Review

Next Generation Testing Strategy for Assessment of Genomic Damage: A Conceptual Framework and Considerations

Kerry L. Dearfield,[1] B. Bhaskar Gollapudi,[2] Jeffrey C. Bemis,[3] R. Daniel Benz,[4] George R. Douglas,[5] Rosalie K. Elespuru,[6] George E. Johnson,[7] David J. Kirkland,[8] Matthew J. LeBaron,[9] Albert P. Li,[10] Francesco Marchetti,[5] Lynn H. Pottenger,[11] Emiel Rorije,[12] Jennifer Y. Tanir,[13]* Veronique Thybaud,[14] Jan van Benthem,[15] Carole L. Yauk,[5] Errol Zeiger,[16] and Mirjam Luijten[15]

[1]U.S. Department of Agriculture, Food Safety and Inspection Service, Washington, District of Columbia
[2]Exponent® Inc, Center for Toxicology and Mechanistic Biology, Midland, Michigan
[3]Litron Laboratories, Rochester, New York
[4]OmnyCorp, Rockville, Maryland
[5]Environmental Health Science and Research Bureau, Health Canada, Ottawa, ON K1A 0K9, Canada
[6]U.S. Food and Drug Administration, CDRH/OSEL DBCMS, Silver Spring, Maryland
[7]Institute of Life Science, College of Medicine, Swansea University, Swansea SA2 8PP, United Kingdom
[8]Kirkland Consulting, Tadcaster, LS24 OAS, United Kingdom
[9]The Dow Chemical Company, Molecular, Cellular, and Biochemical Toxicology, Midland, Michigan
[10]In Vitro ADMET Laboratories LLC, Columbia, Maryland
[11]Formerly of The Dow Chemical Company, Toxicology & Environmental Research and Consulting now with Olin Corporation, Midland, Michigan
[12]National Institute for Public Health and the Environment (RIVM), Center for Safety of Substances and Products, Bilthoven, 3720 BA, The Netherlands
[13]ILSI Health and Environmental Sciences Institute (HESI), Washington, District of Columbia
[14]Sanofi, Drug Disposition, Safety and Animal Research, Vitry-sur-Seine, France
[15]National Institute for Public Health and the Environment (RIVM), Center for Health Protection, Bilthoven, 3720 BA, The Netherlands
[16]Errol Zeiger Consulting, Chapel Hill, North Carolina

*Correspondence to: Jennifer Young Tanir, PhD, Scientific Program Manager, ILSI Health and Environmental Sciences Institute (HESI), 1156 15th Street NW, 2nd Floor, Washington, DC 20005, USA. E-mail: jtanir@hesiglobal.org

Disclaimer: This manuscript has been reviewed by the agencies and organizations of the authors and approved for publication. The views expressed in the manuscript do not necessarily reflect the policy of these agencies and organizations. The findings and conclusions in this paper have not been formally disseminated by the Food and Drug Administration and should not be construed to represent any agency determination or policy. The mention of commercial products, their sources, or their use in connection with material reported herein is not to be construed as either an actual or implied endorsement of such products by the Department of Health and Human Services.

Conflict of interest: J.C.B. and A.P.L. are employed by companies that offer genetic toxicology testing services on a contract basis.

Received 21 June 2016; provisionally accepted 5 August 2016; and in final form 8 August 2016

DOI 10.1002/em.22045

Published online 00 Month 2016 in Wiley Online Library (wileyonlinelibrary.com).

