 2022; Dertinger et al., 2019; Fortin et al., 2023; Thienpont et al., 2023; Wilson et al., 2021), visualization techniques such as those described herein will often represent an important compliment that helps stakeholders interpret and better understand the algorithms' output. Thus, in the case of MultiFlow assay results, the visuals help explain why a particular genotoxic MoA classification has been made.

While we have focused on visualization techniques that help explain ML model predictions, it is important to recognize that they offer additional benefits. For instance, they can also help facilitate the construction and/or optimization of ML models (see Supplemental File 3). That is, by analyzing the distribution of biomarker responses in reduced dimensionality space, one can prioritize certain biomarkers for further investigation or feature selection. Clustering can be useful for ML model evaluation by comparing clustering of data points in the reduced space with the ML classifications to assess how well the model captures the inherent structure of the data. Also, by visualizing data points that are far from the main clusters, these techniques may also be useful for detecting outliers/anomalies. Thus, in the context of a chemical's biomarker response pattern, these dimensionality reduction techniques can be useful for qualifying a ML model's prediction, even when that prediction has been represented as a high probability value.

In summary, in an era where multiplexed assays and ML algorithms are becoming the norm, stakeholders should find the visualization strategies described here useful for efficiently and effectively interpreting their high-dimensional data. Furthermore, dimensionality reduction techniques can offer valuable insights beyond visualization, as they can be leveraged at various stages of the ML pipeline to improve the interpretability, performance, and reliability of a classification task based on biomarker response profiles.

AUTHOR CONTRIBUTIONS

All authors contributed to the outlining of the manuscript. Statistical analyses were conducted by EB, YH, SB, AW and SD. SD, AW, EB and YH provided an initial draft, and all authors contributed to the extensive editing of the manuscript.

ACKNOWLEDGMENTS

This work was funded by a grant from the National Institute of Health/National Institute of Environmental Health Sciences (NIEHS; grant no. R44 ES033138). The contents are solely the responsibility of the authors, and do not necessarily represent the official views of the NIEHS.

CONFLICT OF INTEREST STATEMENT

Several authors are employed by Litron Laboratories. Litron holds patents for flow cytometry-based analyses describe herein and sells the MultiFlow DNA Damage Kit—p53, gamma-H2AX, Phospho-Histone H3 that was used to generate the data described herein. Health Canada does not endorse or recommend the products or services of any commercial entity, including Litron Laboratories.

DATA AVAILABILITY STATEMENT

The data that supports the findings of this study are available in the supplementary material of this article.

ORCID

George E. Johnson https://orcid.org/0000-0001-5643-9942

REFERENCES

Attia, S.M., Aleisa, A.M., Bakheet, S.A., Al-Yahya, A.A., Al-Rejaie, S.S., Ashour, A.E. et al. (2009) Molecular cytogenetic evaluation of the mechanism of micronucleus formation induced by camptothecin, topotecan, and irinotecan. *Environmental and Molecular Mutagenesis*, 50, 145–151.

Avlasevich, S., Pellegrin, T., Godse, M., Bryce, S., Bemis, J., Bajorski, P. et al. (2021) Biomarkers of DNA damage response improve in vitro micronucleus assays by revealing genotoxic mode of action and reducing the occurrence of irrelevant positive results. *Mutagenesis*, 36, 407–418.

Aydemir, N. & Bilaloğlu, R. (2003) Genotoxicity of two anticancer drugs, gemcitabine and topotecan, in mouse bone marrow in vivo. *Mutation Research*, 537, 43–51.

Baker, S.J., Cosenza, S.C., Athuluri-Divakar, S., Reddy, M.V.R., Vasquez-Del Carpio, R., Jain, R. et al. (2020) A contaminant impurity, not rigosertib, is a tubulin binding agent. *Molecular Cell*, 79, 180–190.

Basso, E., Fiore, M., Leone, S., Degrassi, F. & Cozzi, R. (2013) Effects of resveratrol on topoisomerase II-a activity: induction of micronuclei and inhibition of chromosome segregation in CHO-K1 cells. *Mutagenesis*, 28, 243–248.

Bernacki, D.T., Bryce, S.M., Bemis, J.C. & Dertinger, S.D. (2019) Aneugen molecular mechanism assay: proof-of-concept with 27 reference chemicals. *Toxicological Sciences*, 170, 382–393.

Bollag, D.M., McQueney, P.A., Zhu, J., Hensens, O., Koupal, L., Liesch, J. et al. (1995) Epothilones, a new class of microtubule-stabilizing agents with a taxol-like mechanisms of action. *Cancer Research*, 55, 2325–2333.

Borg, I. & Groenen, P. (2005) *Modern multidimensionality scaling: theory and applications*. New York: Springer.

Bryce, S.M., Bernacki, D.T., Bemis, J.C. & Dertinger, S.D. (2016) Genotoxic mode of action predictions from a multiplexed flow cytometric assay and a machine learning approach. *Environmental and Molecular Mutagenesis*, 57, 171–189.

Bryce, S.M., Bernacki, D.T., Bemis, J.C., Spellman, R.A., Engel, M.E., Schuler, M. et al. (2017) Interlaboratory evaluation of a multiplexed high information content in vitro genotoxicity assay. *Environmental and Molecular Mutagenesis*, 58, 146–161.

Bryce, S.M., Bernacki, D.T., Smith-Roe, S.L., Witt, K.L., Bemis, J.C. & Dertinger, S.D. (2018) Investigating the generalizability of the MultiFlow® DNA damage assay and several machine learning models with a set of 103 diverse test chemicals. *Toxicological Sciences*, 162, 146–166.

Buick, J.K., Rowan-Carroll, A., Gagné, R., Williams, A., Chen, R., Li, H.H. et al. (2022) Integrated genotoxicity testing of three anti-infective drugs using the TGx-DDI transcriptomic biomarker and high-throughput CometChip® assay in TK6 cells. *Frontiers in Toxicology*, 4, 991590.

Buick, J.K., Williams, A., Gagné, R., Swartz, C.D., Recio, L., Ferguson, S.S. et al. (2020) Flow cytometric micronucleus assay and TGx-DDI transcriptomic biomarker analysis of ten genotoxic and non-genotoxic chemicals in human HepaRG™ cells. *Genes and Environment*, 42, 5.

10982280, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/em.22604 by Test, Wiley Online Library on [17/05/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Buick, J.K., Williams, A., Meier, M.J., Swartz, C.D., Recio, L., Gagné, R. et al. (2021) A modern genotoxicity testing paradigm: integration of the high-throughput CometChip® and the TGx-DDI transcriptomic biomarker in human HepaRG™ cell cultures. *Frontiers in Public Health*, 9, 694834.

Camptosar® (Irinotecan) package insert (2014) Pfizer Inc, NY, NY. Available from: https://www.accessdata.fda.gov/drugsatfda_docs/label/2014/020571s048lbl.pdf Accessed February 22, 2024.

Carpinelli, P., Ceruti, R., Giorgini, M.L., Cappella, P., Gianellini, L., Croci, V. et al. (2007) PHA-739358, a potent inhibitor of aurora kinases with a selective target inhibition profile relevant to cancer. *Molecular Cancer Therapeutics*, 6, 3158–3168.

Chapman, K.E., Wile, E.C., Chapman, F.M., Verma, J.R., Shah, U.-K., Stannard, L.M. et al. (2021) Multiple-endpoint in vitro carcinogenicity test in human cell line TK6 distinguishes carcinogens from non-carcinogens and highlights mechanisms of action. *Archives of Toxicology*, 95, 321–336.

Chen, G., Ning, B. & Shi, T. (2019) Single-cell RNA-seq technologies and related computational data analysis. *Frontiers in Genetics*, 10, 317.

Chinnasamy, N., Rafferty, J.A., Hickson, I., Ashby, J., Tinwell, H., Margison, G.P. et al. (1997) O6-benzylguanine potentiates the in vivo toxicity and clastogenicity of temozolomide and BCNU in mouse bone marrow. *Blood*, 89, 1566–1573.

Clofibrate, IARC monograph (1996) Available from: http://monographs.iarc.fr/ENG/Monographs/vol66/mono66-17.pdf [Accessed 22th February 2024].

Cojocel, C., Novotny, L. & Vachalkova, A. (2006) Mutagenic and carcinogenic potential of menadione. *Neoplasma*, 53, 316–323.

Corton, J.C., Williams, A. & Yauk, C.L. (2018) Using a gene expression biomarker to identify DNA damage-inducing agents in microarray profiles. *Environmental and Molecular Mutagenesis*, 59, 772–784.

de Graaf, A.O., van den Heuvel, L.P., Dijkman, H.B., de Abreu, R.A., Birkenkamp, K.U., de White, T. et al. (2004) Bcl-2 prevents loss of mitochondria in CCCP-induced apoptosis. *Experimental Cell Research*, 299, 533–540.

Degrassi, F., De Salvia, R. & Berghella, L. (1993) The production of chromosomal alterations by β-lapachone, an activator of topoisomerase I. *Mutation Research*, 288, 263–267.

DeMarini, D.M., Brock, K.H., Doerr, C.L. & Moore, M.M. (1987) Mutagenicity and clastogenicity of teniposide (VM-26) in L5178Y/TK +/- 3.7.2C mouse lymphoma cells. *Mutation Research*, 187, 141–149.

Dertinger, S.D., Kraynak, A.R., Wheeldon, R.P., Bernacki, D.T., Bryce, S.M., Hall, N. et al. (2019) Predictions of genotoxic potential, mode of action, molecular targets, and potency *via* a tiered MultiFlow® assay data analysis strategy. *Environmental and Molecular Mutagenesis*, 60, 513–533.

Dertinger, S.D., Phonethepswath, S., Avlasevich, S.L., Torous, D.K., Mereness, J., Bryce, S.M. et al. (2012) Efficient monitoring of in vivo *Pig-a* gene mutation and chromosomal damage: Summary of 7 published studies and results from 11 new reference compounds. *Toxicological Sciences*, 130, 328–348.

Dertinger, S.D., Phonethepswath, S., Weller, P., Avlasevich, S., Torous, D.K., Mereness, J.A. et al. (2011) Interlaboratory *Pig-a* gene mutation assay trial: studies of 1,3-propane sultone with immunomagnetic enrichment of mutant erythrocytes. *Environmental and Molecular Mutagenesis*, 52, 748–755.

Dexter, E., Rollwagen-Bollens, G. & Bollens, S.M. (2018) The trouble with stress: a flexible method for the evaluation of nonmetric multidimensional scaling. *Limnology and Oceanography: Methods*, 16, 434–443.

Diaz, D., Scott, A., Carmichael, P., Shi, W. & Costales, C. (2007) Evaluation of an automated in vitro micronucleus assay in CHO-K1 cells. *Mutation Research*, 630, 1–13.

Everitt, B. (1974) *Cluster Analysis*. London: Heinemann Educational Books Ltd.

Floxin® (Ofloxacin) package insert. (2008) Ortho-McNeil, Raritan, NJ. Available from: http://www.accessdata.fda.gov/drugsatfda_docs/label/2008/019735s059lbl.pdf [Accessed 22th February 2024].

Fortin, A.V., Long, A.S., Williams, A., Meier, M.J., Cox, J., Pinsonnault, C. et al. (2023) Application of a new approach methodology (NAM)-based strategy for genotoxicity assessment of data-poor compounds. *Frontiers in Toxicology*, 5, 1098432.

Futami, T., Miyagishi, M. & Taira, K. (2005) Identification of a network involved in thapsigargin-induced apoptosis using a library of small interfering RNA expression vectors. *The Journal of Biological Chemistry*, 280, 826–831.

Gewirtz, D.A. (1999) A critical evaluation of the mechanisms of action proposed for the antitumor effects of the anthracycline antibiotics adriamycin and daunorubicin. *Biochemical Pharmacology*, 57, 727–741.

Gleevec (Imatinib mesylate) package insert (2001) Novartis, East Hanover, NJ. Available from: http://www.accessdata.fda.gov/drugsatfda_docs/label/2008/021588s024lbl.pdf [Accessed 22th February 2024].

Glover, T.W., Berger, C., Coyle, J. & Echo, B. (1984) DNA polymerase alpha inhibition by aphidicolin induces gaps and breaks at common fragile sites in human chromosomes. *Human Genetics*, 67, 136–142.

Gocke, E., Bürgin, H., Müller, L. & Pfister, T. (2009) Literature review on the genotoxicity, reproductive toxicity, and carcinogenicity of ethyl methanesulfonate. *Toxicology Letters*, 190, 254–265.

Gollapudi, P., Hasegawa, L.S. & Eastmond, D.A. (2014) A comparative study of the aneugenic and polyploidy-inducing effects of fisetin and two model Aurora kinase inhibitors. *Mutation Research*, 767, 37–43.

Gulati, D.K., Witt, K., Anderson, B., Zeiger, E. & Shelby, M.D. (1989) Chromosome aberration and sister chromatid exchange tests in Chinese hamster ovary cells in vitro III: results with 27 chemicals. *Environmental and Molecular Mutagenesis*, 13, 133–193.

Hall, N.E., Tichenor, K., Bryce, S.M., Bemis, J.C. & Dertinger, S.D. (2022) In vitro human cell-based aneugen molecular mechanism assay. *Environmental and Molecular Mutagenesis*, 63, 151–161.

Han, C., Nam, M.K., Park, H.J., Seong, Y.M., Kang, S. & Rhim, H. (2008) Tunicamycin-induced ER stress upregulates the expression of mitochondrial HtrA2 and promotes apoptosis through the cytosolic release of HtrA2. *Journal of Microbiology and Biotechnology*, 18, 1197–1202.

Hartigan, J.A. (1975) *Clustering Algorithms*. New York: Wiley.

Hauf, S., Cole, R.W., LaTerra, S., Zimmer, C., Schnapp, G., Walter, R. et al. (2003) The small molecule Hesperadin reveals a role for Aurora B in correcting kinetochore-microtubule attachment and in maintaining the spindle assembly checkpoint. *The Journal of Cell Biology*, 161, 281–294.

Henderson, L., Fedyk, J., Windebank, S. & Smith, M. (1993) Induction of micronuclei in rat bone marrow and peripheral blood following acute and subchronic administrate of azathioprine. *Mutation Research*, 291, 79–85.

Hendriks, G., Atallah, M., Morolli, B., Calleja, F., Ras-Verloop, N., Huijskens, I. et al. (2012) The ToxTracker assay: novel GFP reporter systems that provide mechanistic insight into the genotoxic properties of chemicals. *Toxicological Sciences*, 125, 285–298.

Hernández, L.G., van Benthem, J. & Johnson, G.E. (2013) A mode-of action approach for the identification of genotoxic carcinogens. *PLoS One*, 8, e64532.

Inselberg, A. (1985) The plane with parallel coordinates. *Visual Computing*, 1, 69–91.

Irressa® (Gefitinib) package insert (2003) AstraZeneca Pharmaceuticals LP, Wilmington, DE. Available from: https://www.accessdata.fda.gov/drugsatfda_docs/label/2003/021399lbl.pdf [Accessed 22th February 2024].

Islam, M.M. & Mirza, S.P. (2022) Versatile use of carmofur: a comprehensive review of its chemistry and pharmacology. *Drug Development Research*, 83, 1505–1518.

Jani, J.P., Arcari, J., Bernardo, V., Bhattacharya, S.K., Briere, D., Cohen, B.D. et al. (2010) PF-03814735, an orally bioavailable small molecule

10982280, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/em.22604 by Test, Wiley Online Library on [17/05/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

aurora kinase inhibitor for cancer therapy. *Molecular Cancer Therapeutics*, 9, 883–894.

Kashida, Y., Sasaki, Y.F., Ohsawa, K., Yokohama, N., Takahashi, A., Watanabe, T. et al. (2002) Mechanistic study on flumequine hepatocarcinogenicity focusing on DNA damage in mice. *Toxicological Sciences*, 69, 317–321.

Keshava, C., Keshava, N., Whong, W.Z., Nath, J. & Ong, T.M. (1998) Inhibition of methotrexate-induced chromosomal damage by folinic acid in V79 cells. *Mutation Research*, 397, 221–228.

Kimura, A., Miyata, A. & Honma, M. (2013) A combination of in vitro comet assay and micronucleus test using human lymphoblastoid TK6 cells. *Mutagenesis*, 28, 583–590.

Kirkland, D., Kasper, P., Martus, H.-J., Müller, L., van Benthem, J., Madia, F. et al. (2016) Updated recommended lists of genotoxic and non-genotoxic chemicals for assessment of the performance of new or improved genotoxicity tests. *Mutation Research*, 795, 7–30.

Kirkland, D., Kasper, P., Müller, L., Corvi, R. & Speit, G. (2008) Recommended lists of genotoxic and non-genotoxic chemicals for assessment of the performance of new or improved genotoxicity tests: a follow-up to an ECVAM workshop. *Mutation Research*, 653, 99–108.

Klein, C.B. & King, A.A. (2007) Genistein genotoxicity: critical considerations of in vitro exposure dose. *Toxicology and Applied Pharmacology*, 224, 1–11.

Kondo, Y., Honda, S., Nakajima, M., Miyahana, K., Hayashi, M., Shinagawa, Y. et al. (1992) Micronucleus test with vincristine sulfate and colchicine in peripheral blood reticulocytes of mice using acridine orange supravital staining. *Mutation Research*, 278, 187–191.

Kong, Y., Bender, A. & Yan, A. (2018) Identification of novel aurora kinase A (AURKA) inhibitors via hierarchical ligand-based virtual screening. *Journal of Chemical Information and Modeling*, 58, 36–47.

Krewski, D., Saunders-Hastings, P., Baan, R.A., Barton-Maclaren, T.S., Browne, P., Chiu, W.A. et al. (2022) Development of an evidence-based risk assessment framework. *ALTEX*, 39, 667–693.

Krishna, G., Urda, G., Tefera, W., Lalwani, N.D. & Theiss, J. (1995) Simultaneous evaluation of dexamethasone-induced apoptosis and micronuclei in rat primary spleen cell cultures. *Mutation Research*, 332, 1–8.

Kruskal, J.B. (1964) Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, 29, 1–27.

Kurihara, D., Matsunaga, S., Kawabe, A., Fujimoto, S., Noda, M., Uchiyama, S. et al. (2006) Aurora kinase is required for chromosome segregation in tobacco BY-2 cells. *The Plant Journal*, 48, 572–580.

Lee, F.Y.F., Borzilleri, R., Fairchild, C.R., Kim, S.H., Long, B.H., Reventos-Suarez, C. et al. (2001) BMS-247550: a novel epothilone analog with a mode of action similar to paclitaxel but possessing superior antitumor efficacy. *Clinical Cancer Research*, 7, 1429–1437.

Li, Y., Luan, Y., Qi, X., Li, M., Gong, L., Xue, X. et al. (2010) Emodin triggers DNA double-strand breaks by stabilizing topoisomerase II-DNA cleavage complexes and by inhibiting ATP hydrolysis of topoisomerase II. *Toxicological Sciences*, 118, 435–443.

Lidoderm® (Lidocaine patch) package insert (2004) Endo Pharmaceuticals, Chad's Ford, PA. Available from: https://www.accessdata.fda.gov/drugsatfda_docs/label/2005/020612s007lbl.pdf [Accessed 22th February 2024].

Lotz, A.S., Havla, J.B., Richter, E., Frölich, K., Staudenmaier, R., Hagen, R. et al. (2009) Cytotoxic and genotoxic effects of matrices for cartilage tissue engineering. *Toxicology Letters*, 190, 128–133.

Lu, P.-Z., Lai, C.-Y. & Chan, W.-H. (2008) Caffeine induces cell death via activation of apoptotic signal and inactivation of survival signal in human osteoblasts. *International Journal of Molecular Sciences*, 9, 698–718.

Lynparza™ (Olaparib) package insert (2014) AstraZeneca Pharmaceuticals LP, Wilmington, DE; Available from http://www.accessdata.fda.gov/drugsatfda_docs/label/2014/206162lbl.pdf [Accessed 22th February 2024].

Martelli, A., Allavena, A., Campart, G.B., Canonero, R., Ghia, M., Mattioli, F. et al. (1995) In vitro and in vivo testing of hydralazine genotoxicity. *The Journal of Pharmacology and Experimental Therapeutics*, 273, 113–120.

Marvel, S.W., To, K., Grimm, F.A., Wright, F.A., Rusyn, I. & Reif, D.M. (2018) ToxPi graphical user interface 2.0: dynamic exploration, visualization, and sharing of integrated data models. *BMC Bioinformatics*, 19, 80.

Matsushima, T., Hayashi, M., Matsuoka, A., Ishidate, M., Jr., Miura, K.F., Shimizu, H. et al. (1999) Validation study of the in vitro micronucleus test in a Chinese hamster lung cell line (CHL/IU). *Mutagenesis*, 14, 569–580.

McInnes, L. & Healy, J. (2018) UMAP: uniform manifold approximation and projection for dimension reduction. *ArXiv* abs/1802.03426.

Mevacor® (Lovastatin) package insert. (2012) *Merck Sharp & Dohme Corp*. NJ, USA: Whitehouse Station. Available from: https://www.accessdata.fda.gov/drugsatfda_docs/label/2012/019643s085lbl.pdf [Accessed 22th February 2024].

Moon, J.L., Kim, S.Y., Shin, S.W. & Park, J.W. (2012) Regulation of brefeldin A-induced ER stress and apoptosis by mitochondrial $NADP^+$-dependent isocitrate dehydrogenase. *Biochemical and Biophysical Research Communications*, 417, 760–764.

Nikolova, T., Dvorak, M., Jung, F., Adam, I., Krämer, E., Gerhold-Ay, A. et al. (2014) γH2AX assay for genotoxic and nongenotoxic agents: comparison of H2AX phosphorylation with cell death response. *Toxicological Sciences*, 140, 103–117.

Oliver, J., Meunier, J.R., Awogi, T., Elhajouji, A., Ouldelhkim, M.C., Bichet, N. et al. (2006) SFTG international collaborative study on in vitro micronucleus test V. Using L5178Y cells. *Mutation Research*, 607, 125–152.

Parry, E.M., Parry, J.M., Corso, C., Doherty, A., Haddad, F., Hermine, T.F. et al. (2002) Detection and characterization of mechanism of action of aneugenic chemicals. *Mutagenesis*, 17, 509–521.

Paulsson, B., Kotova, N., Grawé, J., Henderson, A., Granath, F., Golding, B. et al. (2003) Induction of micronuclei in mouse and rat by glycidamide, genotoxic metabolite of acrylamide. *Mutation Research*, 535, 15–24.

Paxil® (Paroxetine) package insert (2011) GlaxoSmithKline, Research Triangle Park, NC. Available from: http://www.accessdata.fda.gov/drugsatfda_docs/label/2011/020031s058s066,020710s022s030lbl.pdf [Accessed 22th February 2024].

Payton, M., Bush, T.L., Chung, G., Ziegler, B., Eden, P., McElroy, P. et al. (2010) Preclinical evaluation of AMG 900, a novel potent and highly selective pan-aurora kinase inhibitor with activity in taxane-resistant tumor cell lines. *Cancer Research*, 70, 9846–9854.

Pearson, K. (1901) On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*, 2, 559–572.

Pepcid® (Famotidine) package insert (2011) Merck Sharp & Dohme Corp., Whitehouse Station, NJ, USA: Available from: http://www.accessdata.fda.gov/drugsatfda_docs/label/2011/019462s037lbl.pdf [Accessed 22th February 2024].

Pognan, F., Beilmann, M., Boonen, H.C.M., Czich, A., Dear, G., Hewitt, P. et al. (2023) The evolving role of investigative toxicology in the pharmaceutical industry. *Nature Reviews Drug Discovery*, 22, 317–335.

Rosefort, C., Fauth, E. & Zanki, H. (2004) Micronuclei induced by aneugens and clastogens in mononucleate and binucleate cells using the cytokinesis block assay. *Mutagenesis*, 19, 277–284.

Salk, J.J. & Kennedy, S.R. (2020) Next-generation genotoxicology: using modern sequencing technologies to assess somatic mutagenesis and cancer risk. *Environmental and Molecular Mutagenesis*, 61, 135–151.

Schuler, M., Muehlbauer, P., Guzzie, P. & Eastmond, D.A. (1999) Noscapine hydrochloride disrupts themitotic spindle inmammalian cells and induces aneuploidy as well as polyploidy in cultured human lymphocytes. *Mutagenesis*, 14, 51–56.

Sehdev, V., Peng, D., Soutto, M., Washington, M.K., Revetta, F., Ecsedy, J. et al. (2012) The aurora kinase A inhibitor MLN8237 enhances

10982280, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/em.22604by Test, Wiley Online Library on [17/05/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

cisplatin-induced cell death in esophageal adenocarcinoma cells. *Molecular Cancer Therapeutics*, 11, 763–774.

Smart, D.J., Halicka, H.D., Schmuck, G., Traganos, F., Darzynkiewicz, Z. & Williams, G.M. (2008) Assessment of DNA double-strand breaks and gammaH2AX induced by the topoisomerase II poisons etoposide and mitoxantrone. *Mutation Research*, 641, 43–47.

Sprycel®. (2010) *Sprycel® (Dasatinib) package insert*. Princeton, NJ: Bristol-Myers Squibb Co. Available from: https://www.accessdata.fda.gov/drugsatfda_docs/label/2010/021986s7s8lbl.pdf [Accessed 22th February 2024].

Sun, X., Rubitski, E., Spellman, R.A., Engel, M. & Schuler, M. (2022) A new imagine plaform (iScreen) allows for the concurrent assessment of micronucleus induction and genotoxic mode of action in human A375 cells. *Environmental and Molecular Mutagenesis*, 63, 230–245.

Tagrisso™ (Osimertinib) package insert (2012) AstraZeneca Pharmaceuticals LP, Wilmington, DE. 2012. Available from: https://www.accessdata.fda.gov/drugsatfda_docs/label/2015/208065s000lbl.pdf [Accessed 22th February 2024].

Tayama, S. & Nakagawa, Y. (2001) Cytogenetic effects of propyl gallate in CHO-K1 cells. *Mutation Research*, 498, 117–127.

Thienpont, A., Verhulst, S., van Grunsven, L.A., Rogiers, V., Vanhaecke, T. & Mertens, B. (2023) Novel prediction models for genotoxicity based on biomarker genes in human HepaRG™ cells. *ALTEX*, 40, 271–286.

Tufte, E. (2001) *The visual display of quantitative information*, 2nd edition. Cheshire, CT: Graphics Press.

Tweats, D.J., Johnson, G.E., Scandale, I., Whitwell, J. & Evans, D.B. (2016) Genotoxicity of flubendazole and its metabolites in vitro and the impact of a new formulation on in vivo aneugenicity. *Mutagenesis*, 31, 309–321.

van der Maaten, L. & Hinton, J. (2008) Visualizing data using t-SNE. *Journal of Machine Learning Research*, 9, 2579–2605.

Van Hummelen, P., Elhajouji, A. & Kirsch-Volders, M. (1995) Clastogenic and aneugenic effects of three benzimidazole derivatives in the in vitro micronucleus test using human lymphocytes. *Mutagenesis*, 10, 23–29.

Verdoodt, B., Decordier, I., Geleyns, K., Cunha, M., Cundari, E. & Kirsch-Volders, M. (1999) Induction of polyploidy and apoptosis after exposure to high concentrations of the spindle poison nocodazole. *Mutagenesis*, 14, 513–520.

Voigt, W., Matsui, S., Yin, M.B., Burhans, W.C., Minderman, H. & Rustum, Y.M. (1998) Topoisomerase-I inhibitor SN-38 can induce DNA damage and chromosomal aberrations independent from DNA synthesis. *Anticancer Research*, 18, 3499–3505.

Wegman, E.J. (1990) Hyperdimensional data analysis using parallel coordinates. *Journal of the American Statistical Association*, 85, 664–675.

Wheeldon, R.P., Bernacki, D.T., Dertinger, S.D., Bryce, S.M., Bemis, J.C. & Johnson, G.E. (2020) Benchmark dose analysis of DNA damage biomarker responses provides compound potency and adverse outcome pathway information for the Topoisomerase II inhibitor class of compounds. *Environmental and Molecular Mutagenesis*, 61, 396–407.

Wilson, A., Grabowski, P., Elloway, J., Ling, S., Stott, J. & Doherty, A. (2021) Transforming early pharmaceutical assessment of genotoxicity: applying statistical learning to a high throughput, multi end point in vitro micronucleus assay. *Scientific Reports*, 11, 2535.

Yang, J., Ikezoe, T., Nishioka, C., Tasaka, T., Taniguchi, A., Kuwayama, Y. et al. (2007) AZD1152, a novel and selective aurora B kinase inhibitor, induces growth arrest, apoptosis, and sensitization for tubulin depolymerizing agent or topoisomerase II inhibitor in human acute leukemia cells in vitro and in vivo. *Blood*, 110, 2034–2040.

Youngblom, J.H., Wiencke, J.K. & Wolff, S. (1989) Inhibition of the adaptive response of human lymphocytes to very low doses of ionizing radiation by the protein synthesis inhibitor cycloheximide. *Mutation Research*, 227, 257–261.

