


## RESEARCH ARTICLE

Environmental and  
Molecular MutagenesisEnvironmental  
Mutagenesis and  
Genomics Society

WILEY

# Visualization strategies to aid interpretation of high-dimensional genotoxicity data

Stephen D. Dertinger<sup>1</sup> | Erica Briggs<sup>1</sup> | Yusuf Hussien<sup>2</sup> | Steven M. Bryce<sup>1</sup> |  
Svetlana L. Avlasevich<sup>1</sup> | Adam Conrad<sup>1</sup> | George E. Johnson<sup>2</sup>  |  
Andrew Williams<sup>3</sup> | Jeffrey C. Bemis<sup>1</sup>

<sup>1</sup>Litron Laboratories, Rochester,  
New York, USA

<sup>2</sup>Institute of Life Sciences, Swansea University,  
Swansea, UK

<sup>3</sup>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](mailto: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<sup>®</sup> 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](https://creativecommons.org/licenses/by-nc-nd/4.0/) 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 analyses 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 FACS-Canto™ 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

TABLE 1 Chemicals and *a priori* classifications.

Chemical, Abbreviation	CAS No., Source if not Sigma-Aldrich	<i>a priori</i> 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]
Danuserib	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
Noscapiene	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

TABLE 1 (Continued)

Chemical, Abbreviation	CAS No., Source if not Sigma-Aldrich	<i>a priori</i> 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 HpT 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

(Continues)

TABLE 1 (Continued)

Chemical, Abbreviation	CAS No., Source if not Sigma-Aldrich	<i>a priori</i> 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.; Camptosol® 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.; Degraasi 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

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; Aydenir & Bilaloglu, 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 <sub>3</sub> 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

(Continues)



TABLE 1 (Continued)

Chemical, Abbreviation	CAS No., Source if not Sigma-Aldrich	<i>a priori</i> 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 <sub>2</sub> 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 $\alpha$ 2 $\mu$ -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	Fluoroquinolone antibiotic; Floxin® FDA approved label, 2008
Osimertinib	1421373-65-0	Non-genotoxic	Bryce et al., 2018	EGFR kinase inhibitor; in vitro and in vivo genotox 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

TABLE 1 (Continued)

Chemical, Abbreviation	CAS No., Source if not Sigma-Aldrich	<i>a priori</i> 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 Hpvt 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; Hpvt, hypoxanthine phosphoribosyl transferase; MLA, mouse lymphoma assay; MN, micronucleus; SCE, sister chromatid exchange; TB, tubulin binder.



established cytotoxicity limit, that is, the top concentration of each chemical had to exhibit  $\leq 80\%$  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,  $\gamma$ H2AX 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  $\mu\text{M}$  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

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](http://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 as 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