For several decades, regulatory testing schemes for genetic damage have been standardized where the tests being utilized examined mutations and structural and numerical chromosomal damage. This has served the genetic toxicity community well when most of the substances being tested were amenable to such assays. The outcome from this testing is usually a dichotomous (yes/no) evaluation of test results, and in many instances, the information is only used to determine whether a substance has carcinogenic potential or not. Over the same time period, mechanisms and modes of action (MOAs) that elucidate a wider range of genomic damage involved in many adverse health outcomes have been recognized. In addition, a paradigm shift in applied genetic toxicology is moving the field toward a more quantitative dose-response analysis and point-of-departure (PoD) determination with a focus on risks to exposed humans. This is directing emphasis on genomic damage that is likely to induce changes associated with a variety of adverse health outcomes. This paradigm shift is moving the testing emphasis for genetic damage from a hazard identification only evaluation to a more comprehensive risk assessment approach that provides more insightful information for decision makers regarding the potential risk of genetic damage to exposed humans. To enable this broader context for examining genetic damage, a next generation testing strategy needs to take into account a broader, more flexible approach to testing, and ultimately modeling, of genomic damage as it relates to human exposure. This is consistent with the larger risk assessment context being used in regulatory decision making. As presented here, this flexible approach for examining genomic damage focuses on testing for relevant genomic effects that can be, as best as possible, associated with an adverse health effect. The most desired linkage for risk to humans would be changes in loci associated with human diseases, whether in somatic or germ cells. The outline of a flexible approach and associated considerations are presented in a series of nine steps, some of which can occur in parallel, which was developed through a collaborative effort by leading genetic toxicologists from academia, government, and industry through the International Life Sciences Institute (ILSI) Health and Environmental Sciences Institute (HESI) Genetic Toxicology Technical Committee (GTTC). The ultimate goal is to provide quantitative data to model the potential risk levels of substances, which induce genomic damage contributing to human adverse health outcomes. Any good risk assessment begins with asking the appropriate risk management questions in a planning and scoping effort. This step sets up the problem to be addressed (e.g., broadly, does genomic damage need to be addressed, and if so, how to proceed). The next two steps assemble what is known about the problem by building a knowledge base about the substance of concern and developing a rational biological argument for why testing for genomic damage is needed or not. By focusing on the risk management problem and potential genomic damage of concern, the next step of assay(s) selection takes place. The work-up of the problem during the earlier steps provides the insight to which assays would most likely produce the most meaningful data. This discussion does not detail the wide range of genomic damage tests available, but points to types of testing systems that can be very useful. Once the assays are performed and analyzed, the relevant data sets are selected for modeling potential risk. From this point on, the data are evaluated and modeled as they are for any other toxicology endpoint. Any observed genomic damage/effects (or genetic event(s)) can be modeled via a dose-response analysis and determination of an estimated PoD. When a quantitative risk analysis is needed for decision making, a parallel exposure assessment effort is performed (exposure assessment is not detailed here as this is not the focus of this discussion; guidelines for this assessment exist elsewhere). Then the PoD for genomic damage is used with the exposure information to develop risk estimations (e.g., using reference dose (RfD), margin of exposure (MOE) approaches) in a risk characterization and presented to risk managers for informing decision making. This approach is applicable now for incorporating genomic damage results into the decision-making process for assessing potential adverse outcomes in chemically exposed humans and is consistent with the ILSI HESI Risk Assessment in the 21st Century (RISK21) roadmap. This applies to any substance to which humans are exposed, including pharmaceuticals, agricultural products, food additives, and other chemicals. It is time for regulatory bodies to incorporate the broader knowledge and insights provided by genomic damage results into the assessments of risk to more fully understand the potential of adverse outcomes in chemically exposed humans, thus improving the assessment of risk due to genomic damage. The historical use of genomic damage data as a yes/no gateway for possible cancer risk has been too narrowly focused in risk assessment. The recent advances in assaying for and understanding genomic damage, including eventually epigenetic alterations, obviously add a greater wealth of information for determining potential risk to humans. Regulatory bodies need to embrace this paradigm shift from hazard identification to quantitative analysis and to incorporate the wider range of genomic damage in their assessments of risk to humans. The quantitative analyses and methodologies discussed here can be readily

applied to genomic damage testing results now. Indeed, with the passage of the recent update to the Toxic Substances Control Act (TSCA) in the US, the new generation testing strategy for genomic damage described here provides a regulatory agency (here the US Environmental Protection Agency (EPA), but suitable for others) a golden opportunity to reexamine the way it addresses risk-based genomic damage testing (including hazard identification and exposure). Environ. Mol. Mutagen. 00:000–000, 2016. © 2016 The Authors. Environmental and Molecular Mutagenesis Published by Wiley Periodicals, Inc.