Zerit® (Stavudine) packet insert (2002) Bristol-Myers Squibb Virology, Princeton, NJ. Available from: http://www.accessdata.fda.gov/drugsatfda_docs/label/2002/20412S017.pdf [Accessed 22th February 2024].

Zou, H.Y., Li, Q., Lee, J.H., Arango, M.E., McDonnell, S.R., Yamazaki, S. et al. (2007) An orally available small-molecule inhibitor of c-Met, PF-2341066, exhibits cytoreductive antitumor efficacy through antiproliferative and antiangiogenic mechanisms. *Cancer Research*, 67, 4408–4417.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Dertinger, S.D., Briggs, E., Hussien, Y., Bryce, S.M., Avlasevich, S.L., Conrad, A. et al. (2024) Visualization strategies to aid interpretation of high-dimensional genotoxicity data. *Environmental and Molecular Mutagenesis*, 1–23. Available from: https://doi.org/10.1002/em.22604

10982280, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/em.22604by Test, Wiley Online Library on [17/05/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Appendix 2. Published Literature (Hussien, Dertinger & Johnson, 2025)

Check for updates

Environmental and Molecular Mutagenesis

WILEY

Environmental and Molecular Mutagenesis | Environmental Mutagenesis and Genomics Society

RESEARCH ARTICLE OPEN ACCESS

Synthesizing Genotoxicity Results in the MultiFlow Assay With Point-of-Departure Analysis and ToxPi Visualization Techniques

Yusuf Hussien[1] | Stephen D. Dertinger[2] | George E. Johnson[1]

[1]Instiutue of Life Sciences, Swansea University, Swansea, UK | [2]Litron Laboratories, Rochester, New York, USA

Correspondence: Yusuf Hussien (957196@swansea.ac.uk)

Received: 13 December 2024 | **Revised:** 10 February 2025 | **Accepted:** 13 February 2025

Accepted by: D. Roberts

Funding: This work was supported by Health and Environmental Sciences Institute, HESI GTTC Fast Fund.

Keywords: benchmark dose analysis | data visualization | flow cytometry | genotoxicity | hierarchical clustering | ToxPi

ABSTRACT

In vitro genotoxicity has historically served a hazard identification role, with simple binary outcomes provided for each of several single endpoint assays. This will need to change, given: (i) efforts to curtail animal testing, (ii) the increased use of multiplexed in vitro assays and the ongoing development of NAMS, and (iii) the desire to holistically consider quantitative results from multiple biomarkers/endpoints that take potency into consideration. To help facilitate more quantitative analyses of multiple biomarkers and/or assay streams, we explored the combined use of PROAST and Toxicological Prioritization Index (ToxPi) software. As a proofofconcept, this investigation employed the MultiFlow DNA damage assay, focusing on γH2AX and p53 biomarkers at two time points, whereby 10 genotoxicants were evaluated in the presence and absence of rat liver S9 metabolic activation. Whereas PROAST was used to calculate BMD point estimates and confidence intervals (CIs), ToxPi synthesized the BMD results into visual, quantitative summaries conveying genotoxicity and metabolic properties. Our analyses suggest that ToxPi's data synthesis and visualization modules provide useful insights into compound response, chemical grouping, and genotoxic mechanisms. By integrating multiple data sources, we find that ToxPi offers a powerful complementary approach to traditional BMD CI graphs, particularly for the simultaneous analysis of multiple biomarkers, enhancing chemical potency analysis of complex datasets.

1 | Introduction

A variety of exogenous and endogenous sources of DNA damage, such as ionizing radiation, reactive oxygen species, replication errors, alkylating agents, and carcinogenic compounds, introduce mutations that lead to defective genetic information, faulty cell development, impaired protein production, damaged cellular function, cellular apoptosis, genetic conditions, and tumorigenesis, ultimately affecting the survival of humans and other organisms (Wilson and Hunt 2002; Song et al. 2015; Chatterjee and Walker 2017). Base excision repair, nucleotide excision repair, and mismatch repair are all mechanisms activated once DNA damage is detected at genetic checkpoints and help to cleave/excise mutations from genetic material (Hakem 2008). Using diverse assays and testing models to detect DNA damage and repair events is crucial for designing effective protocols to identify potential sources of genotoxicity (Kim et al. 2013).

Genotoxicity tests have been used for decades as chemical hazard identification tools, meaning they were used to provide a binary outcome on compound genotoxicity—yes or no. However, there is growing recognition among pharmacologists, toxicologists, and pharmaceutical regulators that this approach overlooks the ability of genotoxicity tests to provide detailed dose–response information

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2025 The Author(s). *Environmental and Molecular Mutagenesis* published by Wiley Periodicals LLC on behalf of Environmental Mutagenesis and Genomics Society.

and insights into mechanisms of action, which can significantly enhance risk assessments. This shift in perspective has been driven by advancements in experimental designs and data acquisition methods, which now enable the production of reliable and quantifiable dose–response information. This makes genotoxicity test results a valuable addition to risk assessments by providing more nuanced and actionable insights surrounding potential chemical hazards and their underlying genotoxic mechanisms (MacGregor et al. 2015; Heflich et al. 2019). Such an expanded use of genotoxicity data, including in vitro data, seems important, given the many initiatives aimed at curtailing animal testing, and integrating quantitative genotoxicity data into risk assessments helps meet the broader scientific and ethical demand for more predictive, human-relevant testing methods. This shift represents a step toward achieving more comprehensive safety evaluations while advancing the vision of an animal-free future in toxicology.

An important requirement for using genetic toxicology results for more than hazard identification purposes is that there must be a way to reliably evaluate dose–response relationships. Modern dose–response analysis metrics, such as the No Observed Genotoxicity Effect Level (NOGEL), offer an approach that identifies the highest dose at which no genotoxic effects are observed in a given experimental setup (FDA 2002). While useful in determining safety exposure levels, it only considers the doses tested in the experiment and does not account for the full dose–response relationship in a continuous manner.

The benchmark dose (BMD) model is a modeling approach that describes the dose (or concentration) that produces a predetermined change in response to a chemical being studied (also known as the critical effect size, or CES) to generate a statistically robust point-of-departure (PoD) metrics. Particularly, it produces upper and lower confidence interval bounds—the BMDU and BMDL, respectively—providing insight into model uncertainty and variability (Davis et al. 2011; Sturla 2018; FDA 2002; Dorato and Engelhardt 2005). The BMD also leverages the entire dose–response dataset to offer greater precision and flexibility regarding experimental design and dose selection, making it a more reliable tool in hazard characterization and information extraction from highly detailed dose–response datasets (Davis et al. 2011).

One of the tools used to extract the BMD from dose–response data is PROAST (v70.3): an R Gui–based (v4.2.2–4.4.0) software package that uses statistical modeling algorithms, such as the nested exponential/hill four-parameter models, to mathematically analyze concentration–response datasets (Wills et al. 2017). Based on a user-defined critical effect size (CES) that describes a predefined change in response above the mean background control, PROAST calculates BMD, BMDL, and BMDU metrics, and provides them in tabular form (RIVM 2021). The package also includes several useful visualizations that aid in the interpretation of datasets, providing quantitative data on compound potency and toxicity.

However, one complication to the use of PROAST and other BMD software packages, and for making effective use of quantitative genotoxicity data in general, is the increasing use of multiplexed in vitro assays, and/or the desire to integrate multiple data streams into an assessment. BMD software does not ordinarily accommodate multiple endpoints/biomarkers; rather, it focuses on a single response at a time. While the so-called covariate analysis option allows a user to simultaneously fit multiple curves, an underlying requirement is that the response metric is the very same, single endpoint/biomarker. This limitation highlights the need for complementary approaches capable of integrating and visualizing data from multiple biomarkers to provide a more holistic assessment, such as the Toxicological Prioritization Index (ToxPi) software.

ToxPi (Version 2.3: Marvel et al. 2018) is a JavaScript visualization tool designed to translate quantitative toxicity data into a single ToxPi score, ranking compounds on a continuous spectrum to better discriminate between potency groups (Reif et al. 2012). One of its greatest strengths is its ability to simplify and effectively integrate results from diverse assays or multiple biomarkers within a multiplexed assay. The tool generates ToxPi profiles—dynamically created pie charts that visually represent the distinct components and features of each dataset. Additionally, ToxPi incorporates hierarchical and K-means clustering algorithms using agglomerative nesting (AGNES) and principal component analysis (PCA) algorithms to group compounds based on similarities across the biomarkers assessed (Marvel et al. 2018).

The work described herein makes use of previously published MultiFlow assay data (Tian et al. 2020) and focuses on γH2AX and p53 biomarkers at two time points across 10 known genotoxicants with varying modes of action. Advantages realized through the combined use of PROAST and ToxPi are discussed to evaluate whether both applications can comprehensively compare and visualize original dose–response metrics for both parent compounds and their metabolites using a unique protocol described here.

2 | Methods

2.1 | In Vitro MultiFlow Datasets

The in vitro MultiFlow assay experiments were performed by Litron laboratories from their previously published article by (Tian et al. 2020) where 15 compounds were analyzed using their MultiFlow DNA damage kit—measuring the activation of p53 as a nuclei-free measurement, phosphorylation of H2AX at serine 139 (γH2aX) to detect double-stranded DNA breaks, and phosphorylation of H3 at serine 10 (p-H3) as a mitotic cell activity endpoint at 4/24-h timepoints, in the presence and absence of low dose 0.25% vol/vol S9 rat liver to measure metabolic activation effects. Methodologies and modeling used are further detailed in previous publications performed by (Slob and Setzer 2014; Bryce et al. 2016, 2017). As described in (Tian et al. 2020), a CES of 0.3 was chosen, representing a 30% change in response relative to the control, with exception to the RNC endpoint which had a value of −0.3. A summary of the 10/15 genotoxicants from the (Tian et al. 2020) study assessed in this analysis is listed in the Table 1, providing further insight into their respective modes of action and metabolic activation properties (including sources/citations).

2.2 | PROAST Parameters—BMD Analysis

The (Slob 2002) paper states that for toxicological continuous endpoints, a single dose–response model is unable to fully

10982280, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/em.70003 by Yusuf Hussien - Shibboleth - member@swan.ac.uk , Wiley Online Library on [14/03/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 1 | Test compounds, sources, modes of action, and metabolic properties.

Chemical (abbreviation)	CAS number, source—Sigma Aldrich	Primary mode of action, metabolic properties and references
Mitomycin C	50-07-7	Clastogen/mutagen, DNA cross-linking (dsDNA breaks), alkylating properties and oxidative damage stress (Kirkland et al. 2016; Higa et al. 1998).
Dibenzol[a,l]pyrene	191-30-0	Mutagen, polycyclic aromatic hydrocarbon, requires metabolic activation (thought to be primarily activated by CYP1A1, forming DNA adducts leading to mutations) (Arif and Gupta 1997).
Resorcinol	108-46-3	Clastogen (in vitro), in vitro mammalian cell positive (MLA assay with and without metabolic activation positive/in vitro human lymphocyte MN positive in absence of metabolic activation); in vitro findings not confirmed in vivo (Negative in mouse MN) (European Food Safety Authority 2010).
2-Aminoathracene	613-13-8	Mutagen with some clastogenic properties, aromatic amine, requires metabolic activation (CYP1B1, 2A family) (Carriére et al. 1992; Helma et al. 2001).
Benzo[a]pyrene	50-32-8	Clastogen/mutagen, polycyclic aromatic hydrocarbon, requires metabolic activation and forms bulky adducts (CYP1B1, 2A family) (Ramesh et al. 2007; Kirkland et al. 2016).
7,12-Dimethylbenzanthracene (DMBA)	57-97-6	Clastogen/Mutagen, requires metabolic activation (CYP1B1), forms bulky adducts (Kuklenyik et al. 2004; Kirkland et al. 2016).
2-amino-1-methyl-6-phenylimidazo [4,5-b] pyridine (PhIP)	105,650-23-5	Clastogen/mutagen, heterocyclic amine, requires metabolic activation (Kirkland et al. 2016; Krais et al. 2016).
Cyclophosphamide Monohydrate	6055-19-2	Mutagen/Clastogen, nitrogen mustard, requires metabolic activation (CYP2B6, CYP2C19, CYP2C9, and CYP3A4/5) (Kirkland et al. 2016; Rodriguez-Antona and Ingelman-Sundberg 2006; Khalil et al. 2018)
2-Actylaminofluorene	53-96-3	Mutagen with some clastogenic properties, requires metabolic activation (CYP1A2), forms C8 adduct on guanine (Otteneder and Lutz 1999; Kirkland et al. 2016; Sugimura et al. 2011).
Diethylnitrosamine	55-18-5	Mutagen with clastogenic properties, nitrosamine, O^6-alkylguanine DNA lesions. Requires metabolic activation (Bartsch and Montesano 1984; Yamazaki et al. 1992).

Note: Any abbreviations for compounds are included in brackets and will be subsequently used in the results/discussion sections.

encapsulate the underlying biological mechanisms, so the implementation of nested models is applied to the original concentration and induction values to map out the dose–response curve, thereby effectively estimating the BMDL and BMDU using the CES. Hence, BMD analysis was performed following the (Tian et al. 2020) protocol using PROAST v67.0 with conditions being outlined in the study and BMD figures from that study were collected for further visualization. The PROAST software package, employs nested Exponential and Hill coefficient models, known as the four-parameter exponential models, to analyze concentration–response datasets for each chemical's biomarker to cultivate BMDL and BMDU (90% CI) values (using bootstrapping), providing a quantitative assessment of compound potency and toxicity (Wills et al. 2017; RIVM 2021). PROAST now supports model averaging, a modern alternative to the traditional Hill or exponential modeling approaches. This method reduces bias by accounting for variability across models, quantifies uncertainty through integrated predictions, and applies model weights based on fit, determined by Akaike or Bayesian Information Criteria. (RIVM 2021). These metrics can then be visualized to explain the dose–response curve of each tested compound, as endorsed by the European Food Safety Authority (RIVM 2021). A CES value of 0.3 was selected for both γH2aX and p53 biomarkers, while polyploidy was set at 5.0 with the S9 conditions being the

10982280, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/em.70003 by Yusuf Hussien - <Shibboleth>-member@swan.ac.uk , Wiley Online Library on [14/03/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

covariate factor in the analysis, as described in (Tian et al. 2020). The dose and induction figures were selected as independent and response variables respectively and "exponential models 3 or 5 from various families of models" within PROAST was selected to fit the dose–response curve. For endpoints where the BMD was unable to be modeled (either due to response not reaching the CES or the true BMD estimate exceeding the tested concentration range), no figure was recorded. Where the BMDL and BMDU were calculated, the values were recorded for each chemical, biomarker, timepoint, and S9 activation properties.

2.3 | ToxPi Visualization

BMDL and BMDU figures were further visualized using ToxPi: A JavaScript program that visualizes and interacts with quantitative mutagenicity data, translating it into a single ToxPi score that scales mutagenicity on a continuous spectrum instead of fixed groups to better differentiate toxicity levels (Reif et al. 2012). BMDL and BMDU data generated through PROAST are initially transformed by fitting the BMDs within a realistic concentration testing range of 0.0001–10,000 μM as recommended by guidelines for various in vitro genotoxicity assays (OECD 2016a; OECD 2016b; ICH S2(R1) 2012). Therefore, all values below 0.0001 μM and any figures exceeding 10,000 μM are subsequently transformed to 0.0001 and 10,000 μM, respectively. Both BMDL and BMDU metrics are incorporated into individual ToxPi endpoint slices, weighting folds percentages can be defined for each endpoint, and a $-\log_{10}(x)$ transformation (to ensure more potent compounds (which would have lower BMD values) show as having higher ToxPi scoring) is applied, yielding ToxPi profiles (a profile that illustrates each compound and their potencies across all endpoints as a singular pie chart). The ToxPi score is a metric that is calculated when the sum of all input values across assays is transformed into a slice/component score for each compound. These scores are normalized

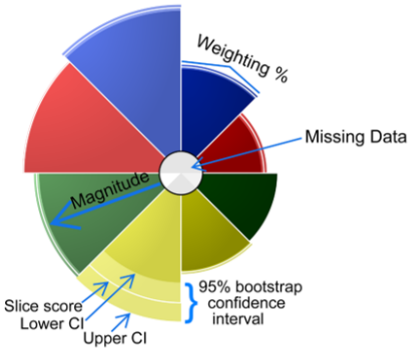


FIGURE 1 | ToxPi Profile key illustrating the features and properties that produce the visual details for individual chemicals in dose–response datasets. The magnitude of each slice represents the relative risk once the $-\log_{10}(x)$ transformation is applied, the width of each slice represents the weighting percentage applied to each slice, and the degree of graying within the center of the ToxPi profile represents the amount of missing data within the original dataset where increased darkening represents more missing individual data points. The 90% CI is represented by the pale section within the outer segment of each slice, where the inner edge represents the lower CI and is equivalent to the BMDU, the outer edge represents the outer CI and is equivalent to the BMDL, and the white line within the center represents the slice score for that endpoint. This profile was produced using MultiFlow data derived from the (Wheeldon et al. 2021; Tian et al. 2020) studies for the chemical resorcinol.

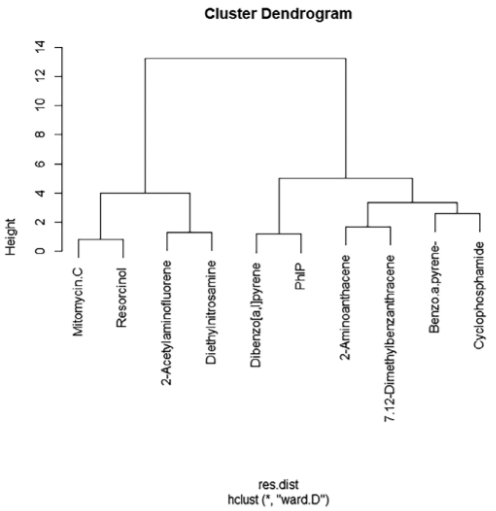


FIGURE 2 | Euclidean Hierarchical clustering generated through R Gui using the *hclust* function ward. D clustering algorithm to show the distinctive groupings of each compound based on their S9 potency ratios CI range values, as described in (Wheeldon et al. 2021) using the divisive difference between the S9− and S9+ of each endpoint tested: 4 h γH2aX, 4 h p53, 24 h γH2aX, and 24 h p53. The branch height signifies the degree of relatedness among compounds, with shorter heights denoting a higher degree of similarity.

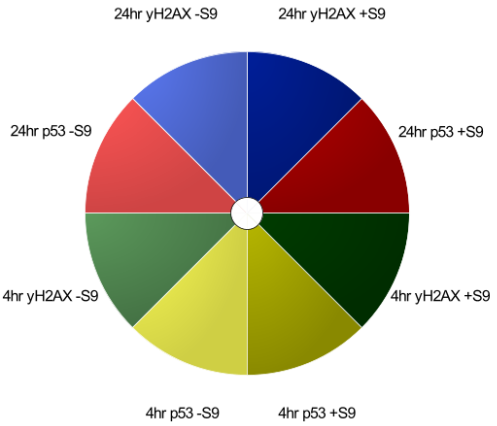


FIGURE 3 | ToxPi profile key illustrating each slice with its respective endpoints, including color-coordinated endpoints for γH2AX and p53 endpoints, at 4/24-h timepoints with S9 present (S9+) and absent (S9−).

10982280, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/em.70003 by Yusuf Hussien - <Shibboleth>-member@swan.ac.uk , Wiley Online Library on [14/03/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

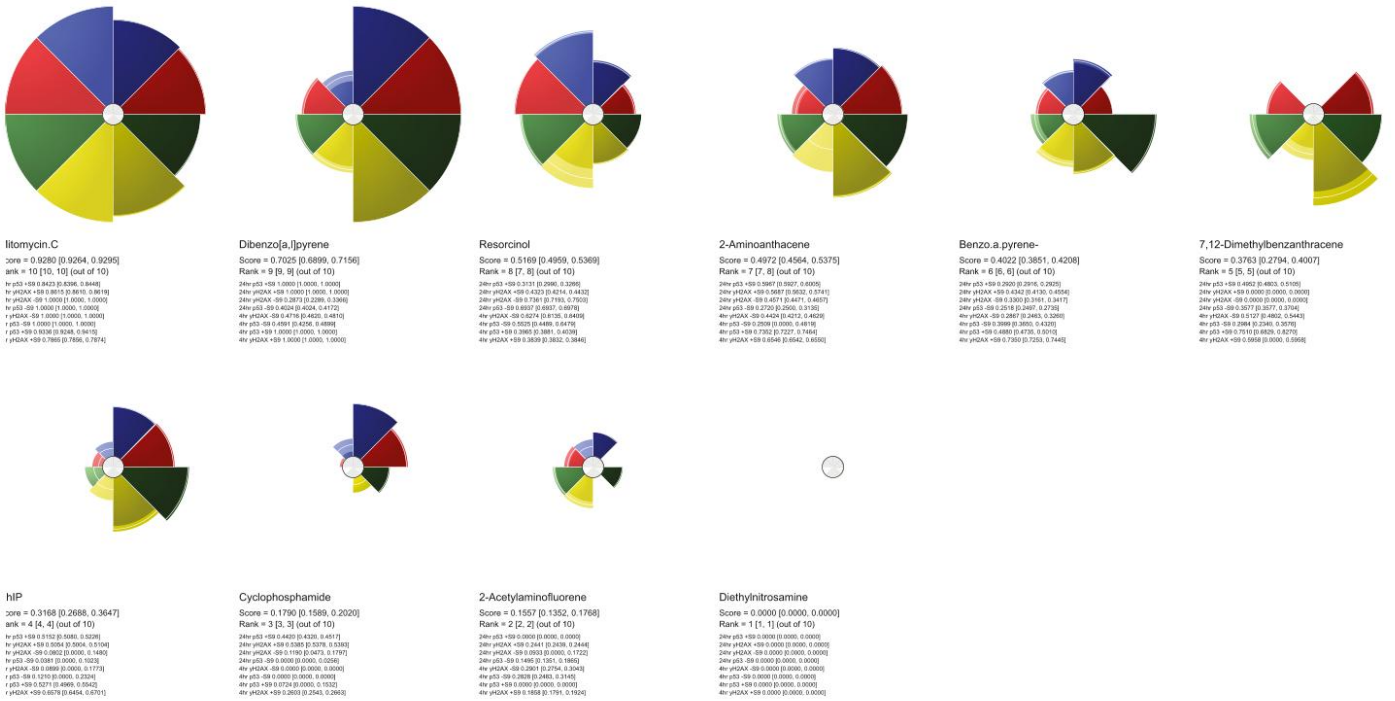


FIGURE 4 | ToxPi profiles (with bootstrap CI) generated from BMD data for each chemical, showing the biomarkers and their ToxPi scoring (alongside bracketed CIs) for each compound across all endpoints. Each slice represents a biomarker at a given time in the presence/absence of S9. Shading in the center of the circle represents uncertainty for each slice (darker = more uncertain). The radius of each slice (from the center of each circle to either the edge or white line toward the edge of the slice) represents the response of each biomarker and the widths of the user-defined weightings. The brighter regions presented in each are the bootstrap CI where the outer edge shows the upper CI and the inner edge is the lower CI for each slice.

10982280, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/em.70003 by Yusuf Hussein - <Schibobeth>-member@swan.ac.uk , Wiley Online Library on [14/03/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

between 0–1 by subtracting the minimum value and dividing by the range. The normalized scores are then multiplied by the user-selected fold-weight (or by 1 ÷ the number of slices if weights are equal), with the resulting component scores summed and ranged between 0–1, producing a final ToxPi score. This score indicates that compounds with values closer to 1 are more potent at inducing toxicity while those with values closer to 0 are less potent (Auerbach et al. 2016). These figures combine overall responses across all figures into an automatically generated visualized pie chart with 90% CI (through 1000 automated bootstrap runs) and features that allow for the toxicity figures for all compounds across all biomarkers/time points/metabolic conditions to be effectively compared, which are illustrated in the below.

3 | Results

See Figures 1–8.

4 | Discussion

Figure S1 presents BMD CI high-low plots that illustrate significant differences in responses among chemical compounds by assessing the presence or absence of overlapping CIs between biomarkers. These figures were adapted from (Wheeldon et al. 2021) and utilized the BMDL and BMDU values to construct the high-low plots. While the figure demonstrates some ability to rank potency and allows for comparison of compounds across different endpoints, its utility is significantly constrained by its limitation to a single endpoint at a time. This means that this approach fails to integrate all biomarkers, time points, and metabolic states into a single combined analysis figure used for comparative analysis, limiting the ability to directly compare overall compound toxicity and assess individual endpoint responses effectively. Instead, multiple endpoint figures must be assessed simultaneously, complicating efficient dose–response and MoA analyses for chemical or drug safety evaluations. Additionally, the use of BMD CIs for biomarker response assessment is debated as a single BMD figure is not determined. Instead, 90% CIs are generated, indicating the concentration range where the true BMD exists. This approach is reflective of the estimative nature of BMD modeling, providing a more robust measure of the PoD. It also fails to highlight how some endpoints are considered higher in significance across regulatory bodies and applies weighting applications to those biomarkers.

These restrictions highlighted here hinder the comparative capacity of multiplexed assays. However, combining BMD CI analysis with ToxPi visualization offers a more sophisticated, efficient, and comprehensive representation of responses across multiple biomarkers, effectively addressing the shortcomings of traditional BMD CI figures and modeling plots. BMD CI and ToxPi analysis allow for datasets modification, enabling the separation of S9+ and S9− figures for each compound. This depicts specific compound potencies in both metabolic states within a single figure. This is demonstrated in Figure 1, where differences in biomarker activation highlighted by the ToxPi analysis reveal response variations between compounds across both metabolic states.

ToxPi also enables users to customize the selection and weighting of individual slices, providing flexibility in generating ToxPi figures and tailoring weightings to specific biomarkers across multiple unique assays. As a result, biomarker responses and compound toxicity metrics can be more precisely aligned with the relevance of each endpoint in the context of each assay. This enhances analytical nuance by incorporating endpoint weightings that are tailored to regulatory frameworks. However, it is important to note that ToxPi does not automatically apply weighting according to regulatory knowledge. Instead, it serves as a platform for users to apply their knowledge and expertise in determining the relative importance of endpoints in different assays and to determine which weightings figures should be applied to each biomarker. This facet enables results that hold relevant weightings across endpoints to be more precisely determined.

Key findings from the clustering analysis include the strong association between Mitomycin C and resorcinol, both of which demonstrated reduced response in the presence of S9, thereby clustering closely together. Meanwhile, compounds such as 7,12-dimethylbenz[a]anthracene (DMBA) and 2-aminoanthracene, which exhibited increased potency in the presence of S9, also formed a distinct cluster. The S9+/S9− ToxPi profile results in Figure 4 reaffirm these findings as both mitomycin C and resorcinol have ToxPi scores higher in the absence of S9− compared to S9+. 2-acetylaminofluorene and diethylnitrosamine, which showed no significant difference in biomarker activation with or without S9, clustered together as expected and ToxPi ranking also reflects this. Interestingly, both benzo[a]pyrene and cyclophosphamide showed similar, modest increases in activity with S9, consistent with ToxPi rankings and Euclidean hierarchical cluster figures. However, despite that ToxPi's hierarchical clustering algorithm (Figure 6) separated these compounds, their ToxPi profiles (Figure 2) suggest that they are more closely related than indicated by the Figure 5 clustering results, highlighting

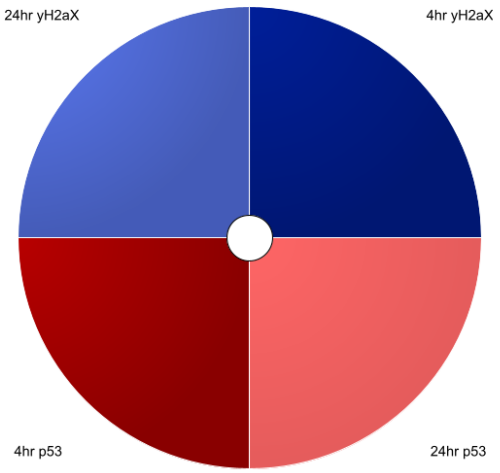


FIGURE 5 | ToxPi profile key illustrating each slice with its respective endpoints, including color-coordinated endpoints for γH2AX and p53 endpoints at 4/24-h time points.