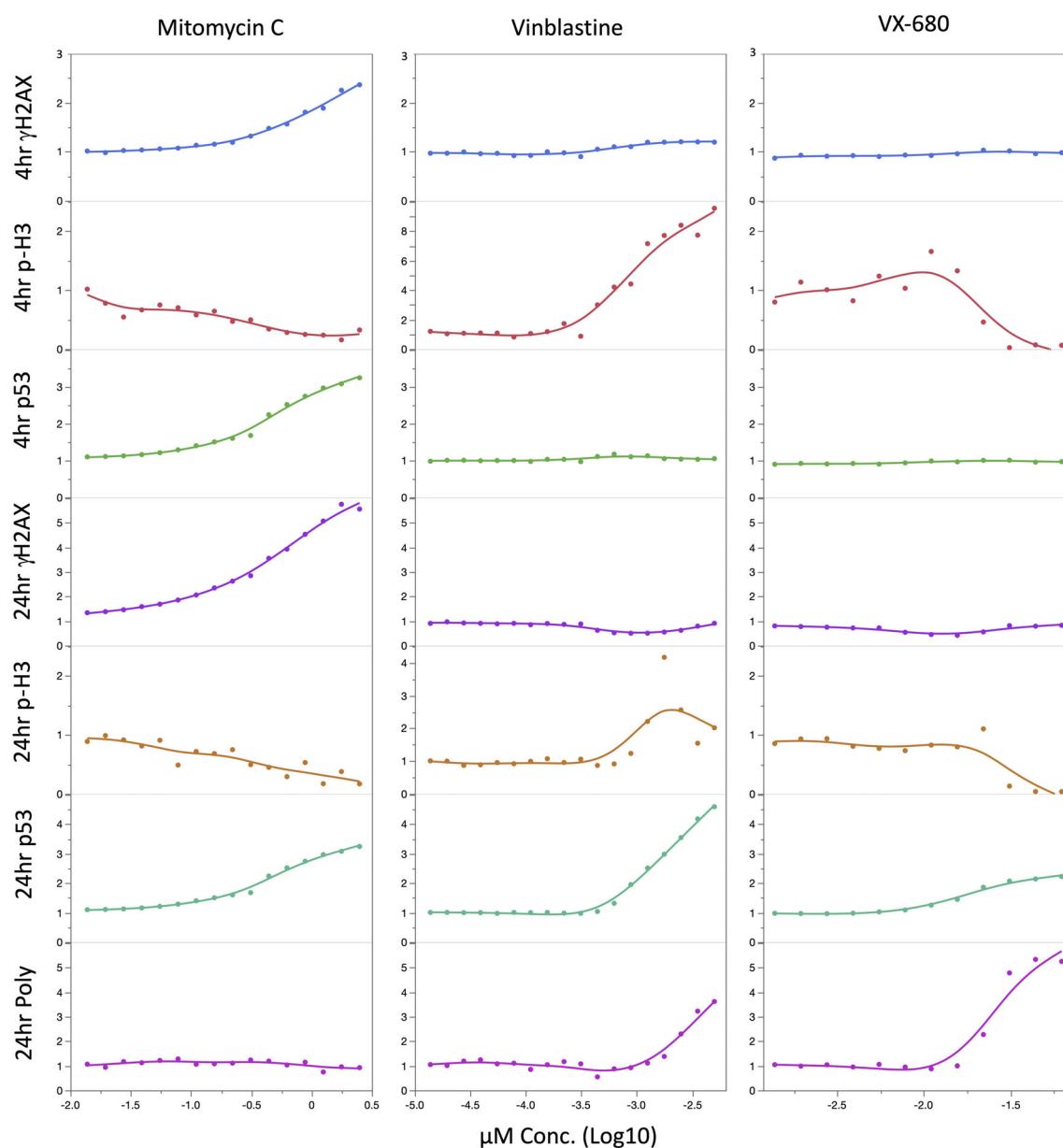
Uniform Manifold Approximation and Projection (UMAP) is a non-linear dimensionality reduction technique that efficiently captures both local and global structures in high-dimensional data (McInnes & Healy, 2018). UMAP aims to create a low-dimensional representation where similar points are modeled as 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;

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.