Key words: genetic toxicology; risk assessment; exposure assessment; integrated testing strategy; mutagenicity

INTRODUCTION

Testing for mutagenicity endpoints has been in practice for over half a century with both in vitro and in vivo test systems [for review see Cimino, 2006]. For the past several decades, a fairly standardized approach has been in place for regulatory testing of substances for mutagenic activity [e.g., see Dearfield et al., 1991; Müller et al., 1999; Eastmond et al., 2009]. The standard battery incorporates a bacterial mutagenicity test ("Ames" assay), usually one or two in vitro mammalian cell tests selected for gene mutations (e.g., mouse lymphoma assay) and for chromosome damage (e.g., micronucleus or chromosomal aberration test), and one or more in vivo tests depending on the results of the in vitro tests and the regulatory guidance being followed. This standard battery has served the regulatory purposes well when, during the early years of mutagenicity testing, most of the substances being tested were generally electrophilic and therefore able to damage DNA. In more recent times, because of smarter structural design during development of new substances to avoid reactive moieties in the molecule, many substances are not typical "direct" DNA-damaging substances. Some of these non-electrophilic substances may induce damage to the genetic material, but not be readily detected in the standard battery. It is clear now that many other types of genomic alterations exist besides gene mutation, clastogenicity, and aneugenic events that can impact human health. Thus, it is time to address this wider range of alterations and the related perturbation of toxicity pathways. For genetic toxicity testing to evolve and take into account emerging technologies and recent scientific knowledge, a different approach is needed to capture the potential of a substance for inducing functional genomic alterations. Due to the broader scale of genomic alterations and functional outcomes, we must clearly move away from a "one-size-fits-all" standard test battery approach and develop a more flexible approach that also includes an understanding of the underlying mechanisms affecting adverse outcomes.

Historically, the genetic toxicology testing community has based its testing strategy on a simple yes/no paradigm in which the main question was whether the substance being tested has genomic damaging capability (i.e., hazard identification). The next generation approach described in the following discussion takes into account many aspects of accepted risk assessment practices (e.g., planning and scoping, quantitative analysis of dose-response results, exposure assessment) that the genetic toxicology testing community has not usually embraced. This entails greater emphasis on estimating the potential risk of a substance if and when people are exposed rather than applying genotoxicity testing data only for hazard identification. Adhering more closely to these risk assessment practices will provide more informative data for risk management of exposures to substances that have potential to damage the genome. Many risk guidances and documents have been written over the years and this discussion will refer to those materials for fuller descriptions rather than provide extensive detail here, except to highlight how genomic damage testing fits into these practices.

Further, the more flexible approach will allow for assessment of a greater diversity of genomic damage than is currently being measured. Initially, cancer was the endpoint for which most genotoxicity assays were used to predict hazard. However, there is now much greater knowledge of the contribution or association of genomic damage and other endpoints (e.g., adverse outcome pathways, epigenetics) with various health outcomes besides cancer [Milic et al., 2015]. This widening focus includes heritable genetic damage (germ cell risk), aging (accumulation of genetic damage) [Moskalev et al., 2012; Wolters and Schumacher, 2013], cardiovascular disease [Uryga et al., 2016], and the specific contribution of different types of genomic damage (e.g., specific mutations in tumor suppressor genes and oncogenes) to cancer risk and the various stages of cancer (e.g., [Dycaico et al., 1996; Pottenger et al., 2014]). The use of a more flexible approach allows deployment of new tools, such as high-throughput assays and advanced sequencing approaches, in ways that permit integration of the new knowledge on both genomic endpoints and the potential health consequences of genomic interactions.

To provide direction as to what a "next generation" approach might look like, the Genetic Toxicology Technical Committee (GTTC) of the International Life Sciences Institute (ILSI) Health and Environmental Sciences Institute (HESI) formed a work group to formulate a flexible strategy. This work group had various discussions on this