10982280, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/em.70003 by Yusuf Hussien - <Shibboleth>-member@swan.ac.uk , Wiley Online Library on [14/03/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

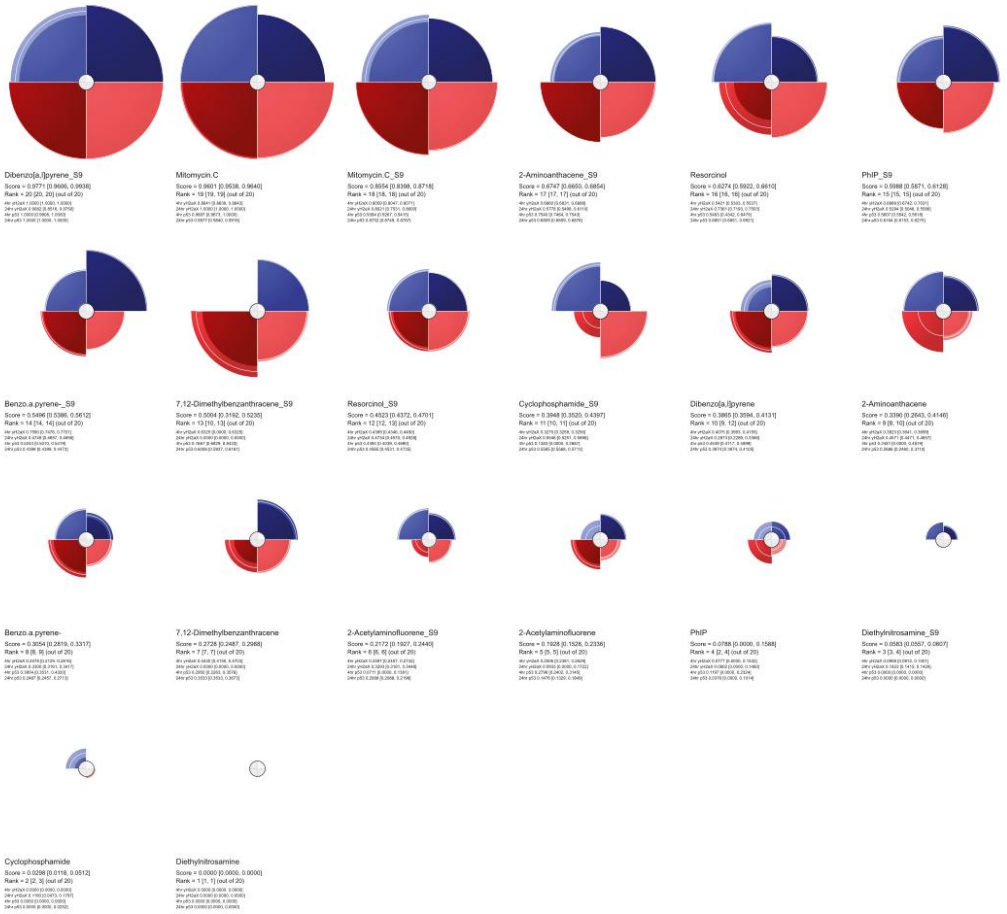


FIGURE 6 | ToxPi profiles (with bootstrap CI) generated from BMD data for each chemical, showing the biomarkers and their ToxPi scoring (alongside bracketed CIs) for each compound across all endpoints. Shading in the center of the circle represents uncertainty for each slice (darker = more uncertain). The radius of each slice (from the center of each circle to the either the edge or white line toward the edge of the slice) represents the response of each biomarker and the widths the user-defined weightings. The brighter regions presented in each is the bootstrap CI where the outer edge shows the upper CI, and the inner edge is the lower CI for each slice.

differences between ToxPi's divisive top-down approach and the AGNES bottom-up approach of the *hclust* function.

Similar to the compounds, PhIP and dibenzo[a,l]pyrene both show a significant increase in signal when metabolically activated, a trend reflected in both the hclust function and ToxPi ranking and clustering, despite what the ToxPi hierarchical clustering suggests, which left PhIP/dibenzo [a,l]pyrene and benzo[a]pyrene/cyclophosphamide in different grouped clusters. Additionally, it is worth noting that BMD CI limits were established in this analysis, and unlike the original study by (Wheeldon et al. 2021), a subtraction approach was utilized instead of division to assess S9 potency ratio CI differences. The absence of recorded RNC BMD CIs in this analysis, as opposed to the referenced study, likely further contributed to the variations observed in the clustering outcomes between the two studies. Further clustering analysis requires the integration of further programs to be performed effectively, and hence, direct raw figures can be compared using either R functions, JMP software's MDS platform, XLSTAT in excel, etc. The process further results in the transformation of the original data to fit the desired visualization and clustering purposes; hence, Gene ToxPi offers a robust alternative to these current approaches.

Despite a singular system being implemented to produce visualized ToxPi figures for the MultiFlow and MicroFlow assays, multiple assays are covered in the field of toxicology, pharmacy, pharmacology, and genotoxicology. This means unique approaches are required for the protocols described in this study when a different assay is considered for ToxPi visualization due

10982280, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/em.70003 by Yusuf Hussien - <Shibboleth>-exember@swan.ac.uk , Wiley Online Library on [14/03/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

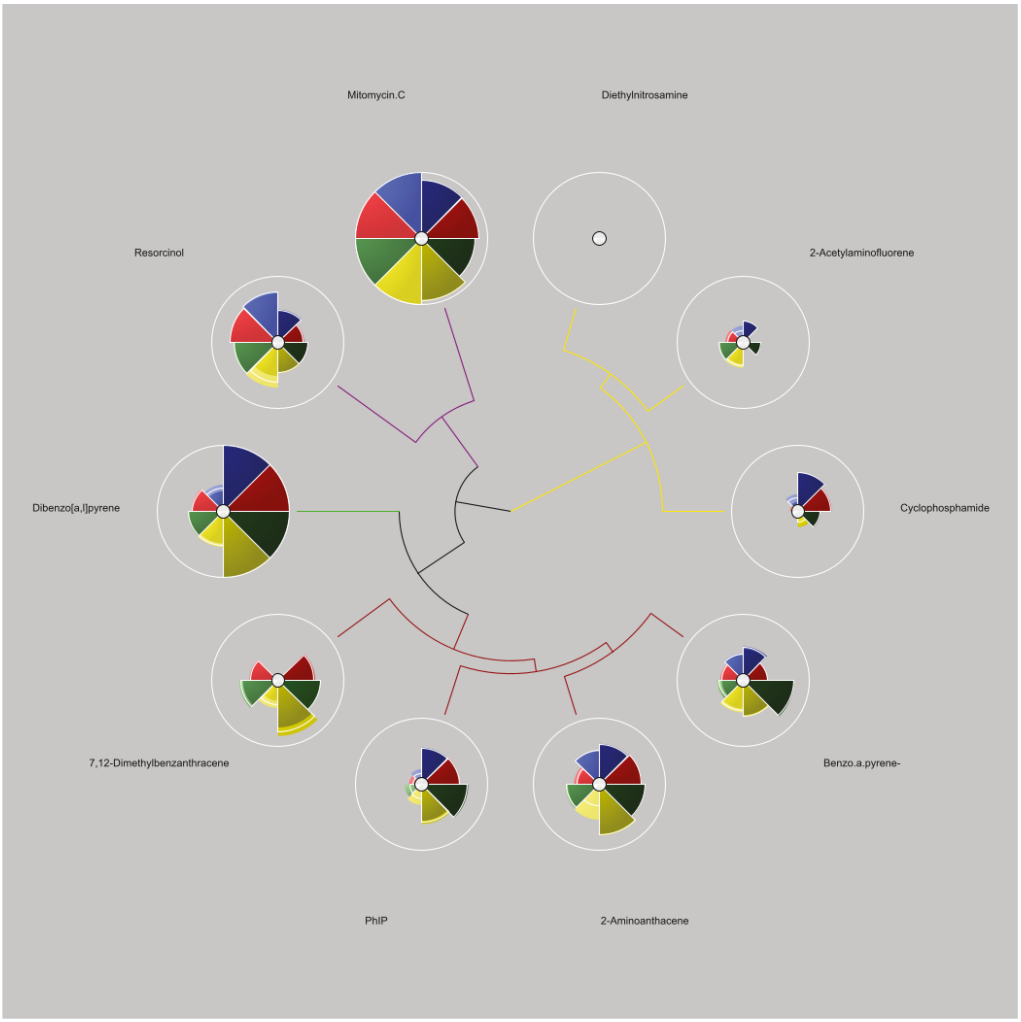


FIGURE 7 | ToxPi profiles hierarchical clustering generated through complete divisive clustering to show grouping of each chemical, which are color-coordinated. Green indicates compounds where genotoxicity was dramatically increased in the presence of S9 (average ratio of 2.6-fold in ToxPi across all endpoints), red for compounds where S9 increased genotoxicity (average ratio of 1.6-fold in ToxPi across all compounds/endpoints), yellow for no potency changes when S9 is introduced and purple, where S9 presence reduced genotoxic capacity of a compound; correlating with the S9 potency ratio CIs generated by (Wheeldon et al. 2021).

to the different meanings behind the numerical values produced in raw figures, whether any significant differences are present in numerical units, the study design behind the analysis, and the formatting/transformations that are applied.

Diethylnitrosamine is a potent genotoxicant, yet its ToxPi score of 0.000 might initially appear to suggest a lack of response across all biomarkers. However, this outcome highlights the importance of marker selection and coverage when designing multiplexed assays, rather than the limitations surrounding the ToxPi framework itself. In this case, the lower ToxPi score stems from the absence of relevant gene mutation endpoints for diethylnitrosamine detection, resulting in elevated BMD values across endpoints. This necessitated the adjustment of some biomarkers to 10,000 μM, as detailed in the methods. The specific nature of O6-alkylguanine DNA lesions, a hallmark of diethylnitrosamine's mutagenicity, likely necessitates specialized detection techniques, including probes, antibodies, or microbial assays,

10982280, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/em.70003 by Yusuf Hussien - <Shibboleth>-member@swan.ac.uk , Wiley Online Library on [14/03/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

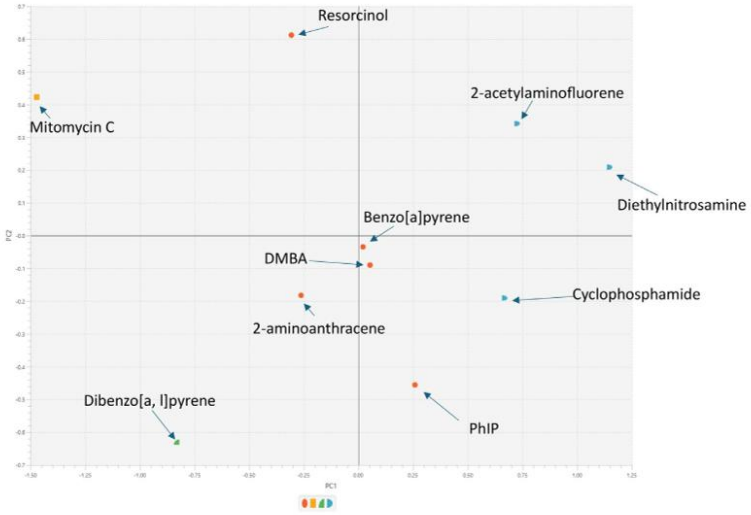


FIGURE 8 | ToxPi K-means AGNES/PCA clustering where each compound is grouped based on variation across the ToxPi scores/genotoxicity and biomarker sensitivity. Each point represents a compound that is automatically given unique colors/shapes based on a user-selected number of clusters (4 clusters in this case) and presented in the graph legend. Each color/shape represents each unique cluster that compounds have been placed into. The pattern located is that PC1 compounds with higher ToxPi scores are toward the left, and for PC2, compounds with increased response in the presence of S9+ are at the bottom. Furthermore, the compounds cluster similarly in K-means colors/shapes according to their metabolic properties as described in the hierarchical cluster chart, with the exception of mitomycin C and resorcinol: Both compounds that have reduced potency with S9 presence.

that fall outside the scope of the MultiFlow assay (Bartsch and Montesano 1984). While γ-H2AX is capable of detecting DNA damage, it does not specifically target O6-alkylguanine lesions, potentially contributing to the elevated BMD values observed. Consequently, the ToxPi score of 0.000 reflects diethylnitrosamine's comparatively lower potency within the chosen endpoints of this study, rather than a lack of genotoxicity. This suggests that there is a need for follow-up testing (such as the Ames test) and careful endpoint selection to ensure adequate coverage, particularly when designing assays intended to predict responses across a broad spectrum of genotoxicants. However, it is important to note that the aim of this analysis was not to evaluate the predictivity of individual markers but rather to derive insights from BMD estimates across multiple endpoints. Both these additions to the existing methodology could provide additional key information to develop a more robust compound biomarker response profile.

Additionally, the ToxPi results generated are subjective to each dataset, and therefore currently, ToxPi scores themselves cannot be compared across multiple individual ToxPi analyses (you cannot currently compare one set of ToxPi scores to a separate set). However, as groupings are a collective result of all BMD input values and ToxPi profiles, approaches to combining the clustering results may be compared, and extensive results from a collective effort would alleviate these concerns. Furthermore, multiple analyses may be combined into a single assessment for extensive compound analysis, which has already been demonstrated in (Dertinger et al. 2024) which analyzed 126 compounds across 7 biomarkers. While the MultiFlow assessment process is straightforward, analyzing additional compounds and endpoints can complicate clustering by modes of action, reflecting study design challenges rather than limitations of ToxPi.

Weighting percentages are configured such that increasing the percentage enhances the priority or importance assigned to each endpoint/slice, directing greater focus toward higher-weighted factors in the analysis. A standardized weighting percentage system has not been formally set; hence, it has remained unchanged across all endpoints as 1-fold in this analysis (leaving all weightings as equal), as justifications for numerical selection, biomarker biological importance, and regulatory constraints are required before a standardized system of weighting fold applications may be set. Extensive experimentation is currently underway to refine model weightings and significance by evaluating assay-specific differences in biomarker selection and response information. In the long term, this research seeks to establish a tailored, individualized approach that accounts for the unique contributions of each assay.

Although the data presented here focus on two endpoints (+/− S9), this analysis serves as an initial evaluation of the methodology, demonstrating its potential utility for more complex multiplexed datasets. The approach's adaptability to integrate multiple biomarkers and conditions makes it well-suited for scaling up to multiplexed studies, and while traditional binary hazard assays might achieve clustering of compounds in the presence/

10982280, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/em.70003 by Yusuf Hussien - <Shibboleth>-member@swan.ac.uk , Wiley Online Library on [14/03/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

absence of S9, such approaches lack the nuanced integration of biomarker-specific responses across time points and metabolic states, as observed here. This highlights the added value of combining BMD CI analysis with ToxPi visualization for more comprehensive insight. We recognize that this study is not completely comprehensive and does not represent the full complexity of multiplexed assays. However, the findings provide a proof of concept for applying these tools to datasets where binary outputs alone may overlook subtle patterns and interactions among endpoints. If this approach was applied to older assay data, clustering patterns similar to those observed here may emerge. However, the integration of metabolic state-specific responses and the flexibility of endpoint weighting offered by ToxPi significantly enhance the interpretive power, particularly for comparative potency assessments.

5 | Conclusion

In summary, multiplex assays are incorporating new analysis systems and machine learning algorithms to enhance biological insights and optimize data visualization for maximum information extraction. This methodology adeptly refines compound analysis by leveraging BMD CI to provide robust potency assessments, utilizing ToxPi weightings to capture underlying MoA, and incorporating cluster analysis alongside toxicity ranking to enable nuanced and detailed cross-comparisons between diverse multiplexed datasets. Although further research is necessary to establish a standardized system for applying these methods across multiple assays, the methodology presented here demonstrates a clear and robust ability to compare the analysis of compounds and their potencies across multiple endpoints, with or without metabolic activation. Additionally, the weighting percentages for each biomarker remain the same for these analyses. Ongoing multiplex assay analyses using GeneToxPi with BMD on genotoxicity data aim to address these issues by developing a comprehensive weighting system for multiple endpoints across diverse genotoxic assays.

To establish these weightings, each individual genotoxic assay would require an increase in the number of reference compounds with known modes of action being used per analysis, and more biomarkers would require curation/processing to develop and apply a weight-of-evidence approach to weighing biomarkers. Health Canada performed this across multiple assays in one assessment; however, this focused on weightings of entire assays as opposed to individual biomarkers (Fortin et al. 2023). As alluded to earlier, more data from an extensive collective effort will help develop a biological and regulatory rationale for the weightings applied, especially when considering whether to prioritize in vivo vs. in vitro, atypical vs. early repairable indicator-type responses, immediate vs. 4-h vs. 24-h responses, etc. Interpreting BMD CIs is a crucial aspect of modeling dose–response data across various assays.

The integration of BMD analysis with ToxPi visualization represents a pivotal advancement in the refinement of assay evaluation and compound hazard assessment. Each assay implemented requires a well-designed protocol subjected to rigorous evaluation from statistical, mathematical, biological, and genetic perspectives while aligning with regulatory frameworks. These efforts enable the development of optimized analytical pipelines that leverage ToxPi's visualization capabilities to advance comparative techniques for numerous compounds in future research. Integrating BMD analysis with ToxPi enhances data interpretation, minimizes information loss, and addresses the limitations of current compound hazard assessment methodologies, paving the way for more robust and predictive approaches.

Acknowledgments

Yusuf Hussien is a PhD student at Swansea University and has received partial tuition funding datasets through the fast fund at the Health and Environmental Science Institute Genetic Toxicology Technical Committee. Benchtop work was performed by Litron Laboratories, and raw results were provided by both Dr. Ryan P. Wheeldon and Dr. George Johnson as part of the ongoing PhD program. The contents are solely the responsibility of the authors and do not necessarily represent the official views of NIEHS.

Conflicts of Interest

All original benchtop experiments were performed at Litron Laboratories, which is a pharmaceutical firm that has patents covering the MultiFlow flow-cytometric-based assay assessed in this experiment and is commercially sold to customers as the MultiFlow DNA Damage Kit—p53, γH2AX, Phospho-Histone H3.

References

Arif, J. M., and R. C. Gupta. 1997. "Microsome-Mediated Bioactivation of Dibenzo[a,l]Pyrene and Identification of DNA Adducts by 32P-Postlabeling." *Carcinogenesis* 18: 1999–2007.

Auerbach, S., D. Filer, D. Reif, et al. 2016. "Prioritizing Environmental Chemicals for Obesity and Diabetes Outcomes Research: A Screening Approach Using ToxCast™ High-Throughput Data." *Environmental Health Perspectives* 124, no. 8: 1141–1154. https://doi.org/10.1289/ehp.1510456.

Bartsch, H., and R. Montesano. 1984. "Relevance of Nitrosamines to Human Cancer." *Carcinogenesis* 5, no. 11: 1381–1393. https://doi.org/10.1093/carcin/5.11.1381.

Bryce, S. M., D. T. Bernacki, J. C. Bemis, et al. 2017. "Interlaboratory Evaluation of a Multiplexed High Information Content In Vitro Genotoxicity Assay." *Environmental and Molecular Mutagenesis* 58: 146–161.

Bryce, S., D. Bernacki, J. Bemis, and S. Dertinger. 2016. "Genotoxic Mode of Action Predictions From a Multiplexed Flow Cytometric Assay and a Machine Learning Approach." *Environmental and Molecular Mutagenesis* 57, no. 3: 171–189.

Carriére, V., I. de Waziers, Y. A. Courtois, J. P. Leroux, and P. H. Beaune. 1992. "Cytochrome P450 Induction and Mutagenicity of 2-Aminoanthracene (2AA) in Rat Liver and Gut." *Mutation Research* 268: 11–20.

Chatterjee, N., and G. Walker. 2017. "Mechanisms of DNA Damage, Repair, and Mutagenesis." *Environmental and Molecular Mutagenesis* 58, no. 5: 235–263. https://doi.org/10.1002/em.22087.

Davis, J. A., J. S. Gift, and Q. J. Zhao. 2011. "Introduction to Benchmark Dose Methods and U.S. EPA's Benchmark Dose Software (BMDS) Version 2.1.1." *Toxicology and Applied Pharmacology* 254: 181–191.

10982280, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/em.70003 by Yusuf Hussien - <Shibboleth>-member@swan.ac.uk , Wiley Online Library on [14/03/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Dertinger, S. D., E. Briggs, Y. Hussien, et al. 2024. "Visualization Strategies to Aid Interpretation of High-Dimensional Genotoxicity Data." *Environmental and Molecular Mutagenesis* 65, no. 5: 156–178. https://doi.org/10.1002/em.22604.

Dorato, M., and J. Engelhardt. 2005. "The No-Observed-Adverse-Effect-Level in Drug Safety Evaluations: Use, Issues, and Definition(s)." *Regulatory Toxicology and Pharmacology* 42, no. 3: 265–274.

European Food Safety Authority. 2010. "Scientific Opinion on the Use of Resorcinol as a Food Additive." *EFSA Journal* 8, no. 1: 1411.

Fortin, A.-M. V., A. S. Long, A. Williams, et al. 2023. "Application of a New Approach Methodology (NAM)-Based Strategy for Genotoxicity Assessment of Data-Poor Compounds." *Frontiers in Toxicology* 5: 1098432. https://doi.org/10.3389/ftox.2023.1098432.

Hakem, R. 2008. "DNA Damage Repair: The Good, the Bad, and the Ugly." *EMBO Journal* 27, no. 4: 589–605.

Heflich, R. H., G. E. Johnson, A. Zeller, et al. 2019. "Mutation as a Toxicological Endpoint for Regulatory Decision-Making." *Environmental and Molecular Mutagenesis* 61, no. 1: 34–41. https://doi.org/10.1002/em.22338.

Helma, C., M. Uhl, R. Schulte-Hermann, and S. Knasmüller. 2001. "Clastogenic Effects of 2-Aminoanthracene in the Single-Cell Gel Electrophoresis (Comet) Assay in Mammalian Cells." *Mutagenesis* 16, no. 6: 461–466.

Higa, M., M. Nagata, K. Suzuki, H. Ito, and S. Takeda. 1998. "Mitomycin C-Induced DNA Damage and Its Repair." *Journal of Biochemistry* 123, no. 3: 491–497.

International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH S2(R1)). 2012. "ICH Harmonised Guideline S2(R1): Guidance on Genotoxicity Testing and Data Interpretation for Pharmaceuticals Intended for Human Use." https://www.ich.org/page/safety-guidelines.

Khalil, S. R., H. S. Gharib, H. H. A. El-Rahman, M. H. Hassan, and S. F. Mahmoud. 2018. "Mutagenic and Carcinogenic Potential of Cyclophosphamide: Molecular and Cellular Mechanisms." *Environmental Toxicology and Pharmacology* 63: 64–70.

Kim, H., Y. Park, D. Shin, S. Oh, and K. Chung. 2013. "Appropriate In Vitro Methods for Genotoxicity Testing of Silver Nanoparticles." *Environmental Health and Toxicology* 28: e2013003. https://doi.org/10.5620/eht.2013.28.e2013003.

Kirkland, D., P. Kasper, H.-J. Martus, et al. 2016. "Updated Recommended Lists of Genotoxic and Non-Genotoxic Chemicals for Assessment of the Performance of New or Improved Genotoxicity Tests." *Mutation Research* 795: 7–30.

Krais, A. M., E. N. Speksnijder, J. P. M. Melis, et al. 2016. "Metabolic Activation of 2-Amino-1-Methyl-6-Phenylimidazo[4,5-b]Pyridine and DNA Adduct Formation Depends on p53: Studies in Trp53(+/+), Trp53(+/−), and Trp53(−/−) Mice." *Molecular Cancer Biology* 138: 976–982. https://doi.org/10.1002/ijc.29836.

Kuklenyik, Z., D. L. Ashley, C. H. Watson, and D. B. Barr. 2004. "Mutagenicity of 7,12-Dimethylbenz[a]Anthracene and Its Role in Cancer." *Toxicology Letters* 153, no. 2: 213–220.

MacGregor, J. T., R. Frötschl, P. A. White, et al. 2015. "IWGT Report on Quantitative Approaches to Genotoxicity Risk Assessment I. Methods and Metrics for Defining Exposure–Response Relationships and Points of Departure (Pods)." *Mutation Research, Genetic Toxicology and Environmental Mutagenesis* 783: 55–65. https://doi.org/10.1016/j.mrgentox.2014.09.011.

Marvel, S., K. To, F. Grimm, F. Wright, I. Rusyn, and D. Reif. 2018. "ToxPi Graphical User Interface 2.0: Dynamic Exploration, Visualization, and Sharing of Integrated Data Models." *BMC Bioinformatics* 19, no. 1: 80.

Organisation for Economic Co-operation and Development (OECD). 2016a. *Test No. 476: In Vitro Mammalian Cell Gene Mutation Tests Using the Hprt and Xprt Genes (OECD Guidelines for the Testing of Chemicals, Section 476)*. OECD Publishing. https://doi.org/10.1787/9789264264809-en.

Organisation for Economic Co-operation and Development (OECD). 2016b. *Test No 487: In Vitro Mammalian Cell Micronucleus Test (OECD Guidelines for the Testing of Chemicals, Section 487)*. OECD Publishing. https://doi.org/10.1787/9789264264861-en.

Otteneder, M., and W. K. Lutz. 1999. "Correlation of DNA Adduct Level With Tumor Incidence: Carcinogenic Potency of DNA Adducts." *Mutation Research* 424: 237–247.

Ramesh, A., S. A. Walker, D. B. Hood, M. D. Guillen, K. Schneider, and Y. Li. 2007. "Benzo[a]Pyrene-Induced DNA Damage: A Review." *Environmental Toxicology and Pharmacology* 24, no. 1: 1–16.

Reif, D., M. Sypa, E. Lock, et al. 2012. "ToxPi GUI: An Interactive Visualization Tool for Transparent Integration of Data From Diverse Sources of Evidence." *Bioinformatics* 29, no. 3: 402–403. https://doi.org/10.1093/bioinformatics/bts686.

RIVM. 2021. *PROAST*. Dutch National Institute for Public Health and the Environment. https://www.rivm.nl/en/proast.

Rodriguez-Antona, C., and M. Ingelman-Sundberg. 2006. "Cytochrome P450 Pharmacogenetics and Cancer." *Oncogene* 25: 1679–1691.

Slob, W. 2002. "Dose-Response Modeling of Continuous Endpoints." *Toxicological Sciences* 66, no. 2: 298–312. https://doi.org/10.1093/toxsci/66.2.298.

Slob, W., and R. W. Setzer. 2014. "Shape and Steepness of Toxicological Dose–Response Relationships of Continuous Endpoints." *Critical Reviews in Toxicology* 44: 270–297. https://doi.org/10.3109/10408444.2013.853726.

Song, P., L. Wu, and W. Guan. 2015. "Dietary Nitrates, Nitrites, and Nitrosamines Intake and the Risk of Gastric Cancer: A Meta-Analysis." *Nutrients* 7, no. 12: 9872–9895.

Sturla, S. 2018. "Point of Departure." *Chemical Research in Toxicology* 31, no. 1: 2–3.

Sugimura, T., K. Wakabayashi, M. Nagao, and H. Esumi. 2011. "Carcinogenic and Mutagenic Properties of Heterocyclic Amines and 2-Acetylaminofluorene: A Review." *Journal of Cancer Research and Clinical Oncology* 137, no. 6: 849–857.

Tian, S., A. Cyr, K. Zeise, et al. 2020. "3Rs-Friendly Approach to Exogenous Metabolic Activation That Supports High-Throughput Genetic Toxicology Testing." *Environmental and Molecular Mutagenesis* 61: 408–432.

U.S. Department of Health and Human Services Food and Drug Administration FDA. 2002. *Guidance for Industry and Reviewers (Draft). Estimating the Safe Starting Dose in Clinical Trials for Therapeutics in Adult Healthy Volunteers*. Draft Guidance.

Wheeldon, R. P., S. D. Dertinger, S. M. Bryce, J. C. Bemis, and G. E. Johnson. 2021. "The Use of Benchmark Dose Uncertainty Measurements for Robust Comparative Potency Analyses." *Environmental and Molecular Mutagenesis* 62, no. 3: 203–215. https://doi.org/10.1002/em.22422.

Wills, J., G. Johnson, H. Battaion, W. Slob, and P. White. 2017. "Comparing BMD-Derived Genotoxic Potency Estimations Across Variants of the Transgenic Rodent Gene Mutation Assay." *Environmental and Molecular Mutagenesis* 58, no. 9: 632–643. https://doi.org/10.1002/em.22137.

Wilson, J., and T. Hunt. 2002. *Molecular Biology of the Cell*. 4th ed. Garland Science.

10982280, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/em.70003 by Yusuf Hussien - Shibboleth - ymember@swan.ac.uk , Wiley Online Library on [14/03/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Yamazaki, H., Y. Oda, Y. Funae, et al. 1992. "Participation of Rat Liver Cytochrome P450 2E1 in the Activation of n-Nitrosodimethylamine and n-Nitrosodiethylamine to Products Genotoxic in an Acetyltransferase-Overexpressing *Salmonella typhimurium* Strain (NM2009)." *Carcinogenesis* 13: 979–985.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.