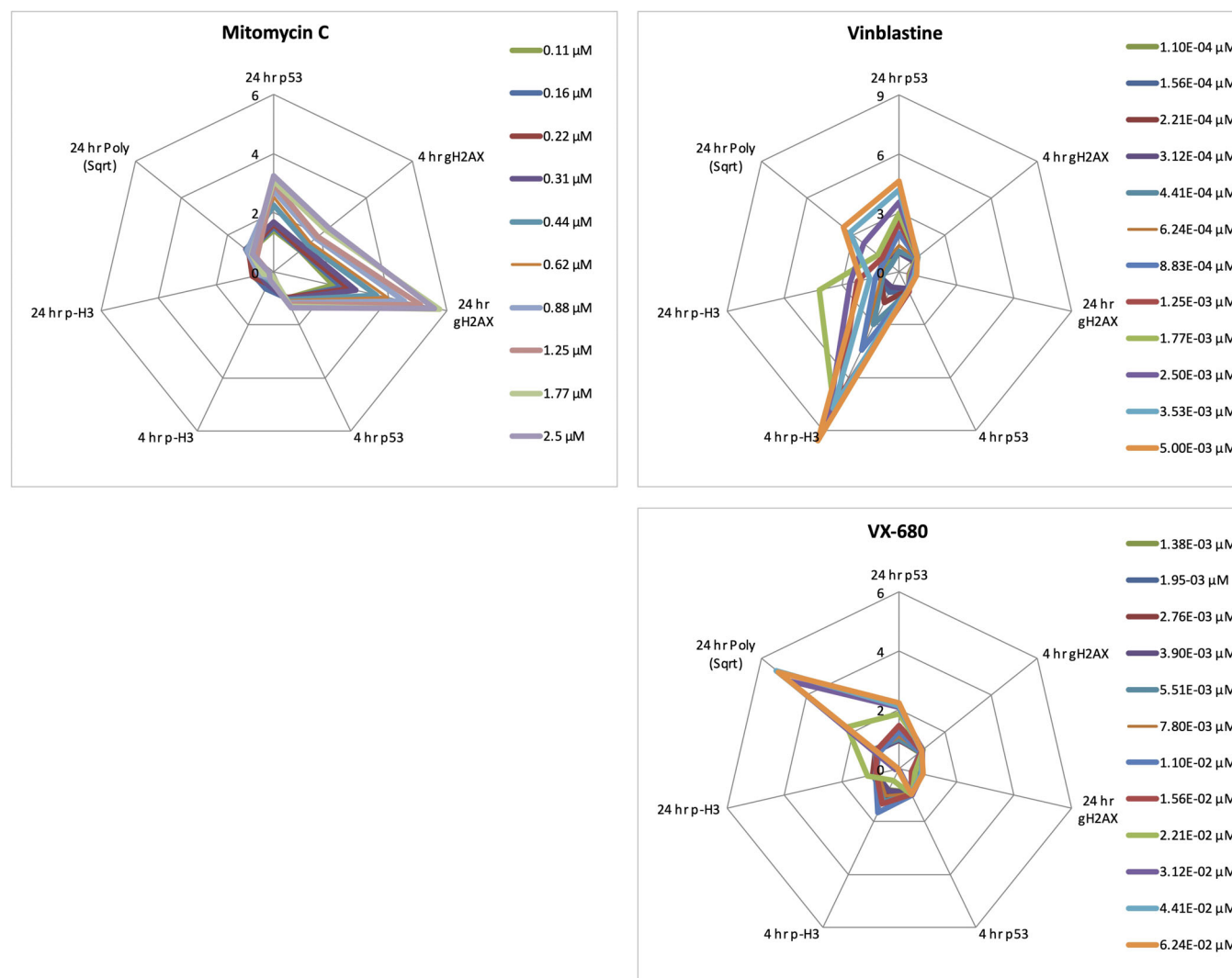
### 3 | RESULTS AND DISCUSSION

#### 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 ( $\mu\text{M}$ ). The left panels correspond to mitomycin C treatment, and show a prototypical clastogenic response pattern whereby  $\gamma\text{H2AX}$  and p53 are elevated at both time points, phospho-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  $\gamma\text{H2AX}$  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  $\gamma\text{H2AX}$  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.



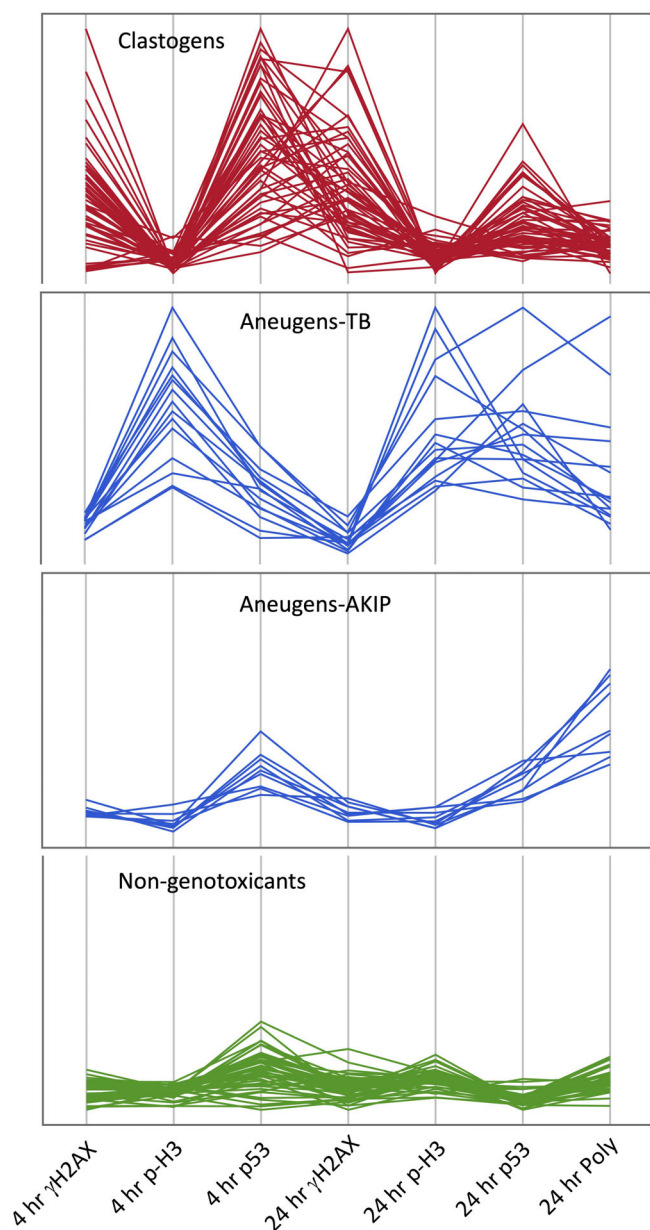
**FIGURE 2** Spider plots show seven biomarker response profiles for three genotoxicants, the clastogen mitomycin C, the tubulin binder-type aneugen vinblastine, and aurora kinase inhibitor phenotype-type aneugen VX-680. The biomarker response data are fold-increase values, and as specified at the right 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