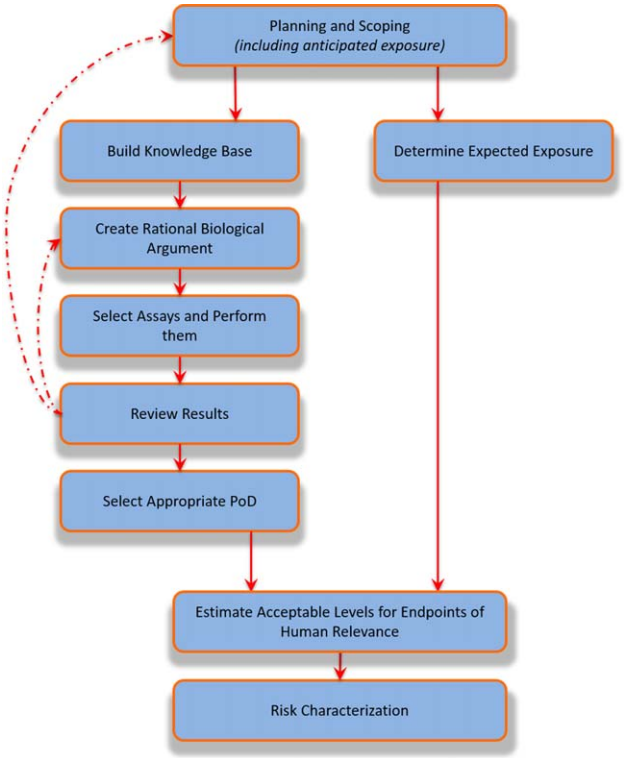

Fig. 1. Strategy for examining genomic damage

flexible approach, including workshops at the 6th International Workshop on Genotoxicity Testing (IWGT) in Iguassu, Brazil, on October 31, 2013 and at the GTTC Annual Meeting in Washington DC, on April 13, 2015. The outcomes from these discussions are the basis for the present manuscript. Since this topic directly relates to other GTTC work groups addressing quantitative analysis, germ cell testing, and emerging technologies, output and results from discussions held in those work groups are also taken into account.

Work group members generally agreed that important aspects (discussed in the following sections in more detail) to be considered and/or incorporated when designing a next generation approach include: (i) advances in mechanistic understanding of toxicity of a given substance; (ii) the human relevance of the mechanism(s) involved; (iii) the need to expand from genetic to genomic alterations, and to include other endpoints of genomic damage associated with human diseases besides cancer; (iv) the need for a flexible, efficient, and animal-sparing approach to assess more substances, with greater speed and accuracy; (v) the need to consider likely human exposure; (vi) the need to take into account potentially susceptible populations and different life stages in humans; and (vii) to contextualize the data in terms of risk rather than simply identifying a hazard. These challenges may be met by using pathway-based approaches to characterize the processes by which toxic substances induce adverse health effects. The National Research Council (NRC) presented the concept of "toxicity pathways" in their report on "*Toxicity Testing in the 21st Century: A Vision and a Strategy*" [NRC, 2007]. In this report, toxicity pathways are defined as normal signaling processes that lead to adverse health effects if significantly perturbed. To organize and communicate the available knowledge on toxicity pathways in a structured and consistent manner, the concept of Adverse Outcome Pathways (AOPs) has been proposed [Ankley et al., 2010]. An AOP depicts, in a linear way, linkages between chemically induced adverse effects at various levels of biological organization as they progress from a molecular initiating event (MIE) to an adverse outcome (AO) [Ankley et al., 2010]. AOPs are important tools to enhance the implementation of pathway- and mechanistic-based approaches in risk assessment [Villeneuve et al., 2014], and this has been recognized by the Organisation for Economic Co-operation and Development (OECD) [OECD, 2013].

In the following sections, we describe a stepwise approach for a next generation strategy to assess the risk of genomic damage as a result of exposure to chemical substances. Since there are many variables and potential pathways, the strategy discussed here does not provide a specific decision-tree type of approach (i.e., "a one-size fits all" approach), but rather introduces and discusses the many factors that should be taken into account when designing a rational testing scheme for any particular substance of interest. The applicability of this framework is also highly relevant to the evolving regulatory context; for example, with the recent update to the Toxic Substances Control Act (TSCA), the U.S. Environmental Protection Agency (EPA) will utilize risk-based evaluations to determine safety and priorities [see USEPA, 2016 for details of updated act].

APPROACH

We propose a general defined approach to provide a useful and consistent strategy for examining the genomic damage a substance may induce when a person is exposed (Fig. 1). It will produce the information required to determine the extent of public health risk that may occur; if there is unacceptable risk, the information may help to direct possible mitigation actions. The approach outlined here consists of steps that ultimately provide knowledge about the risk posed by genomic damage and estimates of risk. Such an approach offers consistency while providing a framework for flexibility in the specific actions and testing that can be applied for specific substances or combination of substances. Overall, this approach is consistent with the approach developed in the ILSI HESI RISK21 project [Embry et al., 2014].