10982280, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/em.70003 by Yusuf Hussien - <Shibboleth>-member@swan.ac.uk , Wiley Online Library on [14/03/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Appendix 3. Genetic Toxicology Association (GTA) 2024 Annual Meeting Abstract Award

20

Yusuf Hussien
Swansea University, Swansea, United Kingdom

STUDENT ABSTRACT TRAVEL AWARD WINNER

Yusuf Hussien, a BSc Medical Genetics graduate, is currently a PhD student attending Swansea University working on developing new methodologies to identify potential hazards to our genomes and utilising visualisation techniques/programmes to maximise information extracted from multiple endpoint/mechanistic assays. This information is used to assist in regulatory decision-making for genotoxic risk assessment studies related to safeguarding human health from newer drug developments by minimising potential compound loss from information-dense raw data that pharmaceutical companies process. Yusuf is collaborating with various members of the Health and Environmental Science Institute Genetic Toxicology Technical Committee (HESI GTTC) members on multiplex datasets analysis. These include Litron Laboratories US, and Procter & Gamble (P&G) who provided the raw data, alongside Health Canada, the University of Ottawa and Toxys whose involvement refined the strength of these projects as their consultations were critical in the refinement of these genotoxic techniques and Yusuf's ability to develop the scientific skills necessary to further involve himself in pioneering risk assessment studies and pharmaceutical drug analysis. Yusuf is also a 100m/200m track and field athlete who has competed for Swansea University Athletics Club against various universities in the British Universities Championships, being part of the team that broke the indoor university record for the 4x200m relay.

An investigation on the mutagenicity of compounds using Benchmark Dose and Dose-Response data for research in genetic toxicology - Workshop I

Limitations in quantitative analytical techniques for genotoxicity studies restrict researchers from numerically analysing, and explaining how various environmental, chemical, and physical sources inflict cellular genetic damage to human tissue. Benchmark dose (BMD) analysis serves as an important metric when gathering dose-response data from genotoxic compounds and multiple studies analyse BMD data to appropriately determine doses that result in genetic abnormalities for potency visualisation, interpretation, and review purposes. Effectively mapping chemicals according to their potency allows for efficient comparative analysis of genotoxic compounds to be expressed, normalised, and formatted coherently, retaining depth and complexity by minimising information loss during the dataset processing phase. This investigation highlights the importance of multiple endpoint/mechanistic assays as a genotoxic tool in chemical risk assessment, allowing us to quantitatively measure the genotoxic effects of various mutagenic compounds on cellular biomarkers and their respective metabolites. PROAST and BMDExpress is used to calculate BMD confidence intervals (CIs) and provide information on potency, and ToxPi will be used to analyse the BMD CI across each endpoint and chemical, using approaches such as potency scoring, hierarchical clustering, and principal component analysis (PCA). This combined analysis, in conjunction with a weight-of-evidence approach, will inform us on the chemical grouping and potency across the multiple genotoxicity endpoints and allows for fast and precise assessments of complex potency datasets, ultimately placing increased importance on the usability of *in vitro* datasets for hazard and risk assessment methods for human safety.

Appendix 4. Genetic Toxicology Association (GTA) 2024 Annual Meeting Presentation

An investigation on the mutagenicity of compounds using Benchmark Dose and Dose-Response data for research in genetic toxicology

Yusuf Hussien BSc (Hons)

Taith

uOttawa

Swansea University
Prifysgol Abertawe

HESI Genetic Toxicology Technical Committee

Health Canada

P&G

LITRON LABORATORIES

EMGS Environmental Mutagenesis and Genomics Society
Synergistic Interactions for a Better World

ToxPi

ToxPi (v2.3) is an interactive visualisation programme that analyses the genotoxic potency of chemicals. Translating BMD data into a single ToxPi score, ranking the mutagenicity on a continuous spectrum instead of fixed groups (Reif et al., 2012).

ToxPi features K-means and Hiearchical clustering uses PCA techniques for dataset dimensionality reduction and AGNES algorithms to group chemicals based on similarities in ToxPi scoring and potency across endpoints (Marvel et al., 2018).

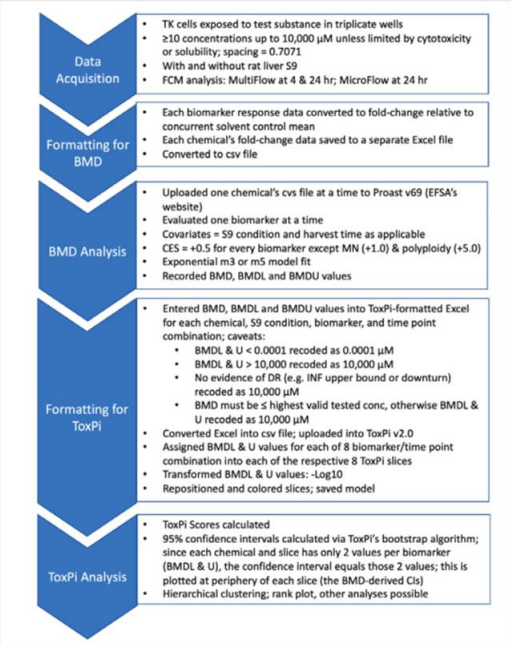
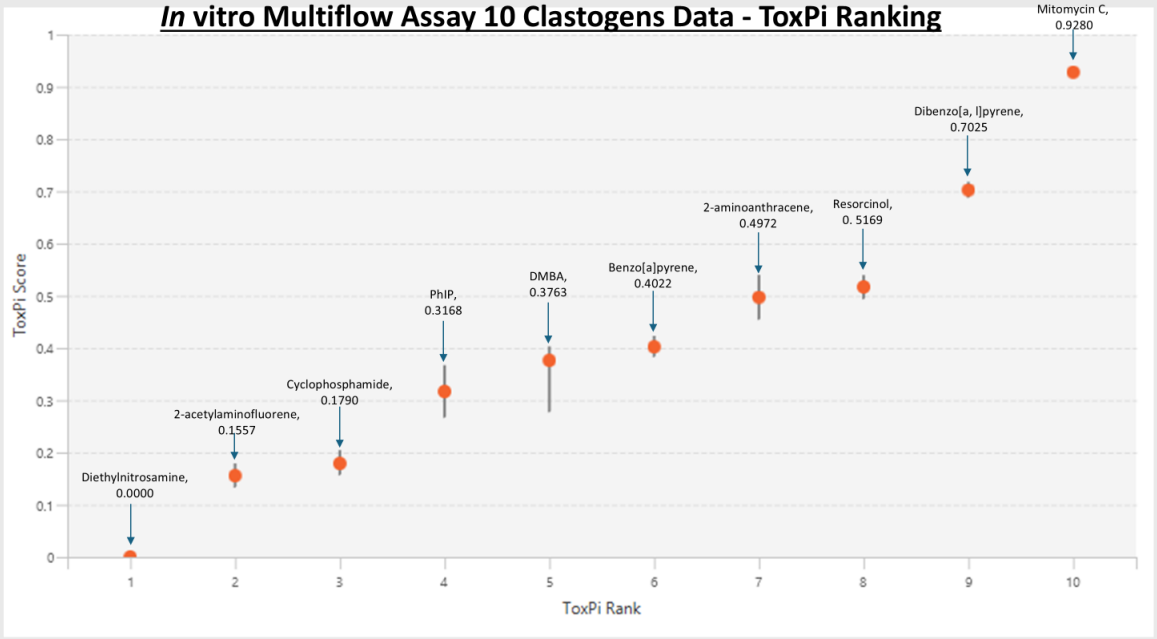


Fig 1: An adaptation of the all procedures by (Avlasenvich Et al., 2021).

Live Demonstration

Row	Model	A_Priori_MoA	Chemical	MN_BMDL	MN_BMDU	yH2AX_BMDL	yH2AX_BMDU	p53_BMDL	p53_BMDU	p-H3_BMDL	p-H3_BMDU
1	3+5	Clastogenic	Methyl_methanesulfonate	7.93	15.2	10.8	19.6	2	38.6	10000	10000
2	3+5	Clastogenic	Genistein	6.08	12	64.5	96.2	38	47.6	10000	10000
3	3+5	Aneugenic	Vinblastine	0.000102	0.000133	10000	10000	0.000166	0.000196	0.000283	0.000595
4	3+5	Aneugenic	Colchicine	0.00944	0.0106	10000	10000	0.12	0.0134	0.00746	0.00913
5	3+5	Non-Geno	D-Mannitol	10000	10000	10000	10000	10000	10000	10000	10000
6	3+5	Non-Geno	Cycloheximide	10000	10000	10000	10000	10000	10000	10000	10000



Abbreviations:
DMBA = 7,12-Dimethylbenzanthracene
PhIP = 2-Amino-1-methyl-6-phenylimidazo[4,5-b]pyridine

Fig 2: Results for Objective 2, showing ToxPi ranking for BMD data on 10 clastogens (1 = lowest potency and 10 = highest potency). ToxPi scores ranging between 0-1 where the higher score translates to higher genotoxic potency and the error bars represent the CI of each compound.

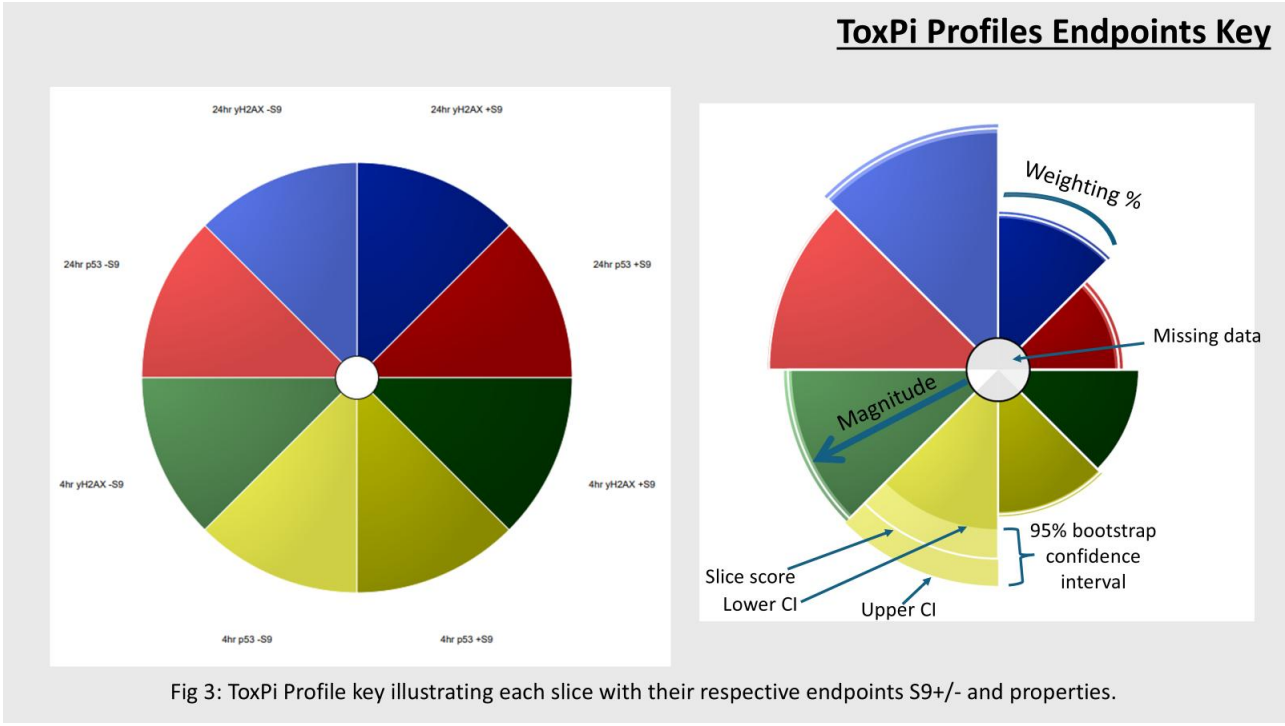


Fig 3: ToxPi Profile key illustrating each slice with their respective endpoints S9+/- and properties.

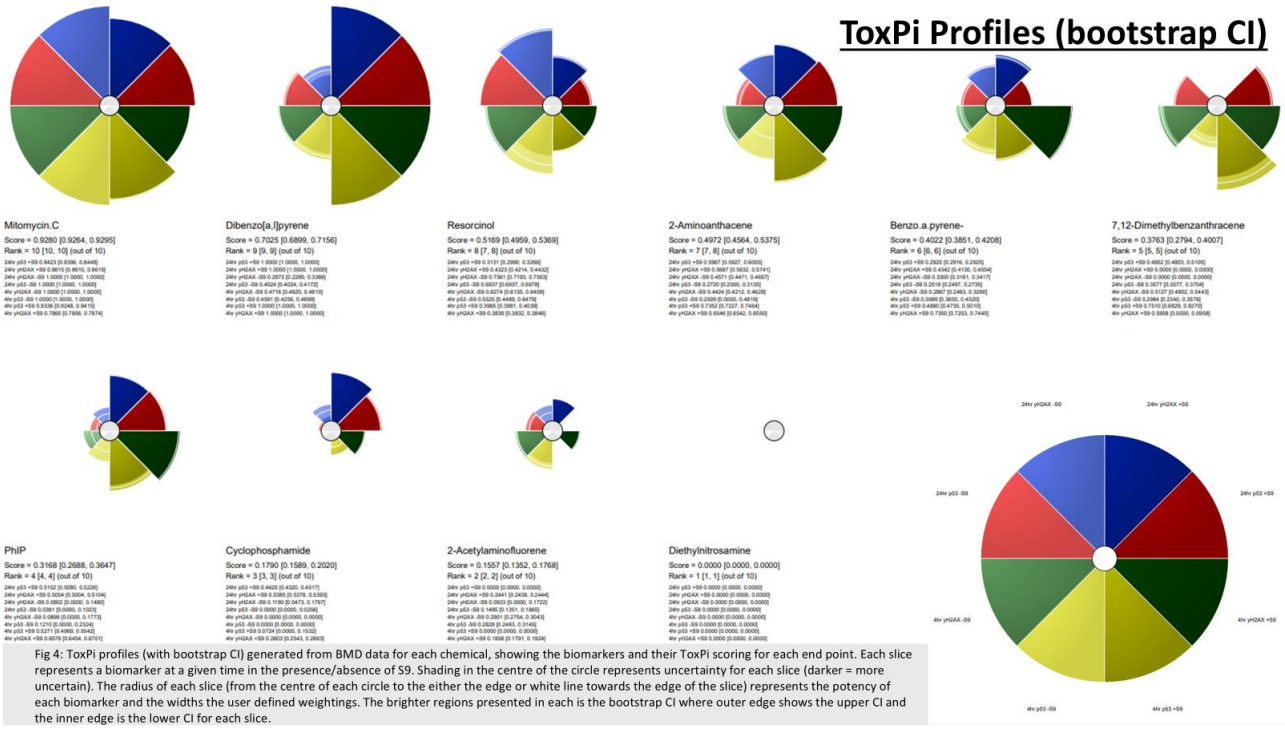
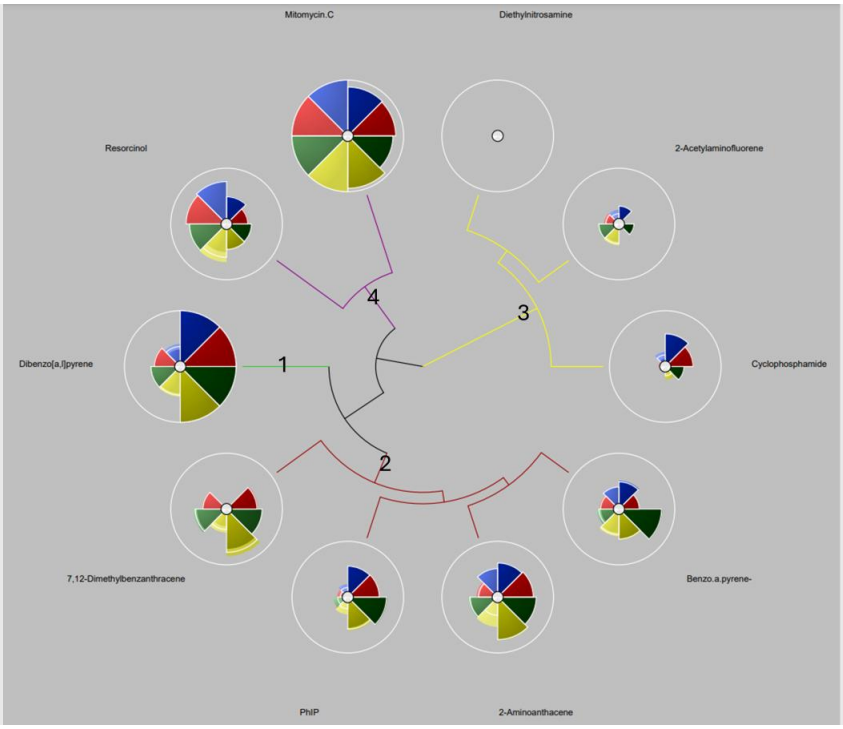


Fig 4: ToxPi profiles (with bootstrap CI) generated from BMD data for each chemical, showing the biomarkers and their ToxPi scoring for each end point. Each slice represents a biomarker at a given time in the presence/absence of S9. Shading in the centre of the circle represents uncertainty for each slice (darker = more uncertain). The radius of each slice (from the centre of each circle to the either the edge or white line towards the edge of the slice) represents the potency of each biomarker and the widths the user defined weightings. The brighter regions presented in each is the bootstrap CI where outer edge shows the upper CI and the inner edge is the lower CI for each slice.



Hierarchical Clustering (complete)

#clusters = 4

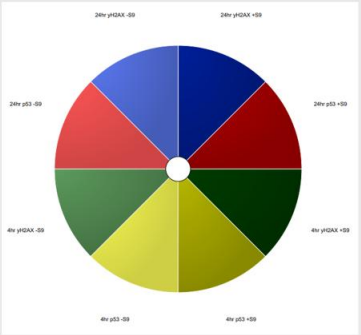
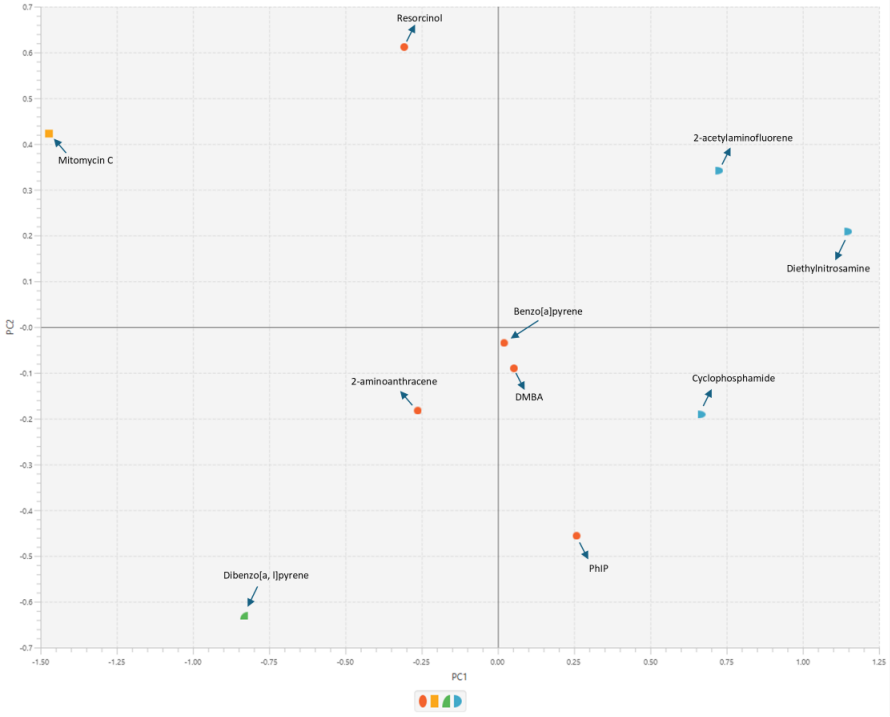


Fig 5: ToxPi profiles hierarchical clustering generated through complete AGNES clustering to show grouping of each chemical.



K-Means Clustering

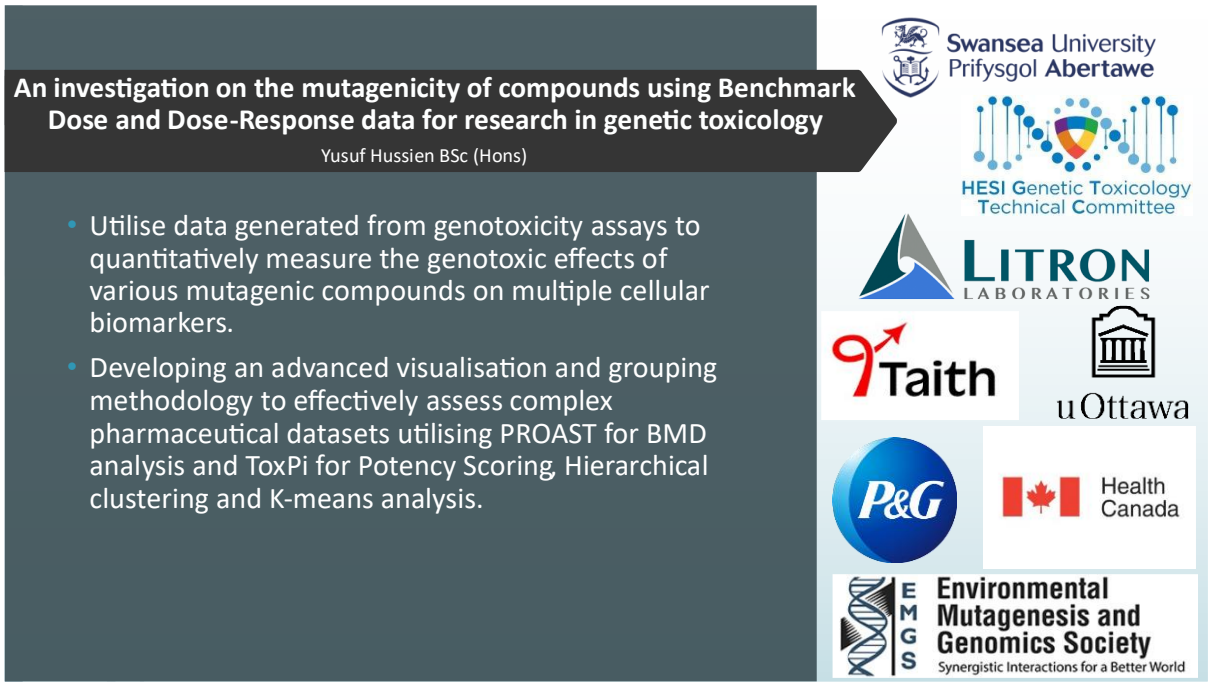
Fig 6: ToxPi K means PCA clustering where each compound is grouped based on variation across the ToxPi scores/genotoxicity and biomarker sensitivity. Each point represents a compound, and each colour/shape represents a cluster they have been placed into.
For PC1 compounds with higher ToxPi scores are towards the left, and for PC2, compounds with increased potency with S9+ are at the bottom.

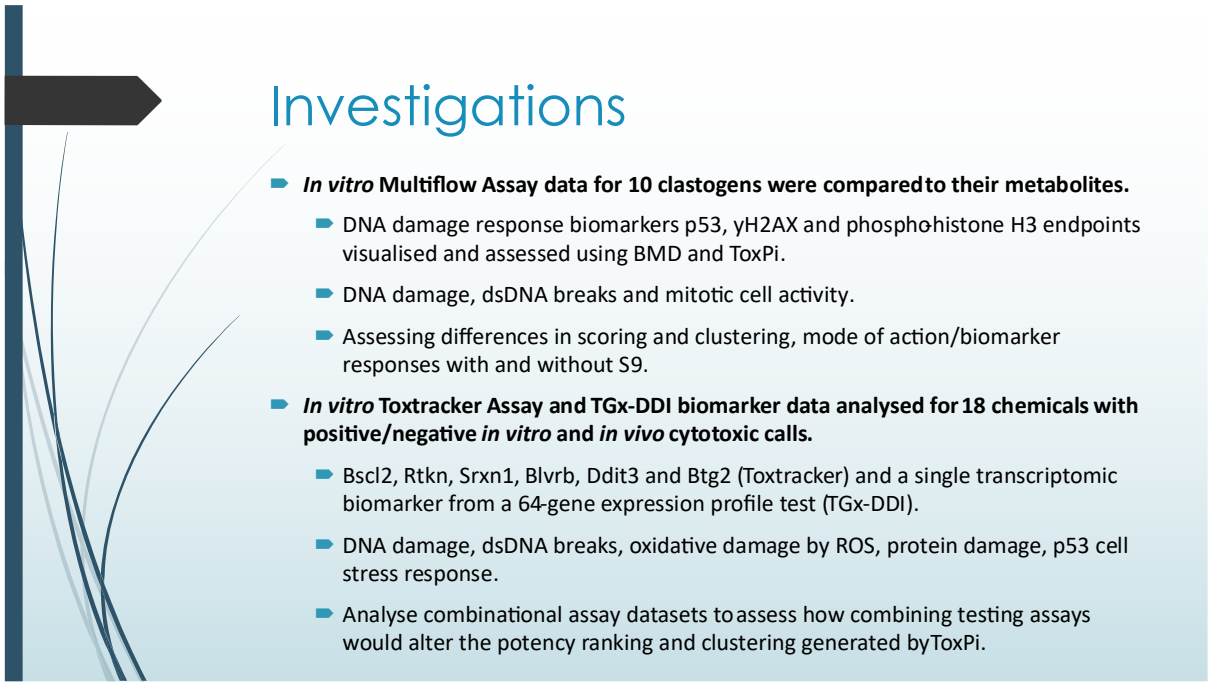
#clusters = 4

Conclusions

- Genotoxic effects of S9+/- on each compound is effectively illustrated in ToxPi profiles, drawing comparisons between parental compounds/metabolites.
- Hierarchical clustering can compartmentalise the clastogens into groups according to similarities in Toxpi scores, biomarker sensitivity and S9 activation properties.
- Clearly highlights the ToxPi profiles displaying the most genotoxic biomarkers at each time point for each clastogens and their BMD CI
- Effectively reduced the number of figures required compared to the main study, easing comparative analysis capabilities, and providing increased genotoxic endpoint information for each compound.

Appendix 5. Health and Environmental Science Institute Genetic Toxicology Technical Committee (HESI GTTC) 2024 Annual Meeting Presentation





ToxPi

- ToxPi (v2.3) is an interactive visualisation programme that analyses the genotoxic potency of chemicals. Translating BMD data into a single ToxPi score, ranking the mutagenicity on a continuous spectrum instead of fixed groups (Reif et al., 2012).
- ToxPi features K-means and Hiearchical clustering uses PCA techniques for dataset dimensionality reduction and AGNES algorithms to group chemicals based on similarities in ToxPi scoring and potency across endpoints (Marvel et al., 2018).

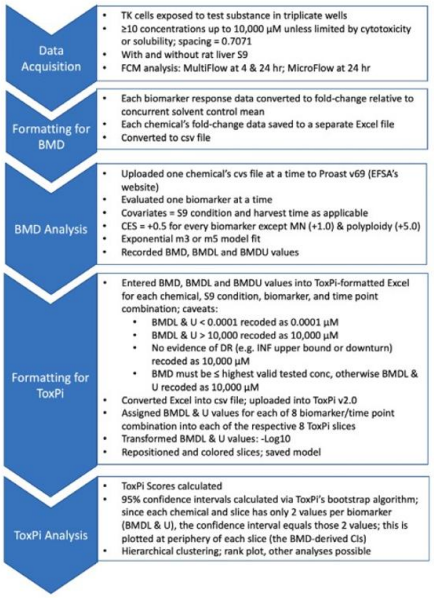


Fig 1: An adaptation of the all procedures by (Avlasenvich Et al., 2021).

ToxPi

- ToxPi (v2.3) is an interactive visualisation programme that analyses the genotoxic potency of chemicals. Translating BMD data into a single ToxPi score, ranking the mutagenicity on a continuous spectrum instead of fixed groups (Reif et al., 2012).
- ToxPi features K-means and Hiearchical clustering uses PCA techniques for dataset dimensionality reduction and AGNES algorithms to group chemicals based on similarities in ToxPi scoring and potency across endpoints (Marvel et al., 2018).