**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.

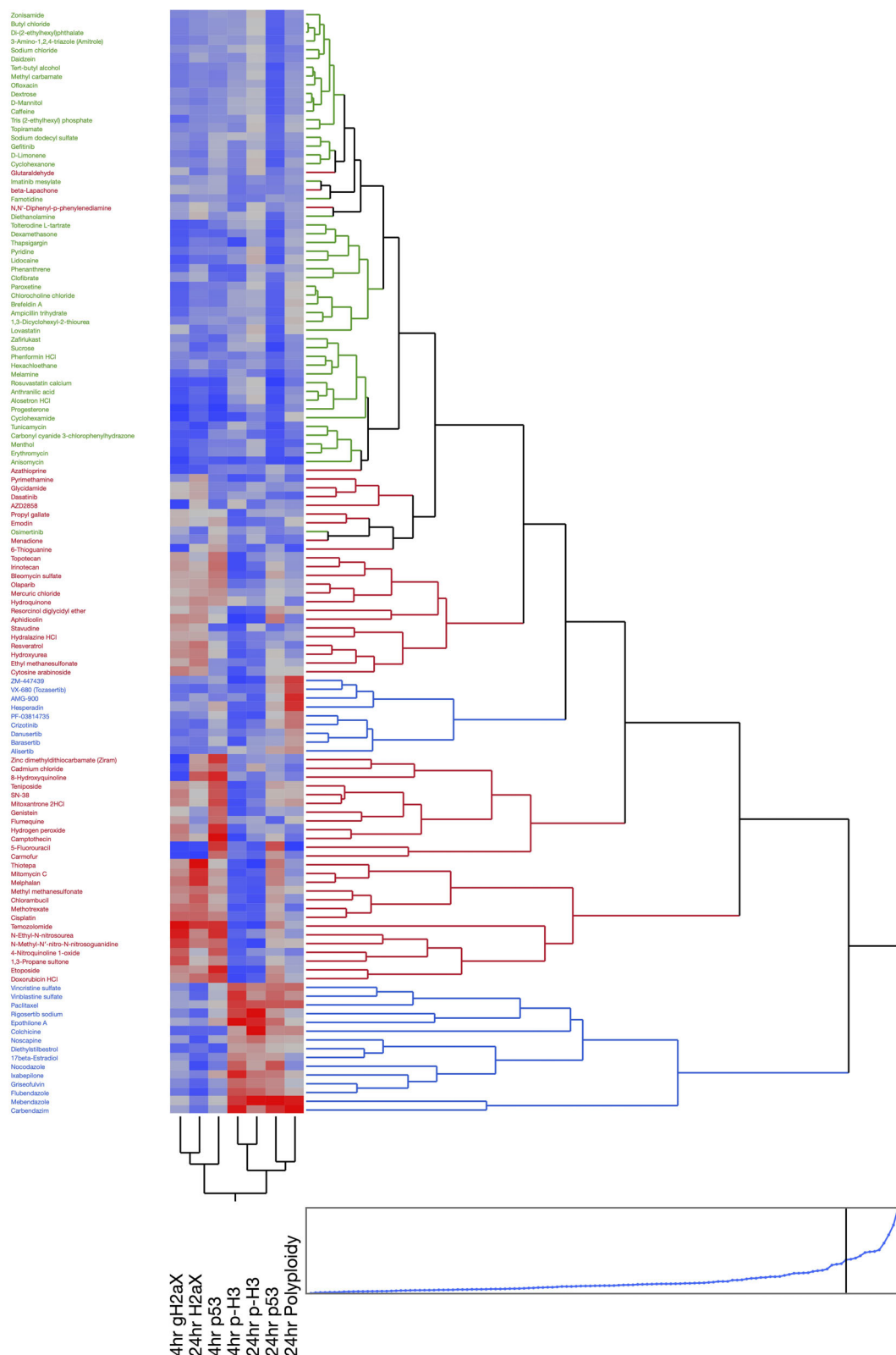
### 3.3 | Parallel coordinate plots

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.