Planning and Scoping

Before attempting to test a substance for genomic damage, it is necessary to establish the reason(s) for the testing; this is the planning and scoping process. From the risk assessment literature, this upfront discussion and thought process is a critical step before embarking on subsequent actions [NRC, 2009]. The planning and scoping process defines the purpose and scope of the testing and focuses on the issues and specific approach(es) involved in the testing [for fuller description see USEPA, 2000—Chapter 2; USEPA, 2014a]. Planning and scoping provide the opportunity to define what is expected to be covered in the risk assessment and to explain the purposes for which the risk assessment information will be used.

The main purpose of planning and scoping is to decide whether testing for genomic damage is necessary and what particular approach(es) should be considered. If testing is required for regulatory purposes, then the relevant regulation(s) must be considered. Alternatively, testing may be necessary as a result of public concern, based on some scientific findings that need clarification, insights for further development of a substance, or other factors. Human relevance must be addressed, including identifying what exposures to people need to be considered. For example, if there is low or no anticipated exposure by any route, the need for any further testing beyond the minimal data/information set should be carefully evaluated. Most importantly, a listing of risk management questions is created that focuses on why testing is necessary and what the path forward should be, i.e., does testing fit with management goals and policies and what is the context for any testing. This is the "fit for purpose" approach the NRC discusses as critical for efficient and effective risk assessment efforts to support risk management decision making [NRC, 2009].

Exposure information can provide an important perspective to define or to refine a test plan, and to help ensure appropriate data on hazard and dose response are available to support risk assessment. These data can inform testing of key elements and thus contribute to resource optimization, including the 3 "Rs" of animal welfare (Replace, Reduce, Refine). While it is unlikely that reliable quantitative exposure data will be available upfront for test plan development of early stage or new substance entities, these substances nonetheless have a target application in focus, which could lend itself to qualitative or semi-quantitative (e.g., levels of concern, ranking) approaches as an early stage exposure assessment.

If the exposure assessment is used as a screening device for setting priorities, the emphasis is probably more on the comparative risk levels, perhaps with the risk estimates falling into broad categories (e.g., semi-quantitative categories such as high, medium, and low) [USEPA, 1992]. Developing a semi-quantitative scheme leading to categorizing ("binning") the likely exposure scenarios, based on a combination of target application arenas and structural information, e.g., quantitative structure–activity relationship (QSAR) or read-across, can provide a general direction toward the appropriate base set of data to assess toxicity to the genome. For example, Health Canada developed a strategy based on three identified "levels of concern" (LOC); this approach relies on a combination of application-type and structural information [HPB, 1993]. Similarly, the International Council of Chemical Associations (ICCA) has proposed such a semi-quantitative scheme based on expected potential for exposure drawn from an application-type approach [ICCA, 2011]. Additionally, not only the extent of exposure would be important, but the nature of the susceptible groups in the population that may be exposed would be just as important to consider and may influence the level of concern.

In concert with the high-throughput toxicity testing efforts of ToxCast and related programs, there is an EPA-led effort to develop high-throughput exposure assessment approaches, i.e., ExpoCast, to provide an early perspective on exposure and risk. A framework is proposed to extend efforts beyond simple models, such as USEtox (UNEP-SETAC toxicity) [Rosenbaum et al., 2008, 2011] and RAIDAR (Risk Assessment IDentification And Ranking) [Arnot et al., 2006], to more complex ones that incorporate Bayesian approaches and reverse toxicokinetics to support exposure estimates based on National Health and Nutrition Examination Survey (NHANES) biomonitoring data [Wambaugh et al., 2013]. Extensive data analysis revealed that this information was helpful in prioritizing substances according to their relative risk to induce adverse human health effects, and that such early incorporation of exposure information can help determine priorities for additional data collection.