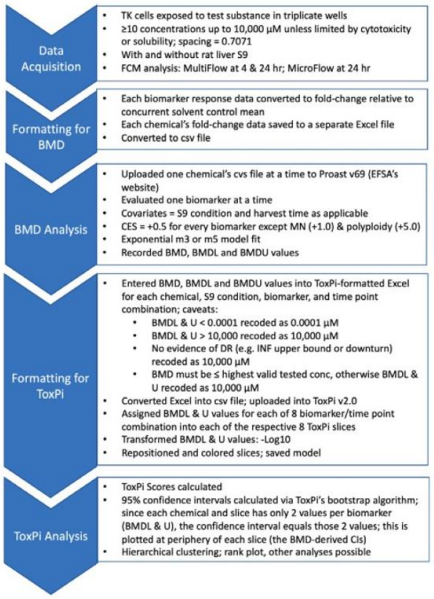
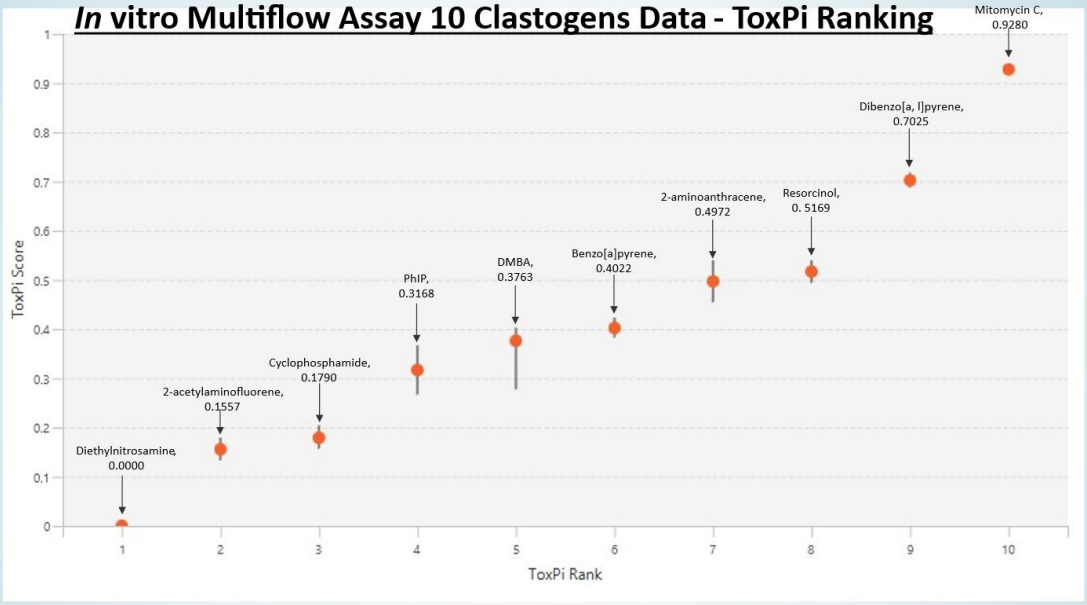


Fig 1: An adaptation of the all procedures by (Avlasenvich Et al., 2021).

The use of benchmark dose uncertainty measurements for robust comparative potency analyses

Ryan P. Wheeldon[1] | Stephen D. Dertinger[2] | Steven M. Bryce[2] | Jeffrey C. Bemis[2] | George E. Johnson[1]

[1]Institute of Life Science, Swansea University Medical School, Swansea University, Swansea, UK
[2]Litron Laboratories, Rochester, New York



Abbreviations:
DMBA = 7,12-Dimethylbenzanthracene
PhIP = 2-Amino-1-methyl-6-phenylimidazo[4,5-b]pyridine

Fig 2: Results for Objective 2, showing ToxPi ranking for BMD data on 10 clastogens (1 = lowest potency and 10 = highest potency). ToxPi scores ranging between 0-1 where the higher score translates to higher genotoxic potency and the error bars represent the CI of each compound.

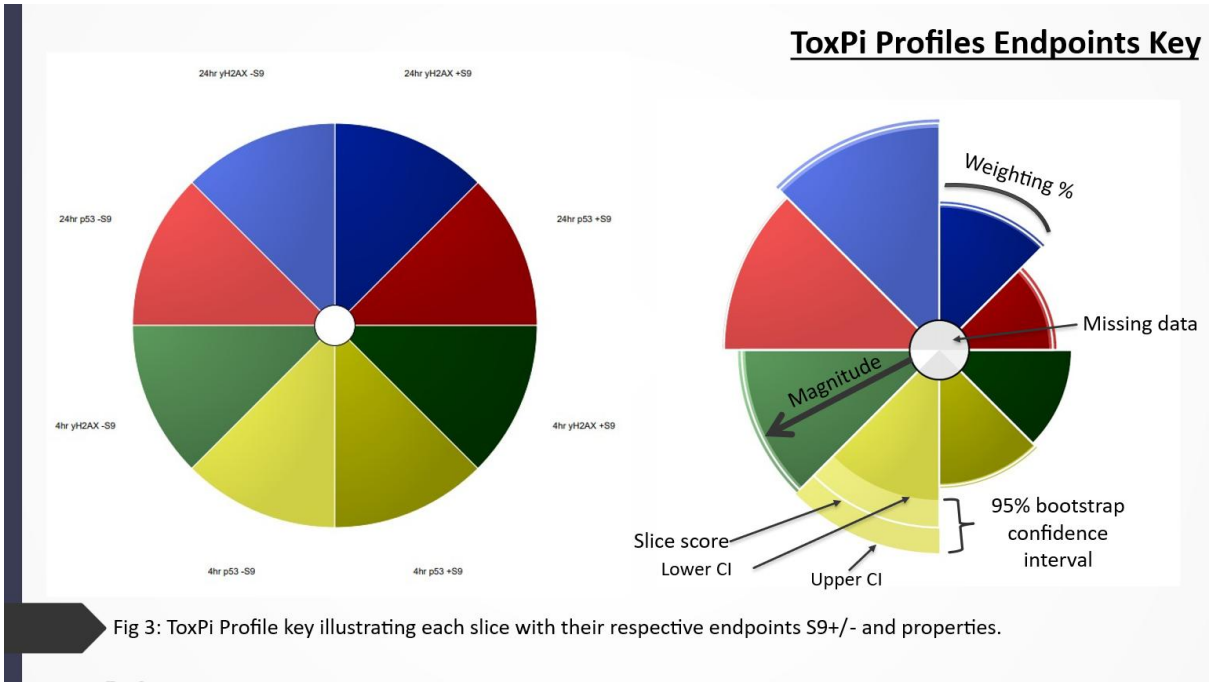


Fig 3: ToxPi Profile key illustrating each slice with their respective endpoints S9+/- and properties.

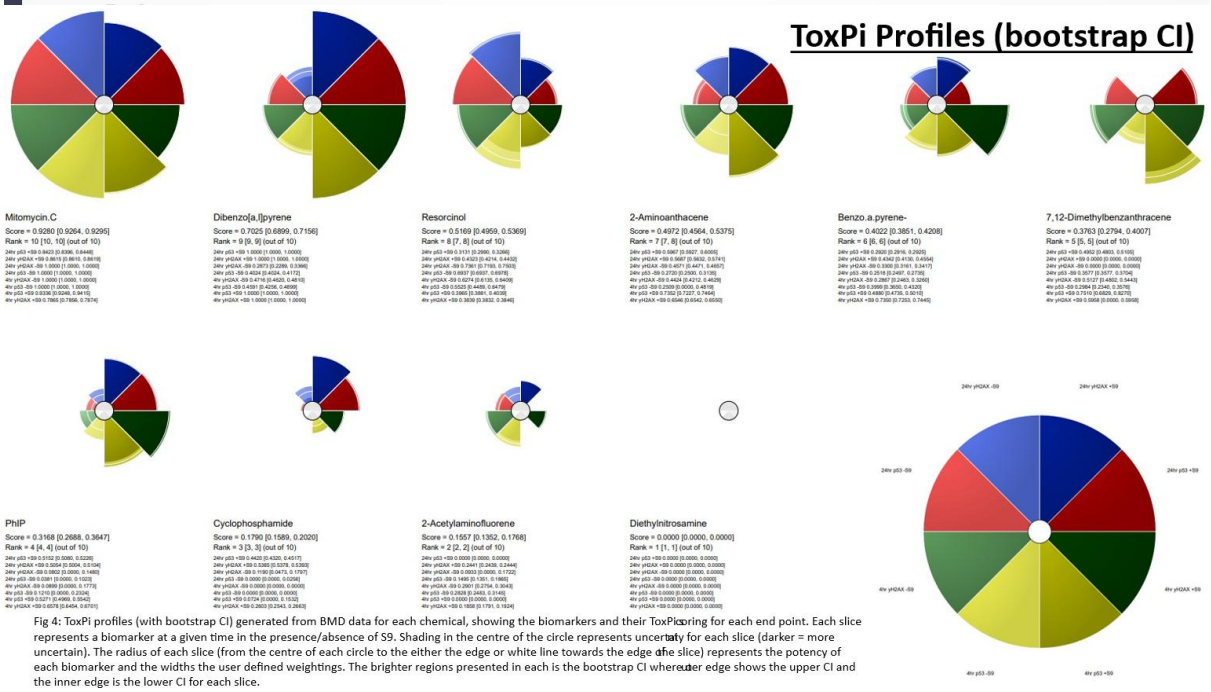
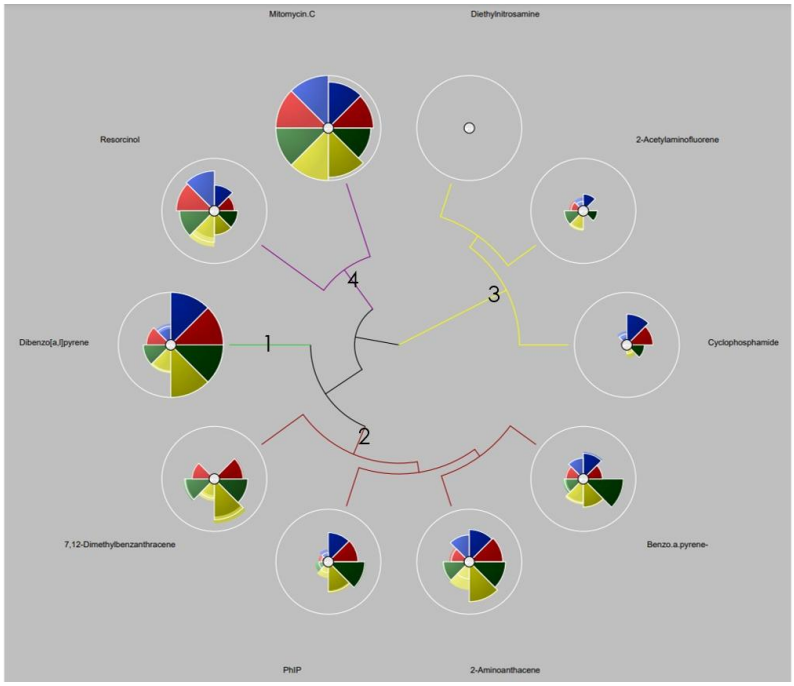


Fig 4: ToxPi profiles (with bootstrap CI) generated from BMD data for each chemical, showing the biomarkers and their ToxPi scoring for each end point. Each slice represents a biomarker at a given time in the presence/absence of S9. Shading in the centre of the circle represents uncertainty for each slice (darker = more uncertain). The radius of each slice (from the centre of each circle to the either the edge or white line towards the edge of the slice) represents the potency of each biomarker and the widths the user defined weightings. The brighter regions presented in each is the bootstrap CI where outer edge shows the upper CI and the inner edge is the lower CI for each slice.



Hierarchical Clustering (complete)

#clusters = 4

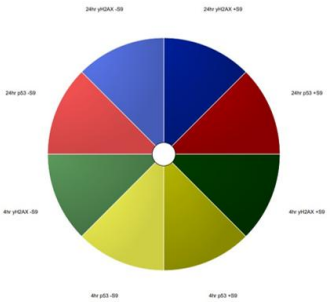
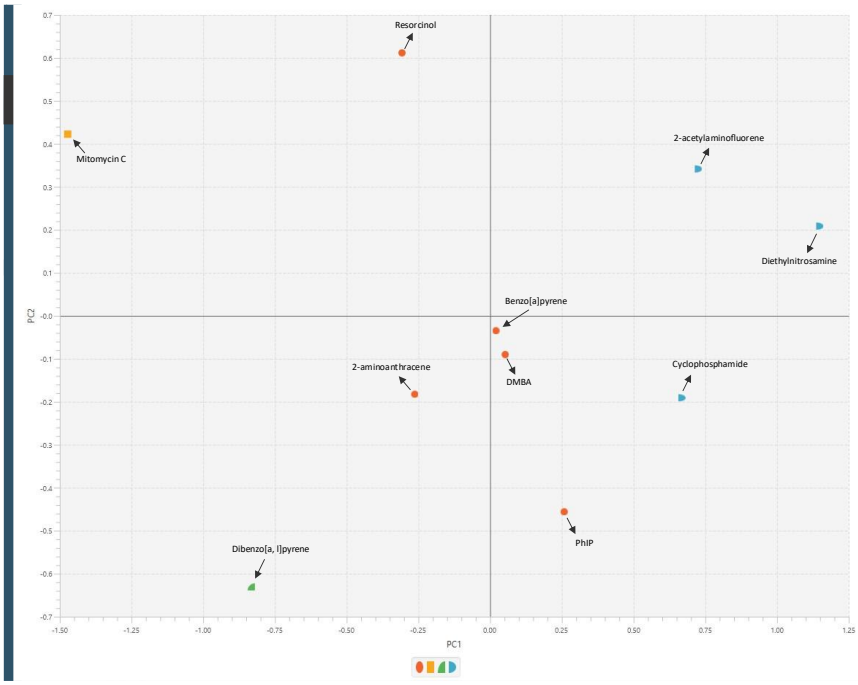


Fig 5: ToxPi profiles hierarchical clustering generated through complete divisive clustering to show grouping of each chemical.



K-Means Clustering

Fig 6: ToxPi K means PCA clustering where each compound is grouped based on variation across the ToxPi scores/genotoxicity and biomarker sensitivity. Each point represents a compound, and each colour/shape represents a cluster they have been placed into.
For PC1 compounds with higher ToxPi scores are towards the left, and for PC2, compounds with increased potency with S9+ are at the bottom.

#clusters = 4

A comparison of classical and 21st century genotoxicity tools: A proof of concept study of 18 chemicals comparing in vitro micronucleus, ToxTracker and genomics-based methods (TGx-DDI, whole genome clustering and connectivity mapping)

Ashley Allemang[1] | K. Nadira De Abrew[1] | Yuqing K. Shan[1] | Jesse M. Krailler[2] | Stefan Pfuhler[1]

[1]Global Product Stewardship, The Procter & Gamble Company, Cincinnati, Ohio

[2]Data and Modeling Sciences, The Procter & Gamble Company, Cincinnati, Ohio

Toxtracker and TGx-DDI Assay Data - ToxPi Ranking

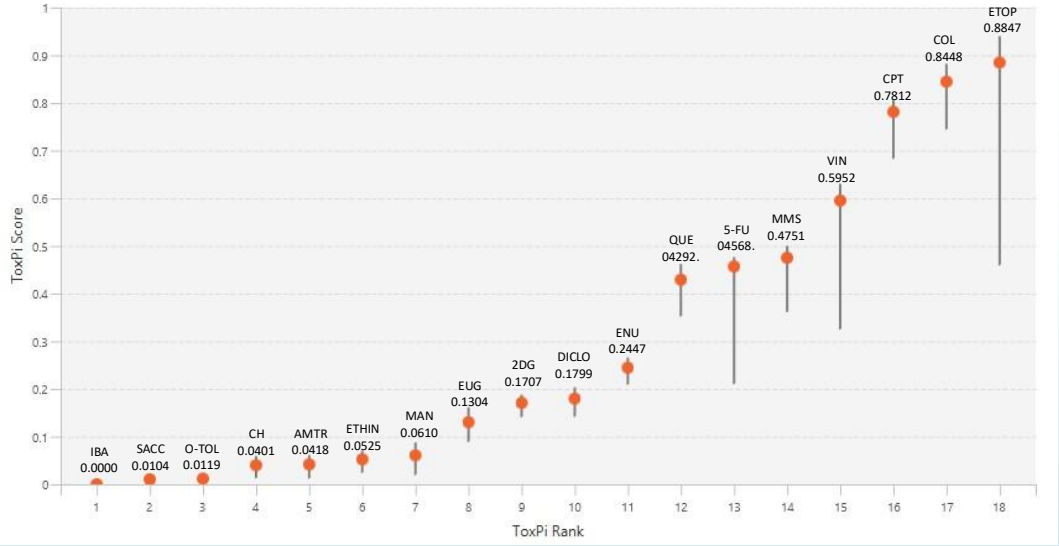


Fig 7: Results for Objective 2, showing ToxPi ranking for BMD data on the 18 test compounds (1 = lowest potency and 18 = highest potency). ToxPi scores ranging between 0-1 where the higher score translates to higher genotoxic potency and the error bars represent the CI of each compound.

ToxPi Profiles Endpoints Key

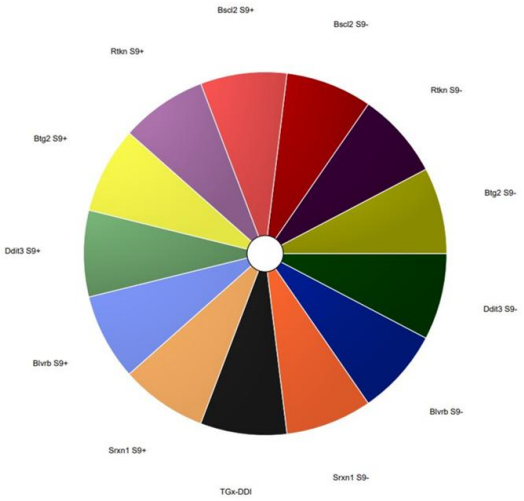
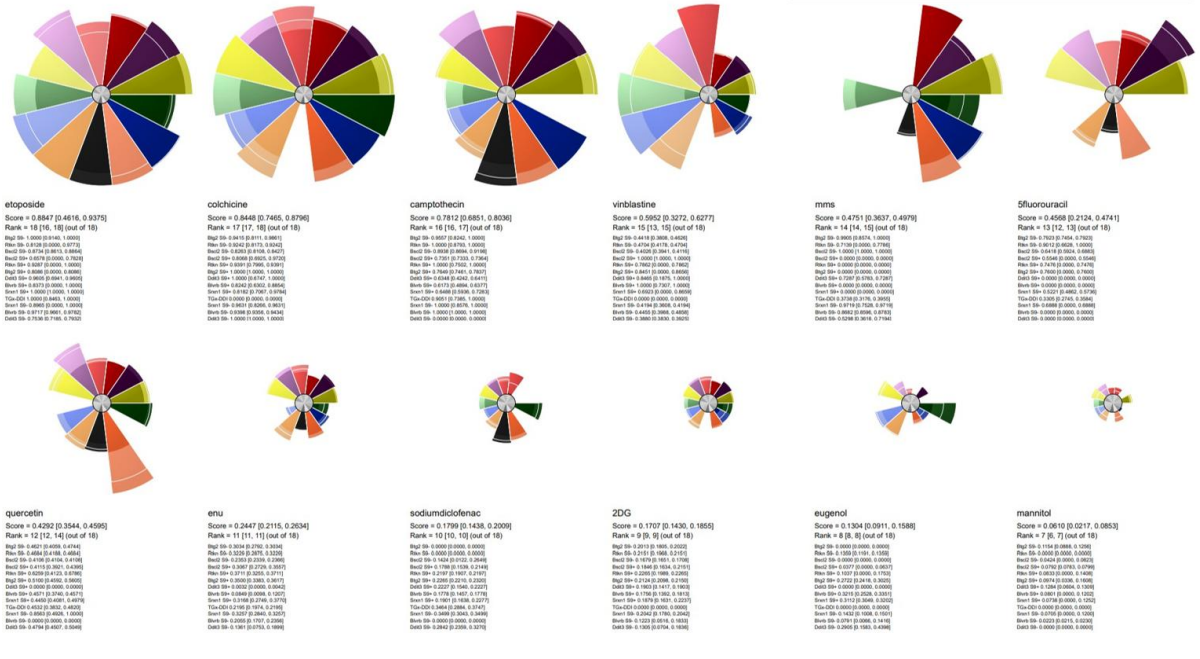


Fig 8: ToxPi Profile key illustrating each slice with each Toxtracker and TGx-DDI endpoint.

ToxPi Profiles (bootstrap CI)



ToxPi Profiles (bootstrap CI)

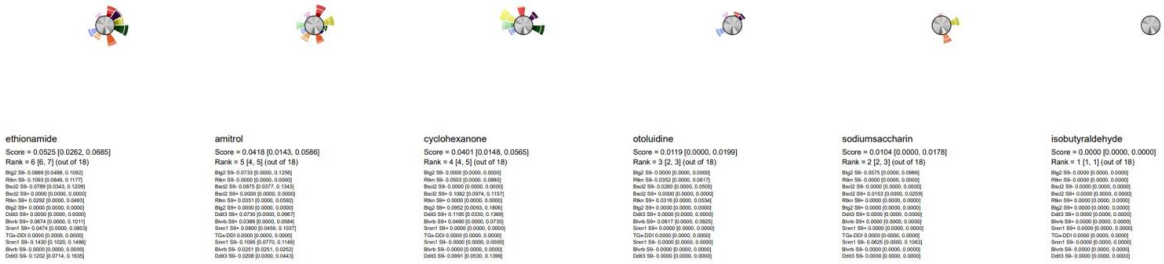
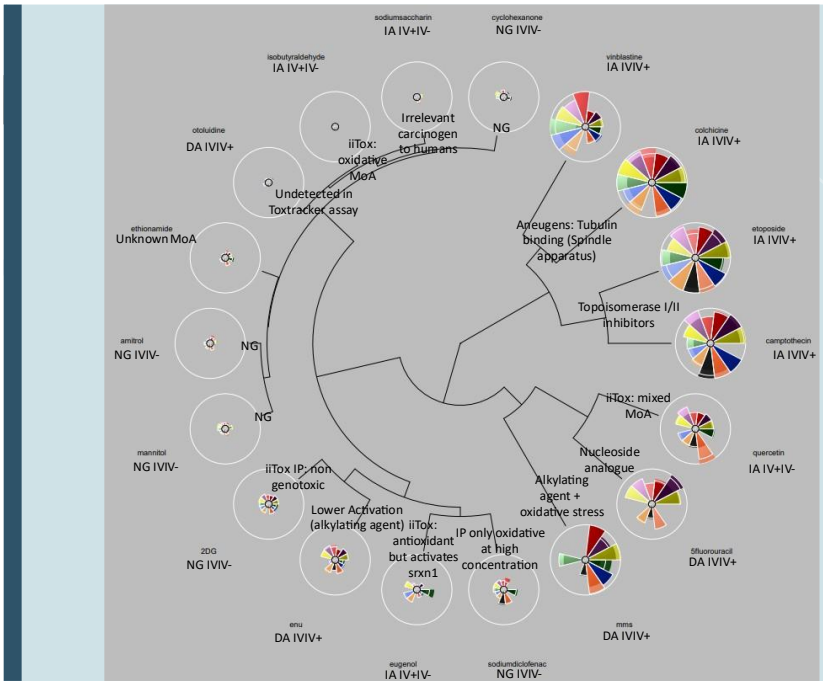


Fig 9: ToxPi profiles (with bootstrap CI) generated from BMD data for each chemical, showing the biomarkers and their ToxPi scoring for each end point. Each slice represents a biomarker at a given time period in the presence/absence of S9. Shading in the centre of the circle represents uncertainty for each slice (darker = more uncertain). The radius of each slice (from the centre of each circle to the either the edge or white line towards the edge of the slice) represents the potency of each biomarker and the widths the user defined weightings. The brighter regions presented in each is the bootstrap CI where outer edge shows the upper CI, and the inner edge is the lower CI for each slice.



Hierarchical Clustering (complete)

Key:

- MoA: Mode of Action
- IP: Irrelevant Positive
- iiTox: Incorrectly Identified (Toxtracker)
- NG: Non-Genotoxic
- DA: Direct Acting
- IA: Indirect Acting
- IVIV+: *in vitro* and *in vivo* positive
- IVIV-: *in vitro* and *in vivo* negative
- IV+IV-: *in vitro* positive and *in vivo* negative

Fig 10: ToxPi profiles hierarchical clustering generated through complete AGNES clustering to show grouping of each chemical.

K-Means Clustering

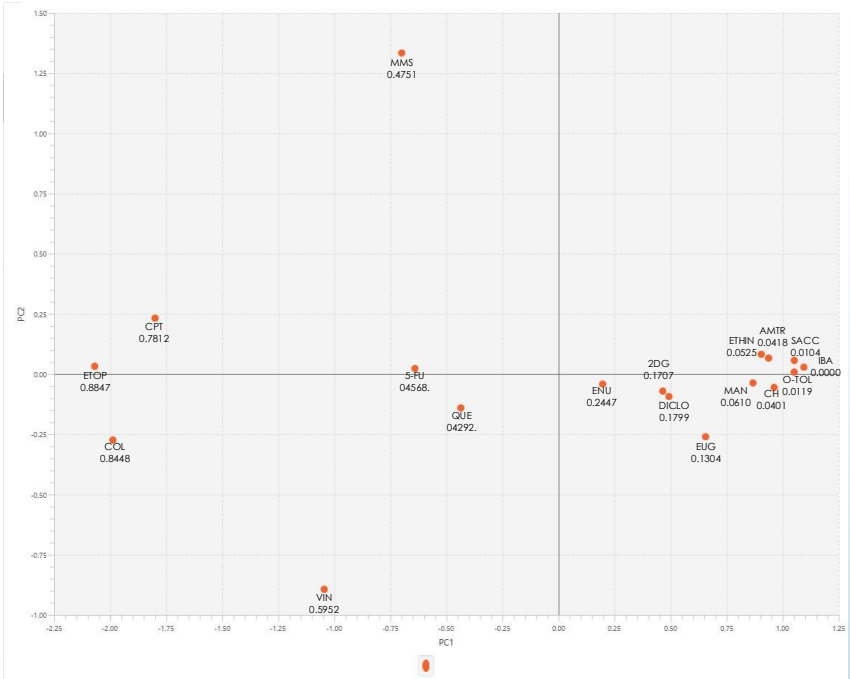


Fig 11: ToxPi K means PCA clustering where each compound is grouped based on variation across the ToxPi scores (PC1), where higher ToxPi scores are further left and biomarker variation (PC2), where compounds with higher potency at S9+ sit lower.

Conclusions

- Genotoxic effects of S9+/- on each compound is effectively illustrated in ToxPi profiles, drawing comparisons between parental compounds/metabolites.
- Hierarchical clustering can compartmentalise the clastogens into groups according to similarities in Toxpi scores, biomarker sensitivity and S9 activation properties.
- Clearly highlights the ToxPi profiles displaying the most genotoxic biomarkers at each time point for each clastogens and their BMD CI
- Effectively reduced the number of figures required compared to the main study, easing comparative analysis capabilities, and providing increased genotoxic endpoint information for each compound.
- Compounds weren't strongly classified mechanistically and/or according to their in vitro/vivo genotoxic positive/negative attributes during the clustering component.

Appendix 6. Beal 2022 IVIVE presentation

Health Canada Santé Canada

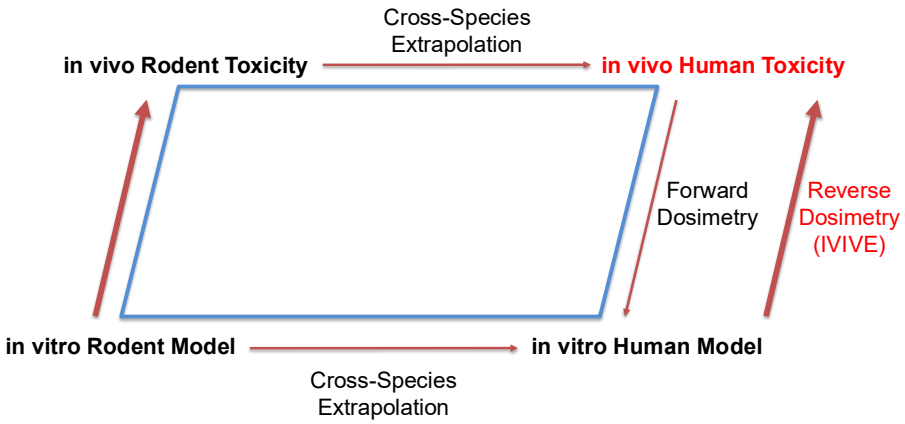
Canada

Relating in vitro Response to in vivo Exposures – The Art of Toxicokinetics

Marc Beal, PhD
Regulatory Toxicology Research Division
Food Directorate
Health Canada

YOUR HEALTH AND SAFETY... OUR PRIORITY.

Different Extrapolations for Predicting Human Toxicity



IVIVE or reverse dosimetry uses toxicokinetic (TK) models to predict the in vivo dose required to achieve an in vitro concentration in blood or target tissue

Appendix 7. Genetic Toxicology Association (GTA) 2025 Annual Meeting Abstract

Title:

Advancing Mechanistic Classification of Genotoxic Compounds: Benchmark Dose Analysis and Visualisation Pipelines for the iScreen high-content imaging platform

Authors:

Yusuf Hussien (Swansea University), Wen Sun (Pfizer), George Johnson (Swansea University)

Abstract:

Modelling algorithms, such as the benchmark dose (BMD), help determine points-of-departure by normalizing dose-response metrics, enabling development of pipelines and assays to visualize and evaluate genotoxic compounds. The iScreen platform uses antibodies and high-throughput imaging to find micronuclei centromere detection/fragmentation and DNA damage response biomarkers. Dose-response potency for 10 genotoxicants was analyzed using PROAST to determine BMD confidence intervals while ToxPi scored, profiled and clustered chemicals based on biomarker sensitivity. ToxPi also uses a biologically relevant weighting system to improve genotoxic potency analysis and combine data across endpoints in compliance with regulatory standards. These techniques optimized genotoxic potency analysis by integrating dose-response data across iScreen endpoints to describe chromosomal instability and group compounds by mechanism. This approach advances *in vitro* dose-response visualization systems by using specific antibody markers to quantify biomarker response, differentiate clastogens/aneugens, and highlight chromosomal breaks and segregation events, evolving the landscape of hazard detection and comparative analysis to safeguard health and safety.