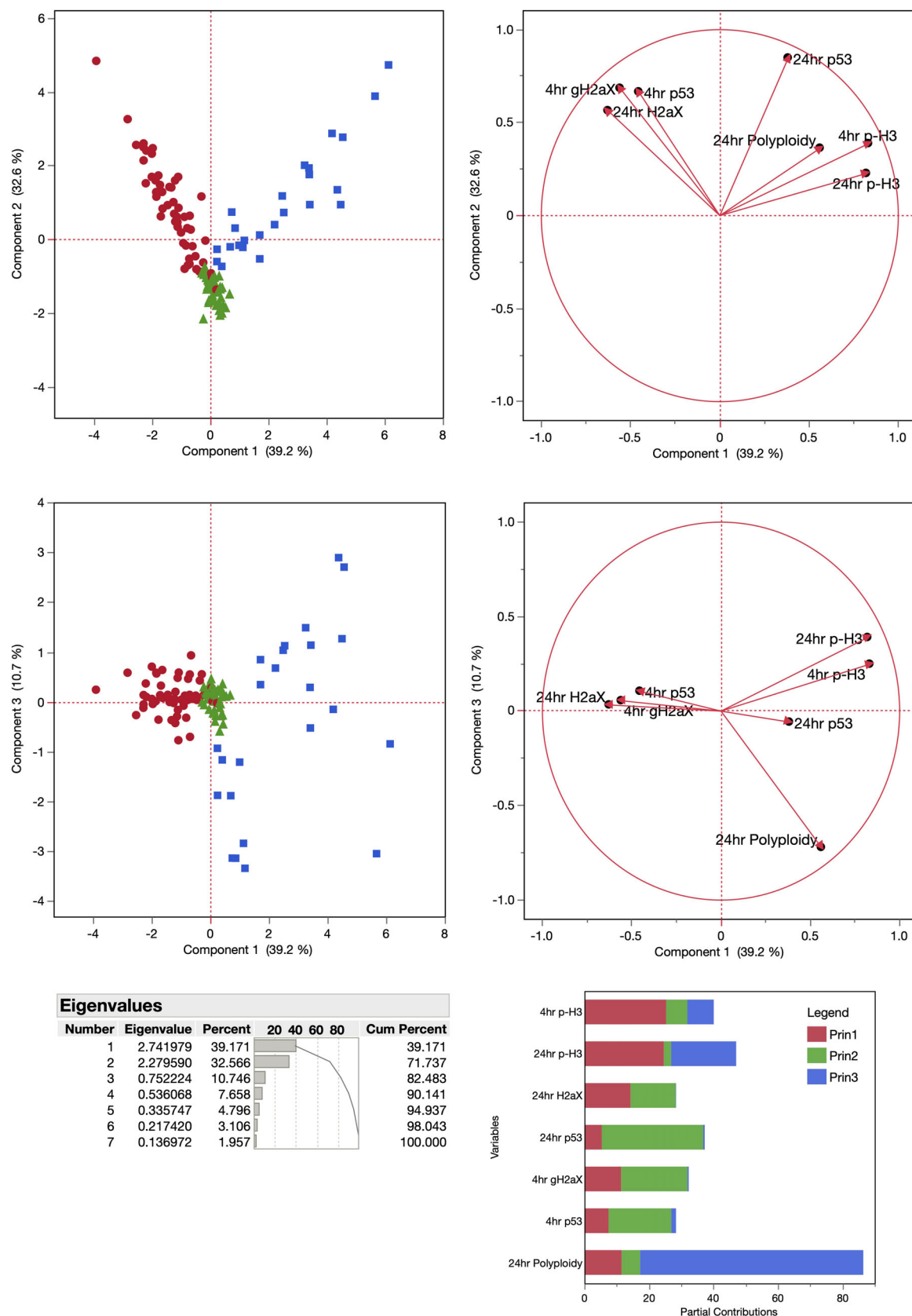
### 3.4 | Hierarchical clustering

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





**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).



**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).



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).