Categorization of exposure scenarios could be applied in order to define an initial base set of genetic toxicity data. Of course, a base set is exactly that—only the base—and as such may require additional data to adequately inform risk assessment. Qualitative categories to inform development of a base set of genetic toxicity testing data are listed in Table I. Such an approach could be further refined by establishing a sliding scale from low or no "discernable" concern to higher levels of concern, and by establishing limits for top concentrations/doses tested in the toxicity assays, which would then further rely on the collection of dose-response data. For example, for a substance anticipated to have a wide-spread human exposure, it may be necessary to examine the genetic toxicity at multiple dose levels that are above and below the human exposure levels to determine a point-of-departure (PoD) based on dose-response modeling [Gollapudi et al.,

TABLE I. Exposure-Based Qualitative Categories to Inform Development of a Base Set of Genetic Toxicity Testing Data

Exposure Group	Exposed Population	Exposure-based Category and Expected Actions
Closed system/Isolated intermediate– Industrial use only	Industrial/Production workers only	Minimal/Low exposure; Expect reliable use of recommended Personal Protective Equipment (PPE)
Incorporation into or onto matrix–Industrial use	Industrial workers only	Low exposure; Expect reliable use of recommended PPE
Non-dispersive/Professional use	Professional workers only	Moderate exposure potential; Expect reliable use of recommended PPE; Dose-response data/determination of PoDs provides perspective on potential for risk
Wide dispersive	Environmental/Human populations	Potential for wide exposure in general population; Industrial chemicals, drugs, *etc*. which are discharged into the environment via waste streams during manufacture or end of life after disposal; No expectations *vis-à-vis* PPE; Dose-response data/determination of PoDs provides perspective on potential for risk; Will likely conduct risk assessment
Wide dispersive/Consumer use	Consumers	Potential for wide/high exposure in general population; Chemicals such as drugs, devices, and food-based substances for consumer use; No expectations *vis-à-vis* PPE; Dose-response data/determination of PoDs provides perspective on potential for risk; Will likely conduct risk assessment

2013; Johnson et al., 2014]. For a substance with minimal exposure potential, a small series of high-throughput tests, or if the regulatory authority desires much more certainty, a small scale in vivo test at a limit dose, for example, 1,000 times greater than the anticipated human exposure (and likely with a reduced animal number), may provide adequate information for prioritization of the substance for any further testing.

The strategy for the selection of limit concentrations for in vitro studies was extensively addressed by the ICH (International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use) S2R1 guidance document for human pharmaceuticals [ICH, 2011] and by OECD Guidance documents [OECD, 2014a,b] for other test substances. These strategies were aimed at increasing the specificity of the test systems while optimizing the sensitivity to detect potential genotoxic agents. Both ICH and OECD recommend that for relatively insoluble test substances, the limit concentration should be the lowest concentration with minimal amount of precipitate. For freely soluble human pharmaceuticals that are not cytotoxic, the ICH recommended a limit concentration of 1 mM or 0.5 mg/mL, whichever is lower. This recommendation was based on a survey of clinical exposures to known human pharmaceuticals and the systemic levels achievable in pre-clinical laboratory animal studies. The OECD, on the other hand, recommended a limit concentration of 10 mM or 2 mg/mL, whichever is lower, to represent the best balance between molarity and mg/mL and to cover substances with a wide range of molecular weights. In the case of test substances that are mixtures or of unknown composition, the OECD recommends a limit concentration of 5 mg/mL. While these recommendations are useful as default upper limits, it is recommended that all available information on a test substance (e.g., toxicokinetics and metabolic saturation) is taken into consideration to select a scientifically justifiable upper limit dose or concentrations for testing.

Planned case studies to further develop a flexible approach for the assessment of genomic damage will, amongst others, be used to define base sets of genetic toxicity data for the various categories of exposure. These case studies will cover different regulations; depending on the particular regulatory requirements, base sets, if needed at all, may therefore differ in composition.

Logistically, planning and scoping also outlines the testing approach and its follow-up. Aspects such as who will do the testing, predicted costs, and what milestones and timeframes are involved are delineated. Planning and scoping also identifies which stakeholders need to be involved and what coordination may be necessary to work with the stakeholders. Additional actions including incorporating peer review and information quality analysis should be part of the planning and scoping discussion. Options regarding how to move forward with any risk management actions, such as mitigations, should be presented as possibilities at this stage in advance of testing and before results are reviewed. A question that might be addressed at the end of planning and scoping is whether there is enough information already available that a decision not to test might be an option. This question may be asked again after building the knowledge base.