Appendix 8. BEMS 2024 Networking Lunch Presentation

An investigation on the mutagenicity of compounds using Benchmark Dose and Dose-Response data for research in genetic toxicology

Yusuf Hussien BSc (Hons)

uOttawa

Swansea University
Prifysgol Abertawe

HESI Genetic Toxicology Technical Committee

Health Canada

P&G

LITRON LABORATORIES

EMGS Environmental Mutagenesis and Genomics Society
Synergistic Interactions for a Better World

ToxPi

ToxPi (v2.3) is an interactive visualisation programme that analyses the genotoxic potency of chemicals. Translating BMD data into a single ToxPi score, ranking the mutagenicity on a continuous spectrum instead of fixed groups (Reif et al., 2012).

ToxPi features K-means and Hiearchical clustering uses PCA techniques for dataset dimensionality reduction and AGNES algorithms to group chemicals based on similarities in ToxPi scoring and potency across endpoints (Marvel et al., 2018).

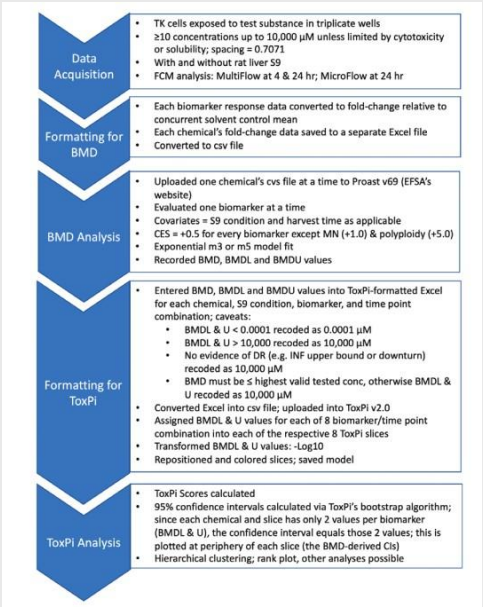
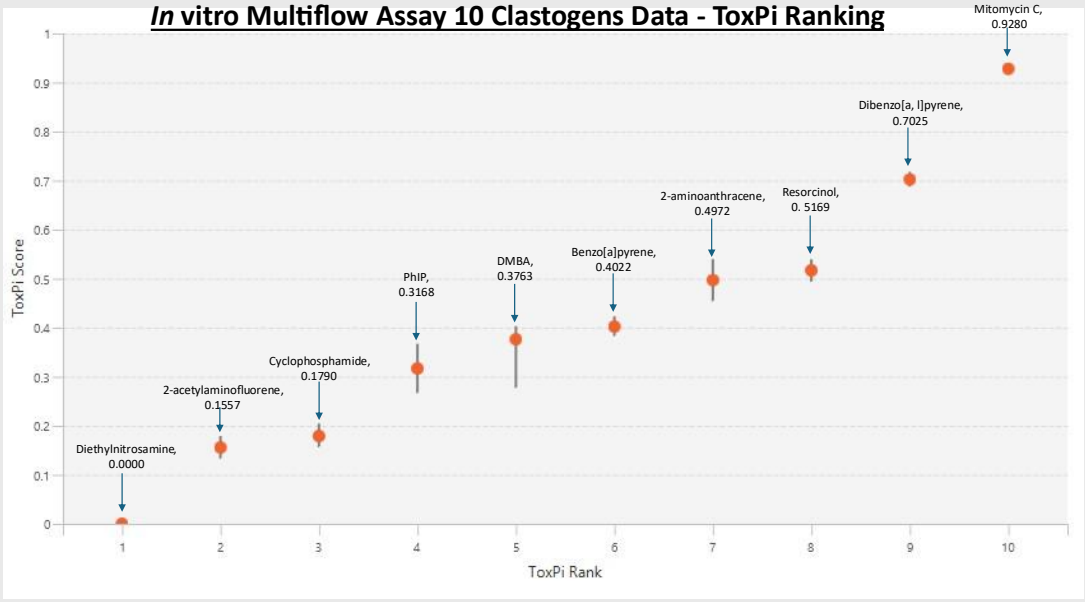


Fig 1: An adaptation of the all procedures by (Avlasenvich Et al., 2021).

Live Demonstration

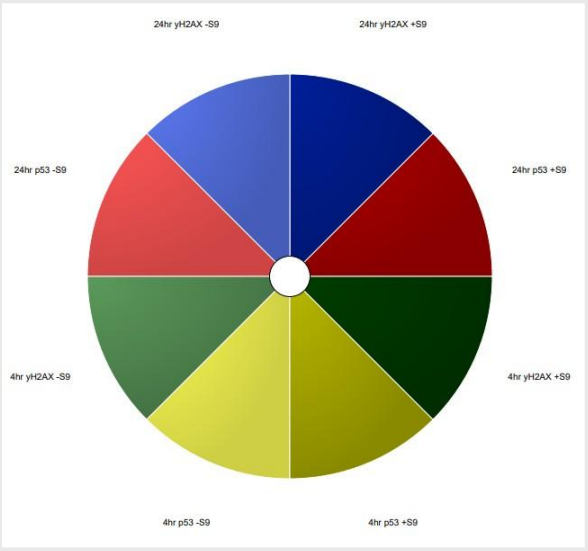
Row	A_Priori_MoA	Classification	Chemical	γH2AX_BMDL	γH2AX_BMDU	p53_BMDL	p53_BMDU	p-H3_BMDL	p-H3_BMDU	MN_BMDL	MN_BMDU
1	Tubulin	Aneugen	Colchicine	10000	10000	0.0138	0.0153	0.00776	0.00975	0.00944	0.0106
2	Tubulin	Aneugen	Vinblastine	10000	10000	0.000219	0.000247	0.000448	0.000953	0.000102	0.000133
3	Aurora	Aneugen	AMG900	10000	10000	0.00956	0.0116	10000	10000	0.00436	0.068
4	Aurora	Aneugen	VX680	10000	10000	0.0351	0.0393	10000	10000	0.045	0.0522
5	TopoII	Clastogen	Genistein	99.29999989	154.9999999	268	1260	15.89999999	50.39999995	6.08	12
6	TopoII	Clastogen	Emodin	89.19999992	137.9999999	314	546.9999996	35.80000003	91.19999993	0.00126	0.001
7	Alkylator	Clastogen	Cisplatin	1.32	2.06	10000	10000	10000	10000	1.21	1.54
8	Alkylator	Clastogen	Melphalan	1.12	1.58	10000	10000	10000	10000	0.0196	0.0485
9	Non_genotoxic	Non-Geno	D-Mannitol	10000	10000	10000	10000	10000	10000	10000	10000
10	Non_genotoxic	Non-Geno	Cycloheximide	10000	10000	10000	10000	10000	10000	10000	10000



Abbreviations:
DMBA = 7,12-Dimethylbenzanthracene
PhIP = 2-Amino-1-methyl-6-phenylimidazo[4,5-b]pyridine

Fig 2: Results for Objective 2, showing ToxPi ranking for BMD data on 10 clastogens (1 = lowest potency and 10 = highest potency). ToxPi scores ranging between 0-1 where the higher score translates to higher genotoxic potency and the error bars represent the CI of each compound.

ToxPi Profiles Endpoints Key



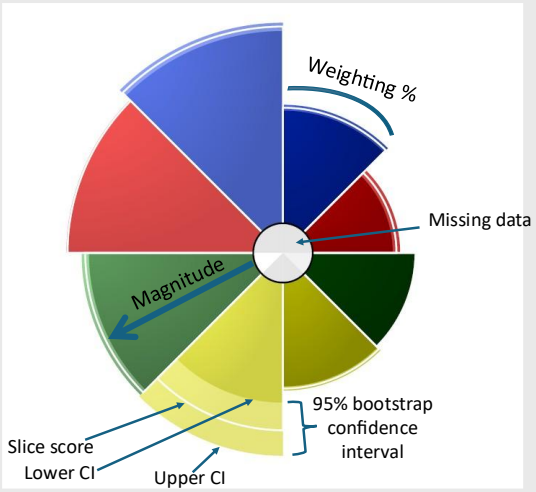


Fig 3: ToxPi Profile key illustrating each slice with their respective endpoints S9+/- and properties.

ToxPi Profiles (bootstrap CI)

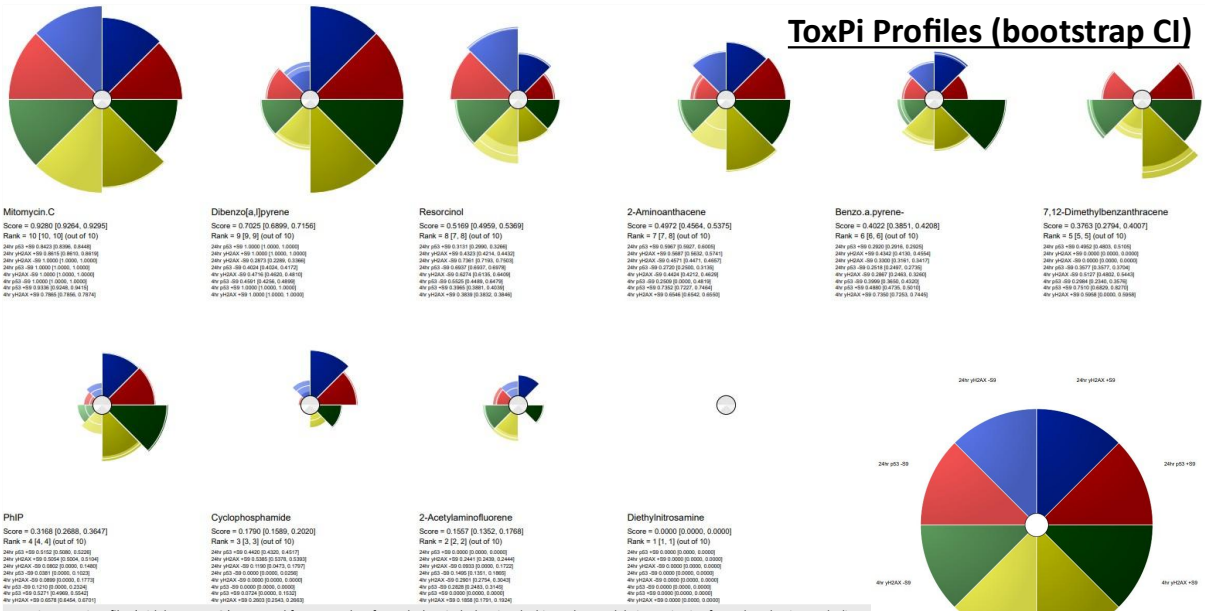
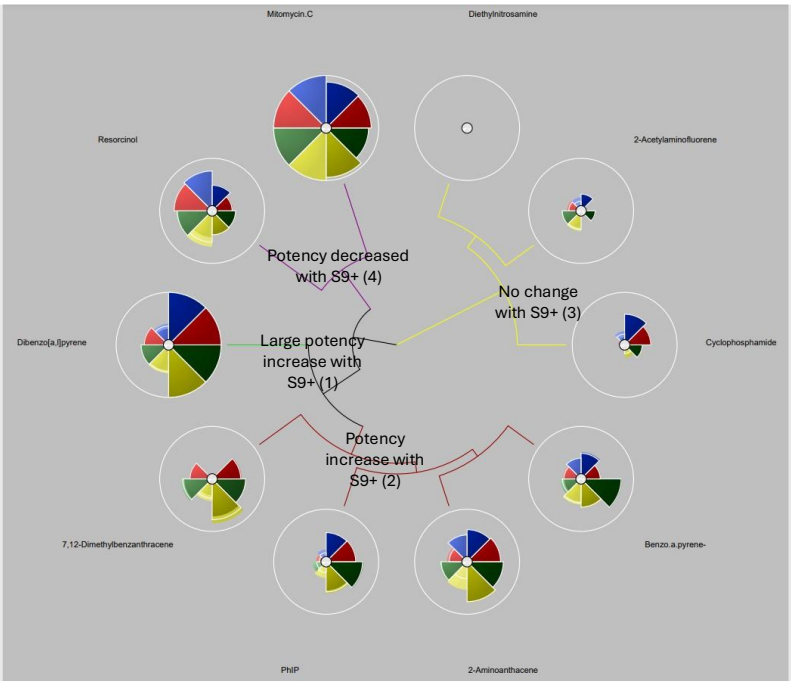
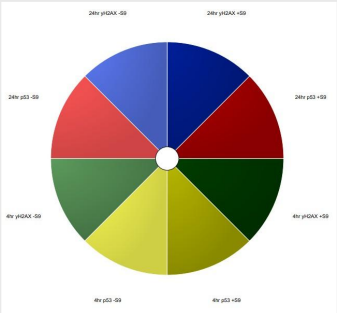


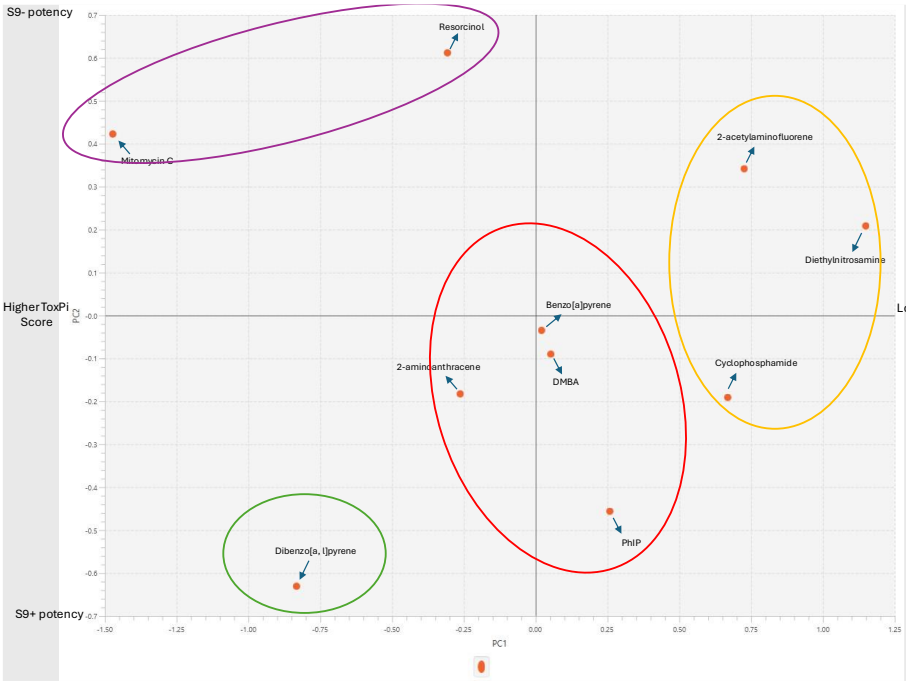
Fig 4: ToxPi profiles (with bootstrap CI) generated from BMD data for each chemical, showing the biomarkers and their ToxPi scoring for each end point. Each slice represents a biomarker at a given time in the presence/absence of S9. Shading in the centre of the circle represents uncertainty for each slice (darker = more uncertain). The radius of each slice (from the centre of each circle to the either the edge or white line towards the edge of the slice) represents the potency of each biomarker and the widths the user defined weightings. The brighter regions presented in each is the bootstrap CI where outer edge shows the upper CI and the inner edge is the lower CI for each slice.



Hierarchical Clustering (complete)

#clusters = 4





K-Means Clustering

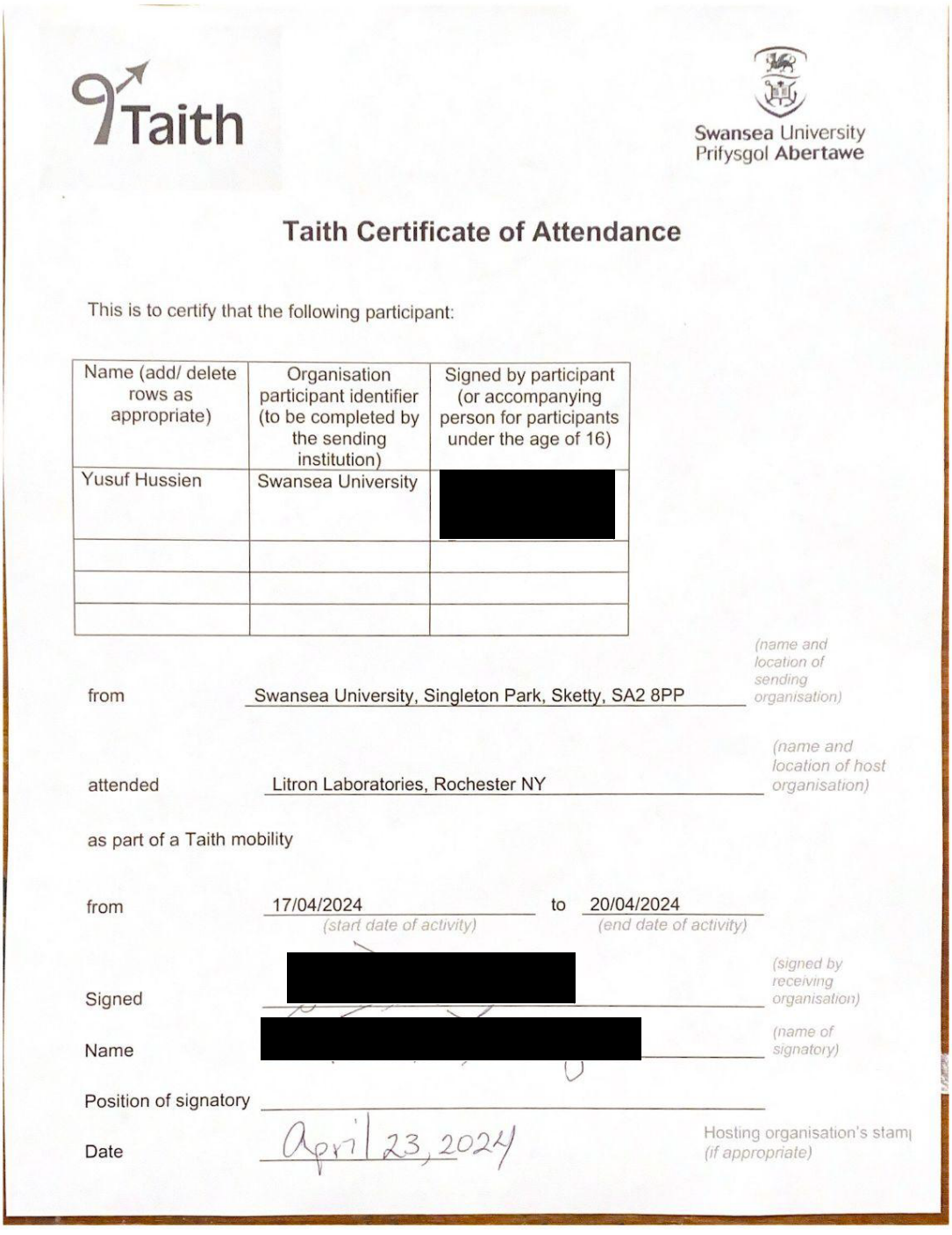
Fig 6: ToxPi K means PCA clustering where each compound is grouped based on variation across the ToxPi scores/genotoxicity and biomarker sensitivity. Each point represents a compound, and each colour/shape represents a cluster they have been placed into.
For PC1 compounds with higher ToxPi scores are towards the left, and for PC2, compounds with increased potency with S9+ are at the bottom.

#clusters = 4

Conclusions

- Genotoxic effects of S9+/- on each compound is effectively illustrated in ToxPi profiles, drawing comparisons between parental compounds/metabolites.
- Hierarchical clustering can compartmentalise the clastogens into groups according to similarities in Toxpi scores, biomarker sensitivity and S9 activation properties.
- Clearly highlights the ToxPi profiles displaying the most genotoxic biomarkers at each time point for each clastogens and their BMD CI
- Effectively reduced the number of figures required compared to the main study, easing comparative analysis capabilities, and providing increased genotoxic endpoint information for each compound.

Appendix 9. Litron Laboratories Collaboration Taith funding Certification of Attendance



Taith

Swansea University
Prifysgol Abertawe

Taith Certificate of Attendance

This is to certify that the following participant:

Name (add/ delete rows as appropriate)	Organisation participant identifier (to be completed by the sending institution)	Signed by participant (or accompanying person for participants under the age of 16)
Yusuf Hussien	Swansea University	[redacted]

from Swansea University, Singleton Park, Sketty, SA2 8PP *(name and location of sending organisation)*

attended Litron Laboratories, Rochester NY *(name and location of host organisation)*

as part of a Taith mobility

from 17/04/2024 *(start date of activity)* to 20/04/2024 *(end date of activity)*

Signed [redacted] *(signed by receiving organisation)*

Name [redacted] *(name of signatory)*

Position of signatory

Date April 23, 2024

Hosting organisation's stamp *(if appropriate)*

Glossary

Administered Equivalent Doses – A normalised dose metric where oral, inhaled, and dermal exposure routes contribute to a specific biological effect which is measured through the combined use of toxicodynamic and toxicokinetic parameters.

Adverse Outcome Pathway – A framework that links molecular initiating events (e.g. DNA damage) to an adverse effect outcome, through key biological events.

Aneugen – An agent that causes abnormalities to the number chromosomes in daughter cells by targeting non-DNA components.

Assay – A laboratory procedure to measure a biological, chemical, or physical property (e.g., enzyme activity, cytotoxicity, gene expression).

Benchmark Dose – A dose/concentration that produces a predetermined change in response derived due to an adverse effect caused by an agent.

Clastogen – An agent that causes chromosomal breaks or structural damage to the DNA directly through mechanisms such as deletions, insertion or rearrangement of the chromosomes.

Confidence Interval – A statistical range indicating the reliability of an estimate, where the true value is expected to lie within the upper and lower bounds within a given probability (e.g. 90% confidence interval).

DNA Damage – Physical alterations to DNA (e.g. breaks, adducts, crosslinks), which can lead to mutations or cell death if unrepaired.

Dose– Response Relationship – The quantitative correlation between the dose/exposure level of a substance and the magnitude of its biological effect.

Hierarchical Clustering – An unsupervised machine learning algorithm that seeks to build hierarchies of grouped data based on similar data points and converting them into nested clusters based on Euclidian distance metrics.

K-means Clustering – An algorithm that partitions data into a user-selected number of k clusters and minimising variation within each cluster.

Mode of Action – The detailed biological mechanism by which a substance produces an effect.

Point-of-Departure – A dose (e.g., BMD, NOAEL) used as a starting point for risk/potency assessment calculations.

Principal Component Analysis – A dimensionality-reduction technique that transforms data into uncorrelated variables (principal components) to highlight patterns.

Quantitative dataset – Numerical data suitable for statistical analysis.

S9 (fraction) – A liver-derived metabolic activation system (containing cytochrome P450 enzymes) used in assays to mimic mammalian metabolism.

References

Abdelhameed, A. S., Attwa, M. W., & Kadi, A. A. (2020). Characterization of stable and reactive metabolites of the anticancer drug, ensartinib, in human liver microsomes using LC-MS/MS: An in silico and practical bioactivation approach. *Drug Design, Development and Therapy, 14*, 5259--5273. https://doi.org/10.2147/DDDT.S274018

Allemang, A., De Abrew, K. N., Shan, Y. K., Krailler, J. M., & Pfuhler, S. (2020). A comparison of classical and 21st century genotoxicity tools: A proof of concept study of 18 chemicals comparing *in vitro* micronucleus, ToxTracker and genomics-based methods (tgx-ddi, whole genome clustering and connectivity mapping). *Environmental and Molecular Mutagenesis, 62*(2), 92--107. https://doi.org/10.1002/em.22418

Allemang, A., Mahony, C., Lester, C., & Pfuhler, S. (2018). Relative potency of fifteen pyrrolizidine alkaloids to induce DNA damage as measured by micronucleus induction in HepaRG human liver cells. *Food and Chemical Toxicology, 121*, 72--81. https://doi.org/10.1016/j.fct.2018.08.003

Anderson, L., Logsdon, D., Ruskie, S., Fox, S., Issaq, H., Kovatch, R., & Riggs, C. (1994). Promotion by polychlorinated biphenyls of lung and liver tumors in mice. *Carcinogenesis, 15*(10), 2245--2248. https://doi.org/10.1093/carcin/15.10.2245

Arif, J. M., & Gupta, R. C. (1997). Microsome-mediated bioactivation of dibenzo[a,l]pyrene and identification of DNA adducts by 32P-postlabeling. *Carcinogenesis, 18*(10), 1999--2007. https://doi.org/10.1093/carcin/18.10.1999

Aronson, J. (2007). Concentration-effect and dose-response relations in clinical pharmacology. *British Journal of Clinical Pharmacology, 63*(3), 255--257. https://doi.org/10.1111/j.1365-2125.2006.02800.x

Attia, S. M. (2010). Deleterious effects of reactive metabolites. *Oxidative Medicine and Cellular Longevity, 3*(4), 238--253. https://doi.org/10.4161/oxim.3.4.13246

Attia, S. M., Kliesch, U., Schriever-Schwemmer, G., Badary, O. A., Hamada, F. M., & Adler, I. -D. (2003). Etoposide and Merbarone are clastogenic and aneugenic in the mouse bone marrow micronucleus test complemented by fluorescence in situ hybridization with the mouse *minor* satellite DNA probe. *Environmental and Molecular Mutagenesis*, *41*(2), 99–103. https://doi.org/10.1002/em.10135

Attia, S. M. (2011). Molecular cytogenetic evaluation of the aneugenic effects of teniposide in somatic and germinal cells of male mice. *Mutagenesis*, *27*(1), 31–39. https://doi.org/10.1093/mutage/ger051

Auerbach, S., Filer, D., Reif, D., Walker, V., Holloway, A. C., Schlezinger, J., ... Thayer, K. A. (2016). Prioritizing environmental chemicals for obesity and diabetes outcomes research: A screening approach using ToxCast™ high-throughput data. *Environmental Health Perspectives, 124*(8), 1141--1154. https://doi.org/10.1289/ehp.1510456

Avlasevich, S., Pellegrin, T., Godse, M., Bryce, S., Bemis, J., Bajorski, P., & Dertinger, S. (2021). Biomarkers of DNA damage response improve *in vitro* micronucleus assays by revealing genotoxic mode of action and reducing the occurrence of irrelevant positive results. *Mutagenesis, 36*(6), 407--418. https://doi.org/10.1093/mutage/geab039

Avlasevich, S. L., Bryce, S. M., Cairns, S. E., & Dertinger, S. D. (2006). *In vitro* micronucleus scoring by flow cytometry: Differential staining of micronuclei versus apoptotic and necrotic chromatin enhances assay reliability. *Environmental and Molecular Mutagenesis, 47*(1), 56--66. https://doi.org/10.1002/em.20170

Avlasevich, S., Bryce, S., De Boeck, M., Elhajouji, A., Van Goethem, F., Lynch, A., Nicolette, J., Shi, J., & Dertinger, S. (2011). Flow cytometric analysis of micronuclei in mammalian cell cultures: Past, present and future. *Mutagenesis, 26*(1), 147--152. https://doi.org/10.1093/mutage/geq061

Baldrick, P. (2021). Genotoxicity Test Battery -- an assessment of its utility in early drug development. *Mutation Research/Genetic Toxicology and Environmental Mutagenesis, 868--869*, 503388. https://doi.org/10.1016/j.mrgentox.2021.503388

Barnes, J., Zubair, M., John, K., Poirier, M., & Martin, F. (2018). Carcinogens and DNA damage. *Biochemical Society Transactions, 46*(5), 1213--1224. https://doi.org/10.1042/BST20180519

Bartsch, H., & Montesano, R. (1984). Relevance of nitrosamines to human cancer. *Carcinogenesis, 5*(11), 1381--1393. https://doi.org/10.1093/carcin/5.11.1381

Beal, M. A., Gagne, M., Kulkarni, S. A., Patlewicz, G., Thomas, R. S., & Barton-Maclaren, T. S. (2022). Implementing *in vitro* bioactivity data to modernize priority setting of chemical inventories. *ALTEX, 39*, 123--139.

Beal, M. A., Gagne, M., Kulkarni, S. A., Patlewicz, G., Thomas, R. S., & Barton-Maclaren, T. S. (2022). Quantitative *in vitro* to *in vivo* extrapolation of genotoxicity data provides protective estimates of *in vivo* dose. *Environmental and Molecular Mutagenesis, 64*(2), 105--122. https://doi.org/10.1002/em.22521

Beal, M. A. (2022). *Relating in vitro response to in vivo exposures -- The art of toxicokinetics* [Presentation].