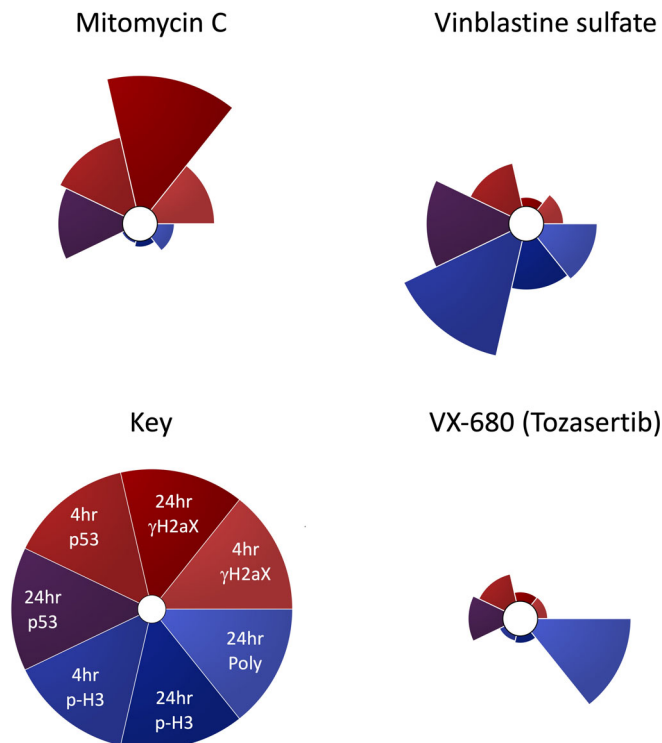
### 3.5 | PCA

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  $\gamma$ H2AX, 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



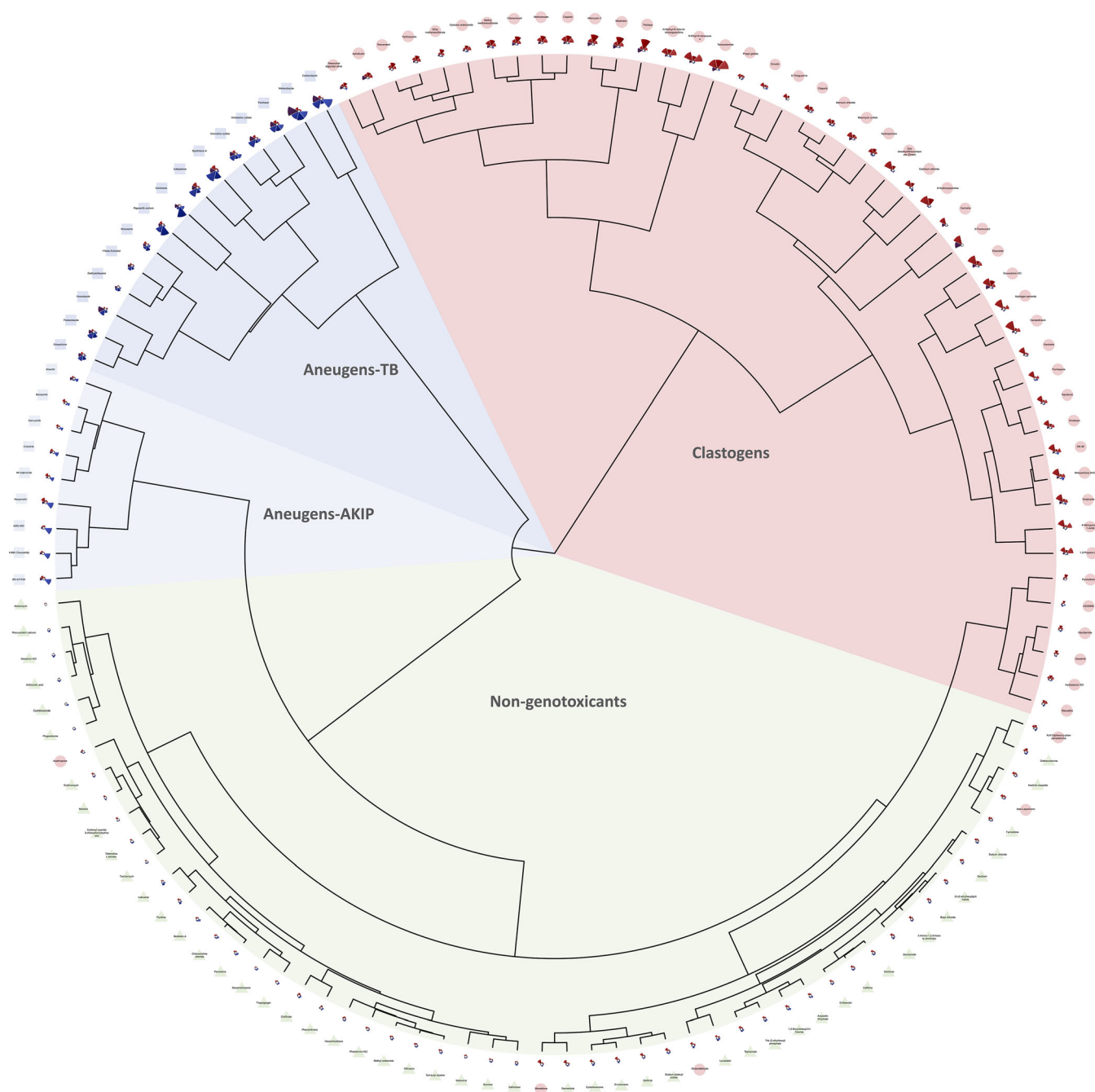
**FIGURE 6** ToxPi profiles are shown for a representative clastogen (mitomycin C), tubulin binder-type aneugen (vinblastine), and aurora kinase inhibitor-type aneugen (VX-680). For each slice, the distance from the origin represents a particular biomarker's AUC value (larger values = longer protrusions). 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.