Build the Knowledge Base

Usually occurring concurrently with planning and scoping is the effort to assemble what is known already about the substance(s) being considered, i.e., building the knowledge base on the substance of interest. There are many data streams to examine and from which to extract

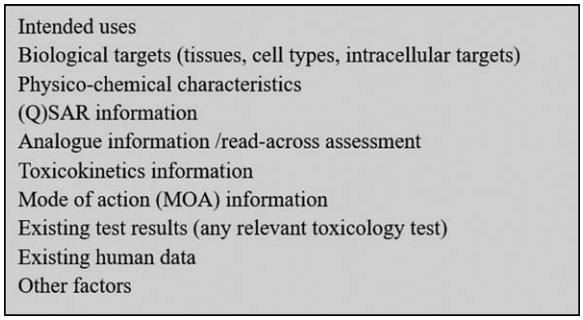

Fig. 2. Available information/data.

any extant information, especially because in many instances substances are being developed or considered for some particular purpose(s) and therefore will likely fall under some regulatory scrutiny. Figure 2 presents some of the pertinent information that can be assembled. It should be noted that existing data do not necessarily need to be exclusively from genetic toxicity tests. This allows the development of a broader range of knowledge of the substance in the absence of apical test data or in vitro testing data.

The resulting knowledge base provides input to the planning and scoping considerations and provides preliminary information in relation to the risk management questions. It may even provide some of the initial answers to those questions. Just as important as identifying what information and data are available, this effort identifies information needs and data gaps that may be important in learning and understanding more about the substance's potential to cause genomic damage. Ultimately, building the knowledge base helps shape the future direction in deciding what testing is needed.

Intended Uses, Biological Targets, Physico-Chemical Characteristics

The intended use can provide information regarding the potential routes of exposure and likely extent of exposure. Further, if a substance is being developed for a specific purpose, its intended target can focus where the potential biological activity may occur (i.e., tissue, cellular, or intracellular targets). This, together with the physico-chemical characteristics (e.g., volatility, pH, solubility), can provide insight into what potential reactions may occur. This could provide ideas regarding what tests would be most appropriate to assess that activity (e.g., electrophilic reactions with DNA sites).

(Quantitative) Structure–Activity Relationships and Analogue Searching (Read-across)

(Quantitative) Structure–Activity Relationships [(Q)SAR] and analogue information provide predictions of potential chemical toxic activity. With the available (Q)SAR software systems and their models, one can distinguish those methods that make a quantitative prediction (QSAR), e.g., a quantitative probability of an adverse effect or a maximum recommended daily dose of a pharmaceutical, and those non-quantitative SAR models that aim to indicate the potential of a substance to exert a toxicological effect. (Q)SAR models can be used for both screening and prioritization. Such a prioritization component could complement in vitro testing, wherein the theoretical models are used to indicate the type of in vitro testing required to establish a more detailed (quantitative) indication of the potential human health hazard of a specific substance.

In many instances, the parent substance is the substance of interest for testing. However, consideration should also be given, when possible, to potential metabolites that may occur which also need to be tested (pharmaceuticals; food additives), particularly if they have structural alerts for possible toxicity that may not be present in the parent substance. There are various in silico methods to predict metabolism, by applying metabolism simulators such as META (MultiCASE Inc.), METEOR (Lhasa Ltd.), or the metabolism simulators included in the freely available OECD QSAR Toolbox. A complicating factor in these metabolism simulators is the absence of a (good) indication of the relevance (expected abundance) of individual predicted metabolites.

The application of most available (Q)SAR models is restricted to organic substances of small to moderate size. Models covering inorganic substances are very rare (although some are now emerging, e.g., aiming at explaining the toxicity of inorganic nanomaterials), and most (Q)SAR models do not cover organo-metallic substances as their chemistry is often very different from "normal" organic chemistry. The areas in human health toxicology where the largest numbers and most successful applications of (Q)SARs can be found (e.g., mutagenicity, sensitization) are mechanistically related to some form of reactivity, which can be seen as an inherent property related to chemical structure. Some (Q)SAR software systems now available are described in Table II.