Beedanagari, S. (2017). Genetic toxicology. *Comprehensive Medicinal Chemistry III*, 195--203. https://doi.org/10.1016/b978-0-12-409547-2.12381-9

Bemis, J. C., Wills, J. W., Bryce, S. M., Torous, D. K., Dertinger, S. D., & Slob, W. (2016). Comparison of *in vitro* and *in vivo* clastogenic potency based on benchmark dose analysis of flow cytometric micronucleus data. *Mutagenesis, 31*(3), 277--285. https://doi.org/10.1093/mutage/gev041

Benigni, R. (2021). In silico assessment of genotoxicity. Combinations of sensitive structural alerts minimize false negative predictions for all genotoxicity endpoints and can single out

chemicals for which experimentation can be avoided. *Regulatory Toxicology and Pharmacology, 126*, 105042. https://doi.org/10.1016/j.yrtph.2021.105042

Bhogal, N., Grindon, C., Combes, R., & Balls, M. (2005). Toxicity testing: Creating a revolution based on new technologies. *Trends in Biotechnology, 23*(6), 299--307. https://doi.org/10.1016/j.tibtech.2005.04.006

Bond, V. P., Benary, V., Sondhaus, C. A., & Feinendegen, L. E. (1995). The meaning of linear dose-response relations, made evident by use of absorbed dose to the cell. *Health Physics, 68*(6), 786--792. https://doi.org/10.1097/00004032-199506000-00006

Boots, A. W., Li, H., Schins, R. P. F., Duffin, R., Heemskerk, J. W. M., Bast, A., & Haenen, G. R. M. M. (2007). The quercetin paradox. *Toxicology and Applied Pharmacology, 222*(1), 89--96. https://doi.org/10.1016/j.taap.2007.02.016

Breen, M., Wambaugh, J. F., Bernstein, A., Sfeir, M., & Ring, C. L. (2022). Simulating toxicokinetic variability to identify susceptible and highly exposed populations. *Journal of Exposure Science & Environmental Epidemiology, 32*(6), 855--863. https://doi.org/10.1038/s41370-022-00491-0

Bryce, S. M., Avlasevich, S. L., Bemis, J. C., Phonethepswath, S., & Dertinger, S. D. (2010). Miniaturized flow cytometric *in vitro* micronucleus assay represents an efficient tool for comprehensively characterizing genotoxicity dose-response relationships. *Mutation Research/Genetic Toxicology and Environmental Mutagenesis, 703*(2), 191--199. https://doi.org/10.1016/j.mrgentox.2010.08.010

Bryce, S. M., Avlasevich, S. L., Bemis, J. C., Tate, M., Walmsley, R. M., Saad, F., ... Dertinger, S. D. (2013). Flow cytometric 96-well microplate-based *in vitro* micronucleus assay with human TK6 cells: Protocol optimization and transferability assessment. *Environmental and Molecular Mutagenesis, 54*(3), 180--194. https://doi.org/10.1002/em.21760

Bryce, S. M., Avlasevich, S. L., Bemis, J. C., & Dertinger, S. D. (2010). Miniaturized flow cytometry-based CHO-K1 micronucleus assay discriminates aneugenic and clastogenic modes of action. *Environmental and Molecular Mutagenesis, 52*(4), 280--286. https://doi.org/10.1002/em.20618

Bryce, S. M., Bemis, J. C., Avlasevich, S. L., & Dertinger, S. D. (2007). *In vitro* micronucleus assay scored by flow cytometry provides a comprehensive evaluation of cytogenetic damage and cytotoxicity. *Mutation Research/Genetic Toxicology and Environmental Mutagenesis, 630*(1--2), 78--91. https://doi.org/10.1016/j.mrgentox.2007.03.002

Bryce, S. M., Bernacki, D. T., Bemis, J. C., Spellman, R. A., Engel, M. E., Schuler, M., Lorge, E., Heikkinen, P. T., Hemmann, U., Thybaud, V., et al. (2017). Interlaboratory evaluation of a multiplexed high information content *in vitro* genotoxicity assay. *Environmental and Molecular Mutagenesis, 58*, 146--161. https://doi.org/10.1002/em.22080

Bryce, S. M., Bernacki, D. T., Bemis, J. C., & Dertinger, S. D. (2016). Genotoxic mode of action predictions from a multiplexed flow cytometric assay and a machine learning approach. *Environmental and Molecular Mutagenesis, 57*(3), 171--189. https://doi.org/10.1002/em.21994

Bryce, S. M., Bernacki, D. T., Smith-Roe, S. L., Witt, K. L., Bemis, J. C., & Dertinger, S. D. (2017). Investigating the generalizability of the multiflow ® DNA damage assay and several companion machine learning models with a set of 103 diverse Test Chemicals. *Toxicological Sciences, 162*(1), 146--166. https://doi.org/10.1093/toxsci/kfx235

Bryce, S. M., Shi, J., Nicolette, J., Diehl, M., Sonders, P., Avlasevich, S., Raja, S., Bemis, J. C., & Dertinger, S. D. (2010). High content flow cytometric micronucleus scoring method is applicable to attachment cell lines. *Environmental and Molecular Mutagenesis, 51*(3), 260--266. https://doi.org/10.1002/em.20544

Buick, J. K., Williams, A., Gagné, R., Swartz, C. D., Recio, L., Ferguson, S. S., & Yauk, C. L. (2020). Flow cytometric micronucleus assay and TGX-DDI transcriptomic biomarker analysis of ten genotoxic and non-genotoxic chemicals in human HepaRG™ cells. *Genes and Environment, 42*(1), 1--12. https://doi.org/10.1186/s41021-019-0139-2

Buick, J. K., Williams, A., Meier, M. J., Swartz, C. D., Recio, L., Gagné, R., ... Yauk, C. L. (2021). A modern genotoxicity testing paradigm: Integration of the high-throughput CometChip® and the TGX-DDI transcriptomic biomarker in human hepaRG™ cell cultures. *Frontiers in Public Health, 9*. https://doi.org/10.3389/fpubh.2021.694834

Campbell, A. E., Hsiung, C. C.-S., & Blobel, G. A. (2014). Comparative analysis of mitosis-specific antibodies for bulk purification of mitotic populations by fluorescence-activated cell sorting. *BioTechniques, 56*(2), 90--94. https://doi.org/10.2144/000114137

Carrière, V., de Waziers, I., Courtois, Y. A., Leroux, J. P., & Beaune, P. H. (1992). Cytochrome P450 induction and mutagenicity of 2-aminoanthracene (2AA) in rat liver and gut. *Mutation Research, 268*(1), 11--20. https://doi.org/10.1016/0027-5107(92)90077-6

Carroll, C., Warnakulasuriyarachchi, D., Nokhbeh, M., & Lambert, I. (2002). Salmonella typhimurium mutagenicity tester strains that overexpress oxygen-insensitive nitroreductases nfsA and nfsB. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis, 501*(1--2), 79--98. https://doi.org/10.1016/S0027-5107(02)00012-7

Chakravarti, S., Radha, S., & Alla, M. (2024). 'Relevant' versus 'similar' structural analogs to support genotoxicity (Q)SAR predictions. *Multicase*. https://multicase.com/wp-content/uploads/2024/04/GTA_2024_poster_final-1.pdf

Chang, J., & Atterwill, C. K. (Eds.). (1998). *In vitro methods in pharmaceutical research*. Academic Press.

Chang, T.S., Jeong, W., Woo, H.A., Lee, S.M., Park, S. and Rhee, S.G., (2004) Characterization of mammalian sulfiredoxin and its reactivation of hyperoxidized

peroxiredoxin through reduction of cysteine sulfinic acid in the active site to cysteine. vol. 279. *J. Biol. Chem*, pp.50994-51001.

Charrad, M., Ghazzali, N., Boiteau, V., & Niknafs, A. (2014). NbClust: An R package for determining the relevant number of clusters in a data set. *Journal of Statistical Software, 61*(6), 1--36. https://doi.org/10.18637/jss.v061.i06

Chatterjee, N., & Walker, G. C. (2017). Mechanisms of DNA damage, repair, and mutagenesis. *Environmental and Molecular Mutagenesis, 58*(5), 235--263. https://doi.org/10.1002/em.22087

 Chinnasamy, N., Rafferty, J. A., Hickson, I., Ashby, J., Tinwell, H., Margison, G. P., ... Fairbairn, L. J. (1997). O6-benzylguanine potentiates the *in vivo* toxicity and clastogenicity of temozolomide and BCNU in mouse bone marrow. *Blood, 89*(5), 1566--1573.

Chow, C., Ong, K., Dougherty, E., & Simons, S. (2011). Inferring mechanisms from dose--response curves. In *Computer Methods, Part C* (pp. 465--483). Academic Press.

Clifford, H., Wessely, F., & Pendurthi, S. (2011). Comparison of clustering methods for investigation of genome-wide methylation array data. *Frontiers in Genetics, 2*, 88. https://doi.org/10.3389/fgene.2011.00088

Cohen Hubal, E. A., Wetmore, B. A., Wambaugh, J. F., El-Masri, H., Sobus, J. R., & Bahadori, T. (2018). Advancing internal exposure and physiologically-based toxicokinetic modeling for 21st-century risk assessments. *Journal of Exposure Science & Environmental Epidemiology, 29*(1), 11--20. https://doi.org/10.1038/s41370-018-0046-9

Committee on Mutagenicity of Chemicals in Food, Consumer Products and the Environment. (2011). *Guidance on a strategy for genotoxicity testing of chemical substances*. https://www.gov.uk/government/publications/guidance-on-a-strategy-for-genotoxicity-testing-of-chemical-substances

Committee on the Design and Evaluation of Safer Chemical Substitutions: A Framework to Inform Government and Industry Decision. (2014). *Toxicological priority index (ToxPi): A framework to guide selection of chemical alternatives*. National Academies Press. https://www.ncbi.nlm.nih.gov/books/NBK253962/

Crump, K. (2011). Use of dose-response models to estimate thresholds for low-dose extrapolation. *Critical Reviews in Toxicology, 41*(8), 637--650. https://doi.org/10.3109/10408444.2011.574236

Davis, J. A., Gift, J. S., & Zhao, Q. J. (2011). Introduction to benchmark dose methods and U.S. EPA's benchmark dose software (BMDS) version 2.1.1. *Toxicology and Applied Pharmacology, 254*(2), 181--191. https://doi.org/10.1016/j.taap.2011.03.007

Delgado, J. L., Hsieh, C.-M., Chan, N.-L., & Hiasa, H. (2018). Topoisomerases as anticancer targets. *Biochemical Journal, 475*(2), 373--398. https://doi.org/10.1042/BCJ20160583

Dertinger, S. D., Avlasevich, S. L., Torous, D. K., Singh, P., Khanal, S., Kirby, C., Drake, A., MacGregor, J. T., & Bemis, J. C. (2019). 3Rs friendly study designs facilitate rat liver and blood micronucleus assays and pig-a gene mutation assessments: Proof-of-concept with 13 reference chemicals. *Environmental and Molecular Mutagenesis, 60*(8), 704--739. https://doi.org/10.1002/em.22312

Dertinger, S. D., Briggs, E., Hussien, Y., Bryce, S. M., Avlasevich, S. L., Conrad, A., Johnson, G. E., Williams, A., & Bemis, J. C. (2024). Visualization strategies to aid interpretation of high-dimensional genotoxicity data. *Environmental and Molecular Mutagenesis, 65*(5), 156--178. https://doi.org/10.1002/em.22604

Dertinger, S. D., Bryce, S. M., Phonethepswath, S., & Avlasevich, S. L. (2011). When Pigs Fly: Immunomagnetic separation facilitates rapid determination of pig-a mutant frequency by flow cytometric analysis. *Mutation Research/Genetic Toxicology and Environmental Mutagenesis, 721*(2), 163--170. https://doi.org/10.1016/j.mrgentox.2011.01.009

Dertinger, S. D., Phonethepswath, S., Avlasevich, S. L., ... & MacGregor, J. T. (2012). Efficient monitoring of *in vivo* pig-a gene mutation and chromosomal damage: Summary of 7 published studies and results from 11 new reference compounds. *Toxicological Sciences, 130*(2), 328--348. https://doi.org/10.1093/toxsci/kfs250

Dertinger, S. D., Phonethepswath, S., Franklin, D., Wills, J. W., Bryce, S. M., Bemis, J. C., ... & Johnson, G. E. (2019). Predictions of genotoxic potential, mode of action, molecular targets, and potency via a tiered MultiFlow® assay data analysis strategy. *Environmental and Molecular Mutagenesis, 60*(6), 513--533. https://doi.org/10.1002/em.22274

Dickinson, D., Warnes, G., Quievryn, G., Messer, J., Zhitkovich, A., Rubitski, E., & Aubrecht, J. (2004). Differentiation of DNA reactive and non-reactive genotoxic mechanisms using gene expression profile analysis. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis, 549*(1--2), 29--41. https://doi.org/10.1016/j.mrfmmm.2003.11.007

Di Ianni, E., Jacobsen, N. R., Vogel, U. B., & Møller, P. (2022). Systematic review on primary and secondary genotoxicity of carbon black nanoparticles in mammalian cells and animals. *Mutation Research/Reviews in Mutation Research, 790*, 108441. https://doi.org/10.1016/j.mrrev.2022.108441

Doak, S. H., Andreoli, C., Burgum, M. J., Chaudhry, Q., Bleeker, E. A., Bossa, C., Domenech, J., Drobne, D., Fessard, V., Jeliazkova, N., Longhin, E., Rundén-Pran, E., Stępnik, M., El Yamani, N., Catalán, J., & Dusinska, M. (2023). Current status and future challenges of genotoxicity OECD Test Guidelines for nanomaterials: A workshop report. *Mutagenesis, 38*(4), 183--191. https://doi.org/10.1093/mutage/gead017

Dorato, M., & Engelhardt, J. (2005). The no-observed-adverse-effect-level in drug safety evaluations: Use, issues, and definition(s). *Regulatory Toxicology and Pharmacology, 42*(3), 265--274. https://doi.org/10.1016/j.yrtph.2005.05.004

Elespuru, R. K., Doak, S. H., Collins, A. R., Dusinska, M., Pfuhler, S., Manjanatha, M., Cardoso., R, Chen, C. L. (2022). Common considerations for genotoxicity assessment of Nanomaterials. *Frontiers in Toxicology, 4*. https://doi.org/10.3389/ftox.2022.859122

European Food Safety Authority. (2010). Scientific opinion on the use of resorcinol as a food additive. *EFSA Journal, 8*(1), 1411. https://doi.org/10.2903/j.efsa.2010.1411

European Food Safety Authority (EFSA). (2009). Guidance of the scientific committee on use of the benchmark dose approach in risk assessment. *EFSA Journal, 7*(12), 1150. https://doi.org/10.2903/j.efsa.2009.1150

European Food Safety Authority (EFSA). (2017). Update: Use of the benchmark dose approach in risk assessment. *EFSA Journal, 15*(1), e04658. https://doi.org/10.2903/j.efsa.2017.4658

European Food Safety Authority (EFSA). (2022). Guidance on the use of the benchmark dose approach in risk assessment. *EFSA Journal, 20*(10), e07584. https://doi.org/10.2903/j.efsa.2022.7584

European Medicines Agency. (2012). *ICH guidelines S2 (R1) on genotoxicity testing and data interpretation for pharmaceuticals intended for human use*. https://www.ema.europa.eu/en/documents/scientific-guideline/ich-guideline-s2-r1-genotoxicity-testing-data-interpretation-pharmaceuticals-intended-human-use-step_en.pdf

Fenech, M. (2000). The *in vitro* micronucleus technique. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis, 455*(1--2), 81--95. https://doi.org/10.1016/S0027-5107(00)00065-8

Fenech, M. (2007). Cytokinesis-block micronucleus cytome assay. *Nature Protocols, 2*(5), 1084--1104. https://doi.org/10.1038/nprot.2007.77

Fisher, W. D. (1958). On grouping for maximum homogeneity. *Journal of the American Statistical Association, 53*(284), 789--798. https://doi.org/10.2307/2281952

Fortin, A.-M. V., Long, A. S., Williams, A., Meier, M. J., Cox, J., Pinsonnault, C., ... White, P. A. (2023). Application of a new approach methodology (NAM)-based strategy for genotoxicity assessment of data-poor compounds. *Frontiers in Toxicology, 5*, 1098432. https://doi.org/10.3389/ftox.2023.1098432

Friedman, K. P., Gagne, M., Loo, L. H., Karamertzanis, P., Netzeva, T., Sobanski, T., et al. (2020). Utility of *in vitro* bioactivity as a lower bound estimate of *in vivo* adverse effect levels and in risk-based prioritization. *Toxicological Sciences, 173*(2), 202--225. https://doi.org/10.1093/toxsci/kfz201

Futami, T., Miyagishi, M., & Taira, K. (2005). Identification of a network involved in thapsigargin-induced apoptosis using a library of small interfering RNA expression vectors. *Journal of Biological Chemistry, 280*(1), 826--831. https://doi.org/10.1074/jbc.M411123200

Galloway, S. (2000). Cytotoxicity and chromosome aberrations *in vitro*: Experience in industry and the case for an upper limit on toxicity in the aberration assay. *Environmental and Molecular Mutagenesis, 35*(3), 191–201. https://doi.org/10.1002/1098-2280(2000)35:3

Gao, C. X., Wang, Y., & Zhang, L. (2023). An overview of clustering methods with guidelines for application in mental health research. *Psychiatry Research, 327*, 115265. https://doi.org/10.1016/j.psychres.2023.115265

Geiss, G. K., Bumgarner, R. E., Birditt, B., Dahl, T., Dowidar, N., Dunaway, D. L., ... Dimitrov, K. (2008). Direct multiplexed measurement of gene expression with color-coded probe pairs. *Nature Biotechnology, 26*(3), 317--325. https://doi.org/10.1038/nbt1385

Gocke, E., Bürgin, H., Müller, L., & Pfister, T. (2009). Literature review on the genotoxicity, reproductive toxicity, and carcinogenicity of ethyl methanesulfonate. *Toxicology Letters, 190*(3), 254--265. https://doi.org/10.1016/j.toxlet.2009.03.019

Goldsmith, J. R., & Kordysh, E. (1993). Why dose-response relationships are often non-linear and some consequences. *Journal of Exposure Analysis and Environmental Epidemiology, 3*(3), 259--276. https://doi.org/10.1038/sj.jea.7500285

Gollapudi, P., Hasegawa, L. S., & Eastmond, D. A. (2014). A comparative study of the aneugenic and polyploidy-inducing effects of fisetin and two model Aurora kinase inhibitors. *Mutation Research/Genetic Toxicology and Environmental Mutagenesis, 767*, 37--43. https://doi.org/10.1016/j.mrgentox.2014.03.004

Grimm, D. (2019). US EPA to eliminate all mammal testing by 2035. *Science, 365*(6459), 1231. https://doi.org/10.1126/science.365.6459.1231

Guha, P., Kaptan, E., Gade, P., Kalvakolanu, D. V., & Ahmed, H. (2017). Tunicamycin induced endoplasmic reticulum stress promotes apoptosis of prostate cancer cells by activating mTORC1. *Oncotarget, 8*(40), 68191--68207. https://doi.org/10.18632/oncotarget.19277

Hakem, R. (2008). DNA-damage repair; the good, the bad, and the ugly. *The EMBO Journal, 27*(4), 589--605. https://doi.org/10.1038/sj.emboj.7601973

Harrington, E. A., Bebbington, D., Moore, J., … & Kramer, R. A. (2004). VX-680, a potent and selective small-molecule inhibitor of the Aurora kinases, suppresses tumor growth *in vivo*. *Nature Medicine, 10*(3), 262–267. https://doi.org/10.1038/nm1003

Hasan, M. N., Begum, K., & Rahman, M. S. (2019). Assessment of drugs toxicity and associated biomarker genes using hierarchical clustering. *Medicina, 55*(8), 451. https://doi.org/10.3390/medicina55080451

Hartung, T. (2009). Toxicology for the twenty-first century. *Nature, 460*(7252), 208--212. https://doi.org/10.1038/460208a

Hassenzahl, D.M. and Finkel, A.M. (2008) 'Risk assessment, environmental/occupational', *International Encyclopedia of Public Health*, pp. 590--600. https://doi.org/10.1016/b978-012373960-5.00286-0

Health Canada. (2021). *Science approach document - Bioactivity exposure ratio: Application in priority setting and risk assessment*.

Henderson, L., Albertini, S., & Aardema, M. (2000). Thresholds in genotoxicity responses. *Mutation Research/Genetic Toxicology and Environmental Mutagenesis, 464*(1), 123--128. https://doi.org/10.1016/S1383-5718(99)00177-3

Hendriks, G., Atallah, M., Morolli, B., Calléja, F., Ras-Verloop, N., Huijskens, I., ... Vrieling, H. (2011). The ToxTracker Assay: Novel GFP reporter systems that provide mechanistic insight into the genotoxic properties of chemicals. *Toxicological Sciences, 125*(1), 285--298. https://doi.org/10.1093/toxsci/kfr281

Hendriks, G., Atallah, M., Morolli, B., Calléja, F. M., Ras-Verloop, N., Huijskens, I., ... Vrieling, H. (2024). Interlaboratory validation of the ToxTracker assay: An *in vitro* reporter assay for mechanistic genotoxicity assessment. *Environmental and Molecular Mutagenesis, 65*(1--2), 4--24. https://doi.org/10.1002/em.22592

Hendriks, G., Derr, R. S., Misovic, B., Morolli, B., Calléja, F. M., & Vrieling, H. (2016). The extended ToxTracker assay discriminates between induction of DNA damage, oxidative stress, and protein misfolding. *Toxicological Sciences, 150*(1), 190--203. https://doi.org/10.1093/toxsci/kfv323

Herrmann, K., Pistollato, F., & Stephens, M. L. (2020). (Q)SAR tools for the prediction of mutagenic properties: Are they ready for application in pesticide regulation? *Pest Management Science, 76*(10), 3316--3325. https://doi.org/10.1002/ps.5828

Hirano, S., Sun, X., & Tsumoto, S. (2004). Comparison of clustering methods for clinical databases. *Information Sciences, 159*(3--4), 155--165. https://doi.org/10.1016/j.ins.2003.08.004

Huang, R. and Zhou, P.-K. (2021) 'DNA damage repair: Historical perspectives, mechanistic pathways and clinical translation for targeted cancer therapy', *Signal Transduction and Targeted Therapy, 6*(1). https://doi.org/10.1038/s41392-021-00648-7

International Agency for Research on Cancer (IARC). (1999). *Monographs on the evaluation of carcinogenic risks to humans: 1,2-dimethylhydrazine* (Vol. 71). IARC Press.

Ito, H., Morishita, R., & Nagata, K. (2018). Functions of rhotekin, an effector of Rho GTPase, and its binding partners in mammals. *International Journal of Molecular Sciences, 19*(7), 2121. https://doi.org/10.3390/ijms19072121

Jackson, S. P., & Bartek, J. (2009). The DNA-damage response in human biology and disease. *Nature, 461*(7267), 1071--1078. https://doi.org/10.1038/nature08467

Jani, J. P., Arcari, J., Bernardo, V., ... & Kramer, R. A. (2010). PF-03814735, an orally bioavailable small molecule aurora kinase inhibitor for cancer therapy. *Molecular Cancer Therapeutics, 9*(4), 883--894. https://doi.org/10.1158/1535-7163.MCT-09-0915

Jenkins, G., Zaïr, Z., Johnson, G., & Doak, S. (2010). Genotoxic thresholds, DNA repair, and susceptibility in human populations. *Toxicology, 278*(3), 305–310. https://doi.org/10.1016/j.tox.2010.05.004

Jenks, G. F. (1977). *Optimal data classification for choropleth maps*. University of Kansas, Department of Geography.

Kavlock, R. J., Bahadori, T., Barton-Maclaren, T. S., Gwinn, M. R., Rasenberg, M., & Thomas, R. S. (2018). Accelerating the pace of chemical risk assessment. *Chemical Research in Toxicology, 31*(4), 287–290. https://doi.org/10.1021/acs.chemrestox.8b00039

Keshava, C., Keshava, N., Whong, W. Z., Nath, J., & Ong, T. M. (1998). Inhibition of methotrexate-induced chromosomal damage by folinic acid in V79 cells. *Mutation Research, 397*(2), 221–228. https://doi.org/10.1016/S0027-5107(97)00222-2

Kim, H., Park, Y., Shin, D., Oh, S., & Chung, K. (2013). Appropriate *in vitro* methods for genotoxicity testing of silver nanoparticles. *Environmental Health and Toxicology, 28*, e2013003. https://doi.org/10.5620/eht.2013.28.e2013003

Kirkland, D., Aardema, M., Henderson, L., & Müller, L. (2005). Evaluation of the ability of a battery of three *in vitro* genotoxicity tests to discriminate rodent carcinogens and non-carcinogens. *Mutation Research/Genetic Toxicology and Environmental Mutagenesis*, *584*(1–2), 1–256. https://doi.org/10.1016/j.mrgentox.2005.02.004

Kirkland, D., Pfuhler, S., Tweats, D., Aardema, M., Corvi, R., Darroudi, F., Elhajouji, A., Glatt, H., Hastwell, P., Hayashi, M., Kasper, P., Kirchner, S., Lynch, A., Marzin, D., Maurici, D., Meunier, J. R., Müller, L., Nohynek, G., Parry, J., Parry, E., Thybaud, V., Tice, R., van Benthem, J., Vanparys, P., & White, P. (2007). How to reduce false positive results when undertaking *in vitro* genotoxicity testing and thus avoid unnecessary follow-up animal tests: Report of an ECVAM workshop. *Mutation Research, 628*(1), 31

Kirkland, D., Pfuhler, S., Tweats, D., Aardema, M., Corvi, R., Darroudi, F., … & Müller, L. (2007). How to reduce false positive results when undertaking *in vitro* genotoxicity testing and thus avoid unnecessary follow-up animal tests: Report of an ECVAM workshop. *Mutation Research/Genetic Toxicology and Environmental Mutagenesis, 628*(1), 31–55. https://doi.org/10.1016/j.mrgentox.2006.11.008

Kirkland, D., Kasper, P., Müller, L., Corvi, R., & Speit, G. (2008). Recommended lists of genotoxic and non-genotoxic chemicals for assessment of the performance of new or improved genotoxicity tests: A follow-up to an ECVAM workshop. *Mutation Research, 653*(1–2), 99–108. https://doi.org/10.1016/j.mrgentox.2008.03.008

Kirkland, D., Reeve, L., Gatehouse, D., & Vanparys, P. (2011). A core *in vitro* genotoxicity battery comprising the Ames test plus the *in vitro* micronucleus test is sufficient to detect rodent carcinogens and *in vivo* genotoxins. *Mutation Research, 721*(1), 27–73. https://doi.org/10.1016/j.mrgentox.2011.01.004

Kirkland, D., Kasper, P., Martus, H.-J., Müller, L., van Benthem, J., Madia, F., & Corvi, R. (2016). Updated recommended lists of genotoxic and non-genotoxic chemicals for

assessment of the performance of new or improved genotoxicity tests. *Mutation Research, 795*, 7–30. https://doi.org/10.1016/j.mrgentox.2016.03.007

Kirsch-Volders, M., Aardema, M., & Elhajouji, A. (2000). Concepts of threshold in mutagenesis and carcinogenesis. *Mutation Research, 464*(1), 3–11. https://doi.org/10.1016/S1383-5742(99)00059-1

Kirsch-Volders, M., Decordier, I., Elhajouji, A., Plas, G., Aardema, M., & Fenech, M. (2010). *In vitro* genotoxicity testing using the micronucleus assay in cell lines, human lymphocytes and 3D human skin models. *Mutagenesis, 26*(1), 177–184. https://doi.org/10.1093/mutage/geq068

Klein, C. B., & King, A. A. (2007). Genistein genotoxicity: Critical consideration of *in vitro* exposure dose. *Toxicology and Applied Pharmacology, 224*(1), 1–11. https://doi.org/10.1016/j.taap.2007.06.022

Kobets, T., Iatropoulos, M., & Williams, G. (2019). Mechanisms of DNA-reactive and epigenetic chemical carcinogens: Applications to carcinogenicity testing and risk assessment. *Toxicology Research, 8*(2), 123–145. https://doi.org/10.1039/C8TX00250A

Krais, A. M., Speksnijder, E. N., Melis, J. P. M., Singh, R., Caldwell, A., de Costa, C. G., … Arlt, V. M. (2016). Metabolic activation of 2-amino-1-methyl-6-phenylimidazo[4,5-b]pyridine and DNA adduct formation depends on p53: Studies in Trp53(+/+), Trp53(+/−), and Trp53(−/−) mice. *Molecular Cancer Biology, 138*(5), 976–982. https://doi.org/10.1158/1541-7786.MCR-15-0453

Krishna, G., Urda, G., Tefera, W., Lalwani, N. D., & Theiss, J. (1995). Simultaneous evaluation of dexamethasone-induced apoptosis and micronuclei in rat primary spleen cell cultures. *Mutation Research, 332*(1–2), 1–8. https://doi.org/10.1016/0027-5107(95)00034-8

Lambert, I. B., Singer, T. M., Boucher, S. E., & Douglas, G. R. (2005). Detailed review of transgenic rodent mutation assays. *Mutation Research/Reviews in Mutation Research*, *590*(1–3), 1–280. doi:10.1016/j.mrrev.2005.04.002

Larsen, A. K., Escargueil, A. E., & Skladanowski, A. (2003). Catalytic topoisomerase II inhibitors in cancer therapy. *Pharmacology & Therapeutics, 99*(2), 167–181. https://doi.org/10.1016/S0163-7258(03)00058-5

Lee, F. Y., Borzilleri, R., Fairchild, C. R., Kim, S. H., Long, B. H., Reventos-Suarez, C., … Kramer, R. A. (2001). BMS-247550: A novel epothilone analog with a mode of action similar to paclitaxel but possessing superior antitumor efficacy. *Clinical Cancer Research, 7*(5), 1429–1437.