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.

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

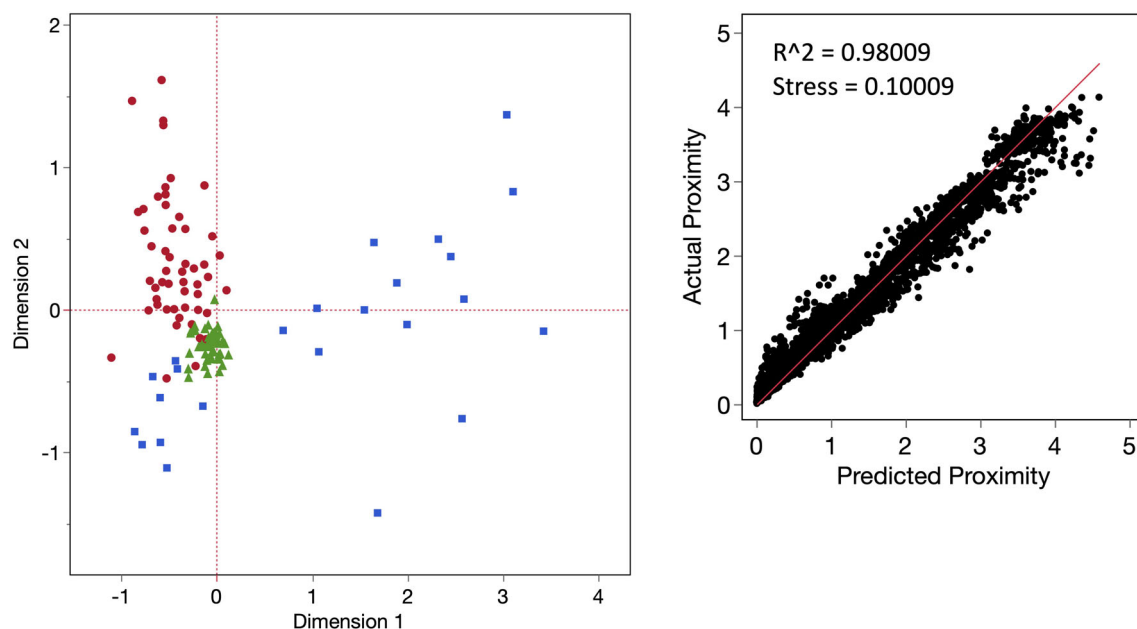
Figure 6 displays three ToxPi profiles, mitomycin C, vinblastine, and VX-680. As seen in the associated key, clastogen responsive



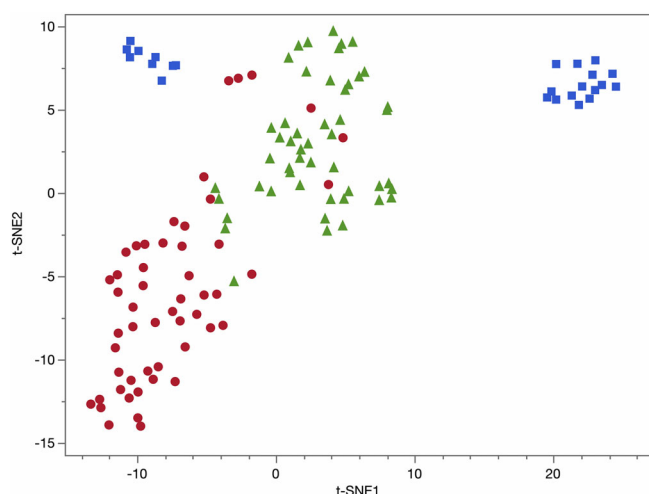
**FIGURE 7** ToxPi profiles are shown for 7 biomarker responses and 126 chemicals. The same biomarker arrangement described in Figure 6 is used here. This image is the result of ToxPi software's hierarchical clustering algorithm, which places like-responding chemicals near each other. The clustering 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  $\gamma$ H2AX, and 24  $\gamma$ H2AX) 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  $\gamma$ H2AX. 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



**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 intention 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.

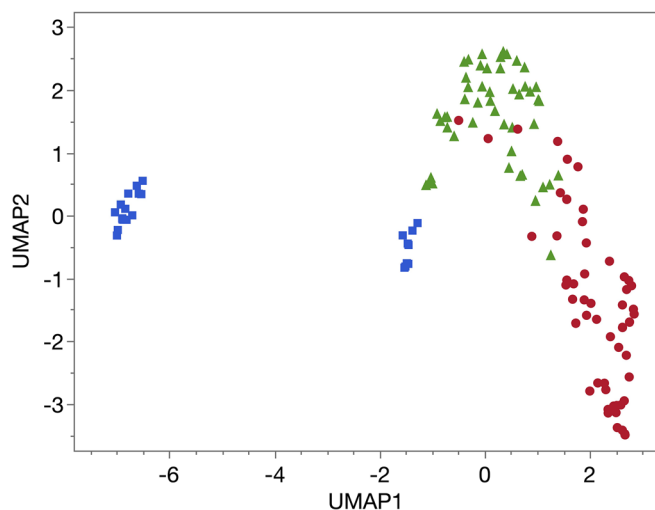
One advantageous feature of ToxPi software is that it provides algorithms for hierarchical clustering. In Figure 7, a circular hierarchical

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



**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 extends 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

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. & Bilaloglu, 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., Degraffi, F. & Cozzi, R. (2013) Effects of resveratrol on topoisomerase II- $\alpha$  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) Epopothilones, 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<sup>®</sup> 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<sup>®</sup> 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<sup>™</sup> cells. *Genes and Environment*, 42, 5.



- 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<sup>®</sup> and the TGx-DDI transcriptomic biomarker in human HepaRG<sup>™</sup> cell cultures. *Frontiers in Public Health*, 9, 694834.
- Camptosar<sup>®</sup> (Irinotecan) package insert (2014) Pfizer Inc, NY, NY. Available from: [https://www.accessdata.fda.gov/drugsatfda\\_docs/label/2014/020571s048lbl.pdf](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  $\beta$ -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<sup>®</sup> assay data analysis strategy. *Environmental and Molecular Mutagenesis*, 60, 513–533.
- Dertinger, S.D., Phonetheswath, 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., Phonetheswath, 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<sup>®</sup> (Ofloxacin) package insert. (2008) Ortho-McNeil, Raritan, NJ. Available from: [http://www.accessdata.fda.gov/drugsatfda\\_docs/label/2008/019735s059lbl.pdf](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](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 administration 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.
- Iressa<sup>®</sup> (Gefitinib) package insert (2003) AstraZeneca Pharmaceuticals LP, Wilmington, DE. Available from: [https://www.accessdata.fda.gov/drugsatfda\\_docs/label/2003/021399lbl.pdf](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

- 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](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](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](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 NAD<sup>+</sup>-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., Ouldeldhkim, 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](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](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 the mitotic spindle in mammalian 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



- 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](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](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 aneugenecity. *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/2002/204125017.pdf](http://www.accessdata.fda.gov/drugsatfda_docs/label/2002/204125017.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 cytotoreductive 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*, 65(5), 156–178. Available from: <https://doi.org/10.1002/em.22604>