Leist, M., Hasiwa, N., Daneshian, M., & Hartung, T. (2012). Validation and quality control of replacement alternatives – current status and future challenges. *Toxicology Research, 1*(1), 8–22. https://doi.org/10.1039/c2tx20011b

Li, H., Hyduke, D. R., Chen, R., Heard, P., Yauk, C. L., Aubrecht, J., & Fornace, A. J. (2015). Development of a toxicogenomics signature for genotoxicity using a dose-optimization and

informatics strategy in human cells. *Environmental and Molecular Mutagenesis*, *56*(6), 505–519. doi:10.1002/em.21941

Li, H.-H., Chen, R., Hyduke, D. R., Williams, A., Frötschl, R., Ellinger-Ziegelbauer, H., … Fornace, A. J. (2017). Development and validation of a high-throughput transcriptomic biomarker to address 21st century genetic toxicology needs. *Proceedings of the National Academy of Sciences*, *114*(51). □ Redon, C. E., Nakamura, A. J., Martin, O. A., Parekh, P. R., Weyemi, U. S., & Bonner, W. M. (2011). Recent developments in the use of γ -H2AX as a quantitative DNA double-strand break biomarker. *Aging*, *3*(2), 168–174. doi:10.18632/aging.100284

Li, H.-H., Yauk, C. L., Chen, R., Hyduke, D. R., Williams, A., Frötschl, R., … Fornace, A. J. (2019). TGX-DDI, a transcriptomic biomarker for genotoxicity hazard assessment of Pharmaceuticals and Environmental Chemicals. *Frontiers in Big Data*, *2*. doi:10.3389/fdata.2019.00036

Li, X., Li, Y., Ning, K. G., Chen, S., Guo, L., Bonzo, J. A., & Mei, N. (2022). The expression of phase II drug-metabolizing enzymes in human B-lymphoblastoid TK6 cells. *Journal of Environmental Science and Health, Part C, 40*(1), 106–118. https://doi.org/10.1080/26896583.2022.2044242

Litron Laboratories. (2017). *In vivo MicroFlow® PLUS-R kit instruction manual v170503*. Rochester, NY, USA. https://litronlabs.com/Products/In-Vivo-Micronucleus/Kits/Plus/MicroFlow-PLUS-R-Kit

Litron Laboratories. (2018). *MicroFlow® in vitro kit instruction manual, v180208*. Rochester, NY, USA. https://litronlabs.com/Products/In-Vitro-Micronucleus/Kits/MicroFlow-In-Vitro-2,000-400-Kit

Litron Laboratories. (2018). *MultiFlow® DNA damage kit instruction manual*. Rochester, NY, USA. https://litronlabs.com/Products/In-Vitro-MultiFlow/Kits/DNA-Damage-Kits/MultiFlow-DNA-Damage-Kit-%E2%80%93-p53,-%CE%B3H2AX,-Phospho-(1)

Litron Laboratories. (2018). *Microflow® In vitro Kit Instruction Manual*. Rochester, NY, USA. https://litronlabs.com/Products/In-Vitro-Micronucleus/Kits/MicroFlow-In-Vitro-2,000-400-Kit

Litron Laboratories. (2021). *Prototype in vivo MicroFlow® PLUS-RL kit instruction manual v211215*. Rochester, NY, USA. https://litronlabs.com/Products/In-Vivo-Liver-Micronucleus/Kits/Plus/MicroFlow-PLUS-RL-Kit

Litron Laboratories. (2022). *Rat in vivo MutaFlow® instruction manual v220922*. Rochester, NY, USA. https://litronlabs.com/Products/In-Vivo-Pig-a-Gene-Mutation/Kits/Basic/MutaFlow-Rat-Blood-Freezing-Kit

Litron Laboratories. (2022). *MutaFlow® Mouse or Rat Blood Thaw and Analysis Kit instruction manual v220726*. https://litronlabs.com/Products/In-Vivo-Pig-a-Gene-Mutation/Kits/Plus/MutaFlow-Rat-Blood-Thawing-and-Analysis-Kit

Lovell, D. P. (2000). Dose-response and threshold-mediated mechanisms in mutagenesis: Statistical models and study design. *Mutation Research, 464*(1), 87–95. https://doi.org/10.1016/S1383-5742(99)00062-1

Macgregor, J. T., Frötschl, R., White, P. A., Crump, K. S., Eastmond, D. A., Fukushima, S., Guérard, M., Hayashi, M., Soeteman-Hernández, L. G., & Thybaud, V. (2015). IWGT report on quantitative approaches to genotoxicity risk assessment I. Methods and metrics for defining exposure–response relationships and points of departure (PoDs). *Mutation Research/Genetic Toxicology and Environmental Mutagenesis, 783*, 55–65. https://doi.org/10.1016/j.mrgentox.2015.06.007

Mahadevan, B., Snyder, R. D., Waters, M. D., Benz, R. D., Kemper, R. A., Tice, R. R., & Richard, A. M. (2011). Genetic toxicology in the 21st Century: Reflections and future directions. *Environmental and Molecular Mutagenesis*, *52*(5), 339–354. https://doi.org/10.1002/em.20653

Markowitz, V. M., Campbell, J., Chen, L.-M. A., Kosky, A., Palaniappan, K., & Topaloglou, T. (2003). Integration challenges in gene expression data management. In *Bioinformatics* (pp. 277–301). Academic Press. https://doi.org/10.1016/B978-155860829-0/50012-7

Marvel, S., To, K., Grimm, F., Wright, F., Rusyn, I., & Reif, D. (2018). ToxPi Graphical User Interface 2.0: Dynamic exploration, visualization, and sharing of integrated data models. *BMC Bioinformatics, 19*(1), 318. https://doi.org/10.1186/s12859-018-2330-z

Maugeri, A., Barchitta, M., Agodi, A., Basile, G., & Agodi, A. (2021). Applying a hierarchical clustering on principal components approach to identify different patterns of the SARS-CoV-2 epidemic across Italian regions. *Scientific Reports, 11*(1), 13187. https://doi.org/10.1038/s41598-021-86703-3

McQuitty, L. L. (1966). Similarity analysis by reciprocal pairs for discrete and continuous data. *Educational and Psychological Measurement, 26*(4), 825–831. https://doi.org/10.1177/001316446602600402

Meek, D. W. (2009). Tumour suppression by p53: A role for the DNA damage response? *Nature Reviews Cancer, 9*(10), 714–723. https://doi.org/10.1038/nrc2716

Menz, J., Götz, M. E., Gündel, U., Gürtler, R., Herrmann, K., Hessel-Pras, S., Kneuer, C., Kolrep, F., Nitzsche, D., Pabel, U., Sachse, B., Schmeisser, S., Schumacher, D. M., Schwerdtle, T., Tralau, T., Zellmer, S., & Schäfer, B. (2023). Genotoxicity assessment: Opportunities, challenges and perspectives for quantitative evaluations of dose–response data. *Archives of Toxicology, 97*(9), 2303–2328. https://doi.org/10.1007/s00204-023-03553-w

Miltenyi Biotec & Toxys (2019). High-thoroughput ToxTracker ACE genotoxicity assay on the MACSQuant X flow cytometer. https://toxys.com/wp-content/uploads/2019/10/AppNote_ToxTracker_Genotoxicity_MACSQuantX.pdf

Mondal, M., Gabriels, J., McGinnis, C., Magnifico, M., Marsilje, T., Urban, L., Collis, A., Bojanic, D., Biller, S., Frieauff, W., Martus, H., Suter, W., & Bentley, P. (2010). High-content micronucleus assay in genotoxicity profiling: Initial-stage development and some

applications in the investigative/lead-finding studies in drug discovery. *Toxicological Sciences, 118*(1), 71–85. https://doi.org/10.1093/toxsci/kfq181

Moon, J. L., Kim, S. Y., Shin, S. W., & Park, J.-W. (2012). Regulation of brefeldin A-induced ER stress and apoptosis by mitochondrial NADP+-dependent isocitrate dehydrogenase. *Biochemical and Biophysical Research Communications, 417*(4), 760–764. https://doi.org/10.1016/j.bbrc.2011.12.034

Møller, P., Danielsen, P. H., Karottki, D. G., Jantzen, K., Roursgaard, M., Klingberg, H., … Loft, S. (2014). Oxidative stress and inflammation generated DNA damage by exposure to air pollution particles. *Mutation Research/Reviews in Mutation Research*, *762*, 133–166. doi:10.1016/j.mrrev.2014.09.001

More, S. J., Bampidis, V., Bragard, C., Halldorsson, T. I., Hernández-Jerez, A. F., Hougaard Bennekou, S., … Benford, D. (2021). Guidance on aneugenicity assessment. *EFSA Journal*, *19*(8). doi:10.2903/j.efsa.2021.6770

Murtagh, F., & Legendre, P. (2014). Ward's hierarchical agglomerative clustering method: Which algorithms implement Ward's criterion? *Journal of Classification, 31*(3), 274–295. https://doi.org/10.1007/s00357-014-9161-z

Nagy, E., Johansson, C., Zeisig, M., & Möller, L. (2005). Oxidative stress and DNA damage caused by the urban air pollutant 3-NBA and its isomer 2-NBA in human lung cells analyzed with three independent methods. *Journal of Chromatography B, 827*(1), 94–103. https://doi.org/10.1016/j.jchromb.2005.03.041

Neubig, R., Spedding, M., Kenakin, T., & Christopoulos, A. (2003). International Union of Pharmacology Committee on Receptor Nomenclature and Drug Classification. XXXVIII. Update on terms and symbols in quantitative pharmacology. *Pharmacological Reviews, 55*(4), 597–606. https://doi.org/10.1124/pr.55.4.4

Nitiss, J. L. (2009). DNA topoisomerase II and its growing repertoire of biological functions. *Nature Reviews Cancer, 9*(5), 327–337. https://doi.org/10.1038/nrc2608

Nohmi, T. (2018). Thresholds of genotoxic and non-genotoxic carcinogens. *Toxicological Research, 34*(4), 281–290. https://doi.org/10.5487/TR.2018.34.4.281

Ochoa, R. (2018). Design of studies and risk management in toxicologic pathology. In *Fundamentals of Toxicologic Pathology* (pp. 105–122). Academic Press. https://doi.org/10.1016/b978-0-12-809841-7.00007-1

Organisation for Economic Co-operation and Development. (2016). *Test No. 487: In vitro mammalian cell micronucleus test*. OECD Guidelines for the Testing of Chemicals. https://doi.org/10.1787/9789264264861-en

Organisation for Economic Co-operation and Development. (2018). *Guidance document on good in vitro method practices (GIVIMP)*. OECD Series on Testing and Assessment, No. 286. https://doi.org/10.1787/9789264304796-en

Organisation for Economic Co-operation and Development. (2025). *OECD Test No. 488: Transgenic rodent somatic and germ cell gene mutation assays. OECD Guidelines for the Testing of Chemicals, Section 4. https://doi.org/10.1787/9789264203907-en*

Otteneder, M., & Lutz, W. K. (1999). Correlation of DNA adduct level with tumor incidence: Carcinogenic potency of DNA adducts. *Mutation Research, 424*(1–2), 237–247. https://doi.org/10.1016/S0027-5107(99)00022-7

Parry, E. M., Parry, J. M., Corso, C., Doherty, A., Haddad, F., Hermine, T. F., … & Kirsch-Volders, M. (2002). Detection and characterization of mechanisms of action of aneugenic chemicals. *Mutagenesis, 17*(6), 509–521. https://doi.org/10.1093/mutage/17.6.509

Payton, M., Bush, T. L., Chung, G., … & Kramer, R. A. (2010). Preclinical evaluation of AMG 900, a novel potent and highly selective pan-aurora kinase inhibitor with activity in taxane-resistant tumor cell lines. *Cancer Research, 70*(24), 9846–9854. https://doi.org/10.1158/0008-5472.CAN-10-3001

Pearce, R. G., Setzer, R. W., Strope, C. L., Sipes, N. S., & Wambaugh, J. F. (2017). httk: R package for high-throughput toxicokinetics. *Journal of Statistical Software, 79*(4). https://doi.org/10.18637/jss.v079.i04

Pramod, K., Ansari, S. H., & Ali, J. (2010). Eugenol: A natural compound with versatile pharmacological actions. *Natural Product Communications, 5*(12), 1999–2006. https://doi.org/10.1177/1934578X1000501236

Prinz, H. (2009). Hill coefficients, dose–response curves and allosteric mechanisms. *Journal of Chemical Biology, 3*(1), 37–44. https://doi.org/10.1007/s12154-009-0029-3

RIVM. (2021). *PROAST*. The Dutch National Institute for Public Health and the Environment. Accessed December 1, 2021. https://www.rivm.nl/en/proast

Reif, D., Sypa, M., Lock, E., Wright, F., Wilson, A., Cathey, T., Judson, R., & Rusyn, I. (2012). ToxPi GUI: An interactive visualization tool for transparent integration of data from diverse sources of evidence. *Bioinformatics, 29*(3), 402–403. https://doi.org/10.1093/bioinformatics/bts686

Reinhardt, H. C., & Yaffe, M. B. (2009). Kinases that control the cell cycle in response to DNA damage: Chk1, CHK2, and MK2. *Current Opinion in Cell Biology*, *21*(2), 245–255. doi:10.1016/j.ceb.2009.01.018

Ren, N., Atyah, M., Chen, W.-Y., & Zhou, C.-H. (2017). The various aspects of genetic and epigenetic toxicology: Testing methods and clinical applications. *Journal of Translational Medicine, 15*(1). https://doi.org/10.1186/s12967-017-1218-4

Rezvanfar, M.A. (2014) 'Benchmark dose', *Encyclopedia of Toxicology*, pp. 402–406. doi:10.1016/b978-0-12-386454-3.00588-1.

Ricci, P. F., Straja, S. R., & Cox, A. L. (2011). Changing the risk paradigms can be good for our health: J-shaped, linear and threshold dose-response models. *Dose-Response, 10*(2), 178–188. https://doi.org/10.2203/dose-response.11-020.ricci

Ritter, C., Culp, S., Freeman, J., Marques, M., Beland, F., & Malejka-Giganti, D. (2002). DNA adducts from nitroreduction of 2,7-dinitrofluorene, a mammary gland carcinogen, catalyzed by rat liver or mammary gland cytosol. *Chemical Research in Toxicology, 15*(4), 536–544. https://doi.org/10.1021/tx010178a

Rodriguez-Antona, C., & Ingelman-Sundberg, M. (2006). Cytochrome P450 pharmacogenetics and cancer. *Oncogene, 25*(11), 1679–1691. https://doi.org/10.1038/sj.onc.1209377

Rotroff, D. M., Wetmore, B. A., Dix, D. J., Ferguson, S. S., Clewell, H. J., Houck, K. A., LeCluyse, E. L., Andersen, M. E., Judson, R. S., Smith, C. M., Sochaski, M. A., Kavlock, R. J., Boellmann, F., Martin, M. T., Reif, D. M., Wambaugh, J. F., & Thomas, R. S. (2010). Incorporating human dosimetry and exposure into high-throughput *in vitro* toxicity screening. *Toxicological Sciences, 117*(2), 348–358. https://doi.org/10.1093/toxsci/kfq220

Rotroff, D. M., Wetmore, B. A., Dix, D. J., Ferguson, S. S., Clewell, H. J., Houck, K. A., … Thomas, R. S. (2010). Incorporating human dosimetry and exposure into high-throughput *in vitro* toxicity screening. *Toxicological Sciences, 117*(2), 348–358. https://doi.org/10.1093/toxsci/kfq220

Russell, W. M. S., & Burch, R. L. (1959). *The principles of humane experimental technique*. Methuen and Co.

Sasaki, J. C., Allemang, A., Bryce, S. M., Custer, L., Dearfield, K. L., Dietz, Y., Elhajouji, A., Escobar, P. A., Fornace, A. J., Jr., Froetschl, R., Galloway, S., Hemmann, U., Hendriks, G., Li, H. H., Luijten, M., Ouedraogo, G., Peel, L., Pfuhler, S., Roberts, D. J., Thybaud, V., van Benthem, J., Yauk, C. L., & Schuler, M. (2019). Application of the adverse outcome pathway framework to genotoxic modes of action. *Environmental and Molecular Mutagenesis, 61*(1), 114–134. https://doi.org/10.1002/em.22339

Sewell, F., Edwards, J., Prior, H., & Robinson, S. (2016). Opportunities to apply the 3Rs in safety assessment programs. *ILAR Journal, 57*(2), 234–245. https://doi.org/10.1093/ilar/ilw024

Slob, W. (2002). Dose-response modeling of continuous endpoints. *Toxicological Sciences, 66*(2), 298–312. https://doi.org/10.1093/toxsci/66.2.298

Slob, W., & Setzer, R. W. (2014). Shape and steepness of toxicological dose–response relationships of continuous endpoints. *Critical Reviews in Toxicology, 44*(3), 270–297. https://doi.org/10.3109/10408444.2013.853726

Smart, D. J., Helbling, F. R., Verardo, M., Huber, A., McHugh, D., & Vanscheeuwijck, P. (2020). Development of an integrated assay in human TK6 cells to permit comprehensive genotoxicity analysis *in vitro*. *Mutation Research/Genetic Toxicology and Environmental Mutagenesis, 849*, 503129. https://doi.org/10.1016/j.mrgentox.2020.503129

Snyder, R., & Green, J. (2001). A review of the genotoxicity of marketed pharmaceuticals. *Mutation Research/Reviews in Mutation Research, 488*(2), 151–169. https://doi.org/10.1016/S1383-5742(01)00055-2

Soeteman-Hernández, L. G., Fellows, M. D., Johnson, G. E., & Slob, W. (2015). Correlation of *in vivo* versus *in vitro* benchmark doses (BMDs) derived from micronucleus test data: A proof of concept study. *Toxicological Sciences, 148*(2), 355–367. https://doi.org/10.1093/toxsci/kfv197

Sommer, S., Buraczewska, I., & Kruszewski, M. (2020). Micronucleus assay: The state of art, and future directions. *International Journal of Molecular Sciences, 21*(4), 1534. https://doi.org/10.3390/ijms21041534

Song, P., Wu, L., & Guan, W. (2015). Dietary nitrates, nitrites, and nitrosamines intake and the risk of gastric cancer: A meta-analysis. *Nutrients, 7*(12), 9872–9895. https://doi.org/10.3390/nu7125505

Stine, K. E., & Brown, T. M. (2015). *Principles of toxicology: Environmental and industrial applications* (3rd ed.). Wiley.

Sturla, S. (2018). Point of departure. *Chemical Research in Toxicology, 31*(1), 2–3. https://doi.org/10.1021/acs.chemrestox.7b00303

Sun, X., Bryce, S. M., Bemis, J. C., & Dertinger, S. D. (2022). A new imaging platform (iScreen) allows for the concurrent assessment of micronucleus induction and genotoxic mode of action in human A375 cells. *Environmental and Molecular Mutagenesis, 63*(5), 230–245. https://doi.org/10.1002/em.22496

Svingen, T., Villeneuve, D. L., Knapen, D., Panagiotou, E. M., Draskau, M. K., Damdimopoulou, P., & O'Brien, J. M. (2021). A pragmatic approach to adverse outcome pathway development and evaluation. *Toxicological Sciences, 184*(2), 183–190. https://doi.org/10.1093/toxsci/kfab113

Terradas, M., Martín, M., & Genescà, A. (2016). Impaired nuclear functions in micronuclei results in genome instability and chromothripsis. *Archives of Toxicology, 90*(11), 2657–2667. https://doi.org/10.1007/s00204-016-1818-4

Tian, S., Cyr, A., Zeise, K., Bryce, S. M., Hall, N., Bemis, J. C., & Dertinger, S. D. (2020). 3Rs-friendly approach to exogenous metabolic activation that supports high-throughput genetic toxicology testing. *Environmental and Molecular Mutagenesis, 61*(5), 408–432. https://doi.org/10.1002/em.22363

Toxys ToxTracker Academy. (2024). *ToxTracker Academy*. https://toxtrackeracademy.com/science/

Tubbs, A., & Nussenzweig, A. (2017). Endogenous DNA damage as a source of genomic instability in cancer. *Cell, 168*(4), 644–656. https://doi.org/10.1016/j.cell.2017.01.002

U.S. Department of Health and Human Services, Food and Drug Administration. (2002). *Guidance for industry and reviewers (draft): Estimating the safe starting dose in clinical trials for therapeutics in adult healthy volunteers*. https://www.fda.gov/media/72309/download

U.S. Environmental Protection Agency. (2023). *Guidance for In vitro to In vivo Extrapolation (IVIVE) in Toxicity Risk Assessment*. Retrieved from https://www.epa.gov

U.S. Food and Drug Administration. (2006). *Pharmacology review of Dasatinib (Sprycel) NDA 21-986*. Center for Drug Evaluation and Research. Retrieved from https://www.accessdata.fda.gov

Uno, Y., Morita, T., Luijten, M., Beevers, C., Hamada, S., Itoh, S., Ohyama, W., & Takasawa, H. (2015). Recommended protocols for the liver micronucleus test: Report of the IWGT working group. *Mutation Research/Genetic Toxicology and Environmental Mutagenesis, 783*, 13–18. https://doi.org/10.1016/j.mrgentox.2015.04.008

Wambaugh, J. F., Wetmore, B. A., Pearce, R., Strope, C., Goldsmith, R., Sluka, J. P., Sedykh, A., Tropsha, A., Bosgra, S., Shah, I., Judson, R., Thomas, R. S., & Setzer, R. W. (2015). Toxicokinetic triage for environmental chemicals. *Toxicological Sciences, 147*(1), 55–67. https://doi.org/10.1093/toxsci/kfv118

Ward, J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association, 58*(301), 236–244. https://doi.org/10.2307/2282967

Ware, R. E., & Dertinger, S. D. (2021). Absence of hydroxyurea-induced mutational effects supports higher utilization for the treatment of sickle cell anaemia. *British Journal of Haematology, 194*(2), 252–266. https://doi.org/10.1111/bjh.17522

Weiss, J. N. (1997). The Hill equation revisited: Uses and misuses. *FASEB Journal, 11*(11), 835–841. https://doi.org/10.1096/fasebj.11.11.9285481

Wetmore, B. A., Wambaugh, J. F., Ferguson, S. S., Sochaski, M. A., Rotroff, D. M., Freeman, K., … Andersen, M. E. (2012). Integration of dosimetry, exposure, and high-throughput screening data in chemical toxicity assessment. *Toxicological Sciences, 125*(1), 157–174. https://doi.org/10.1093/toxsci/kfr254

Wheeldon, R., Bernacki, D., Dertinger, S., Bryce, S., Bemis, J., & Johnson, G. (2020). Benchmark dose analysis of DNA damage biomarker responses provides compound potency and adverse outcome pathway information for the topoisomerase II inhibitor class of compounds. *Environmental and Molecular Mutagenesis, 61*(4), 396–407. https://doi.org/10.1002/em.22358

Wheeldon, R. P., Dertinger, S. D., Bryce, S. M., Bemis, J. C., & Johnson, G. E. (2021). The use of benchmark dose uncertainty measurements for robust comparative potency analyses. *Environmental and Molecular Mutagenesis, 62*(3), 203–215. https://doi.org/10.1002/em.22422

Wheeldon, R. P., Dertinger, S. D., Bryce, S. M., Bemis, J. C., & Johnson, G. E. (2021). The use of benchmark dose uncertainty measurements for robust comparative potency analyses. *Environmental and Molecular Mutagenesis, 62*(3), 203–215. https://doi.org/10.1002/em.22422

Wilde, S., Queisser, N., Holz, C., Raschke, M., & Sutter, A. (2018). Differentiation of aneugens and clastogens in the *in vitro* micronucleus test by kinetochore scoring using automated image analysis. *Environmental and Molecular Mutagenesis*, *60*(3), 227–242. doi:10.1002/em.22259

Williams, G. (1981). Liver carcinogenesis: The role for some chemicals of an epigenetic mechanism of liver-tumour promotion involving modification of the cell membrane. *Food and Cosmetics Toxicology, 19*, 577–583. https://doi.org/10.1016/0015-6264(81)90510-2

Williams, G., & Numoto, S. (1984). Promotion of mouse liver neoplasms by the organochlorine pesticides chlordane and heptachlor in comparison to dichlorodiphenyltrichloroethane. *Carcinogenesis, 5*(12), 1689–1696. https://doi.org/10.1093/carcin/5.12.1689

Wills, J. W., Johnson, G. E., Battaion, H. L., Slob, W., & White, P. A. (2015). Empirical analysis of BMD metrics in genetic toxicology part I: *In vitro* analyses to provide robust potency rankings and support MOA determinations. *Mutagenesis, 31*(3), 255–263. https://doi.org/10.1093/mutage/gev085

Wills, J., Johnson, G., Battaion, H., Slob, W., & White, P. (2017). Comparing BMD-derived genotoxic potency estimations across variants of the transgenic rodent gene mutation assay. *Environmental and Molecular Mutagenesis, 58*(9), 632–643. https://doi.org/10.1002/em.22137

Wilson, A., Grabowski, P., Elloway, J., Ling, S., Stott, J., & Doherty, A. (2021). Transforming early pharmaceutical assessment of genotoxicity: Applying statistical learning to a high throughput, multi end point *in vitro* micronucleus assay. *Scientific Reports, 11*(1), 2535. https://doi.org/10.1038/s41598-021-81875-4

Wilson, J., & Hunt, T. (2002). *Molecular biology of the cell* (4th ed.). Garland Science.

Wilstermann, A. M., & Osheroff, N. (2003). Stabilization of eukaryotic topoisomerase II-DNA cleavage complexes. *Current Topics in Medicinal Chemistry, 3*(12), 1349–1364. https://doi.org/10.2174/1568026033451914

Yamazaki, H., Oda, Y., Funae, Y., Imaoka, S., Inui, Y., Guengerich, F. P., & Shimada, T. (1992). Participation of rat liver cytochrome P450 2E1 in the activation of N-nitrosodimethylamine and N-nitrosodiethylamine to products genotoxic in an acetyltransferase-overexpressing Salmonella typhimurium strain (NM2009). *Carcinogenesis, 13*(9), 979–985. https://doi.org/10.1093/carcin/13.9.979

Yang, J., Ikezoe, T., Nishioka, C., Takeuchi, T., Nobumori, Y., & Yokoyama, A. (2010). AZD1152 inhibits Aurora B kinase and induces endoreduplication in acute myelogenous leukemia cells. *Cancer Science, 101*(6), 1403–1410. https://doi.org/10.1111/j.1349-7006.2010.01554.x

Yauk, C. L., Buick, J. K., Williams, A., Swartz, C. D., Recio, L., Li, H., … Aubrecht, J. (2016). Application of the TGX-28.65 transcriptomic biomarker to classify genotoxic and

non-genotoxic chemicals in human TK6 cells in the presence of rat liver S9. *Environmental and Molecular Mutagenesis, 57*(4), 243–260. https://doi.org/10.1002/em.22004

Youngblom, J. H., Wiencke, J. K., & Wolff, S. (1989). Inhibition of the adaptive response of human lymphocytes to very low doses of ionizing radiation by the protein synthesis inhibitor cycloheximide. *Mutation Research, 227*(2), 257–261. https://doi.org/10.1016/0165-1161(89)90002-5

Zeller, A., Duran-Pacheco, G., & Guérard, M. (2017). An appraisal of critical effect sizes for the benchmark dose approach to assess dose–response relationships in genetic toxicology. *Archives of Toxicology, 91*(12), 3799–3807. https://doi.org/10.1007/s00204-017-2037-3

Zhu, Y., Wang, T., & Jelsovsky, J. Z. H. (2007). Bootstrap estimation of benchmark doses and confidence limits with clustered quantal data. *Risk Analysis, 27*(2), 447–465. https://doi.org/10.1111/j.1539-6924.2007.00897.x

Zhu, A., Wang, Y., & Zhang, L. (2021). An improved K-means algorithm based on evidence distance. *Entropy, 23*(11), 1550. https://doi.org/10.3390/e23111550