Although for certain toxicity endpoints and specific classes of substances (Q)SAR models have been suggested as full replacements for in vitro or even in vivo testing; in general, such predictions are rarely accepted for regulatory purposes. One exception exists for predictions of bacterial mutagenicity under the very specific conditions defined in ICH M7 for impurities in pharmaceuticals [ICH, 2014]. Other regulatory programs have accepted QSAR and/or read-across for impurities which cannot be synthesized/purified in sufficient quantity for use as a test article, generally handled on a case by case basis. (Q)SAR prediction is accepted in the REACH (Registration, Evaluation, Authorisation, and Restriction of Chemicals) regulatory framework as an input into a Weight of Evidence, where model prediction, for example

TABLE II. Examples of Existing (Quantitative) Structure–Activity Relationship ((Q)SAR) Software Systems (Listed in Alphabetical Order by Software Name)

CASE Ultra (MultiCASE, Inc.)	• CASE Ultra is a commercial (Q)SAR software system that provides the two complimentary (Q)SAR methodologies (statistics-based QSAR and expert rule-based SAR) currently suggested for evaluating the potential toxicity of chemicals • CASE Ultra provides test results, identifying alerts, their statistical parameters and explaining the way they are used to arrive at the final prediction call
Derek Nexus and Sarah Nexus (Lhasa Limited)	• Derek Nexus is a commercial human expert-based SAR software system that evaluates the potential toxicity of existing or prospective chemicals by identifying substructures in substances that are known to be related to a specific type of chemistry related to toxicity based on previously acquired human experience • Derek Nexus covers many specific toxicity endpoints, but with varying depth of coverage for each (*e.g.*, it contains over 400 alerts related to bacterial mutagenicity), but only 13 alerts are currently identified that are related to reproductive toxicology (fertility and developmental toxicity) • Derek Nexus makes decisions about which chemicals are likely to have 'more favorable' toxic profiles when the user does not have as much experimental information as desired about the toxicity of each chemical • Sarah Nexus processes submitted chemical structures by fragmentation, after which the fragments are reviewed for activity versus inactivity; the Sarah Nexus model then arranges those 'interesting' fragments into a network of hypotheses (or nodes) and relevant hypotheses are used to inform an overall prediction of toxicity • The Sarah Nexus prediction includes an overall conclusion about the toxicity in a structure, confidence rating in that prediction, as well as supporting examples
Model Applier (Leadscope, Inc.)	• Model Applier is a commercial (Q)SAR software system that provides the two complimentary (Q)SAR methodologies (statistics-based QSAR and expert rule-based SAR) currently suggested for evaluating the potential toxicity of chemicals • Model Applier QSAR models are built using the structural feature and properties as descriptors. The models encode the relationship between these descriptors and the toxicity endpoint, such as the results of the bacterial mutagenesis assay. The modeling technique used to generate these models is referred to as partial logistic regression • When a statistical-based prediction is made on a new chemical, the same structural features and properties in the model are calculated for the test compounds; these descriptors are then used with the models to calculate a probability of a positive result • The Leadscope genetox expert alerts are based on well-defined mutagenicity structural alerts from the literature, validated against a large database of over 7,000 chemicals with *Salmonella* mutagenicity data (the reference set); alongside this list of alerts, deactivating factors as well as active subclasses (which represent possible cohorts of concern) are encoded • A positive expert-based prediction is made where one or more alerts are present with no defined deactivating factor • In addition, the software determines whether the test compound is similar enough to known classes of chemicals such that it is not trying to extrapolate to areas of chemistry the system has never seen • A positive prediction is only made when the compound is within this applicability domain; a negative prediction is made for chemicals that are within the applicability domain that either contains no alert or when the alert is deactivated
Symmetry (Prous Institute for Biomedical Research)	• Symmetry is a commercial statistical-based QSAR software system that applies advanced machine learning techniques to a variety of structural features and physicochemical properties of small molecules to provide quality predictions about biological effects • Available Symmetry algorithms include binary classification for active/inactive data sets, meta-classifiers to achieve consensus predictions for sets of binary models and multi-label learning that yields ranking and probabilistic estimates of the possible outcomes • Symmetry uses models based on data sets comprised of small molecules and associated biological properties
Toolbox (OECD)	• The OECD Toolbox is a freely available eclectic collection of contributed (Q)SAR models of varying quality and usefulness • Toolbox tries to identify similar substances for which toxicity data are available, and that subsequently allows the user to apply read-across, define a grouping approach (with or without trend line), and/or apply existing simple QSARs • Toolbox relies on existing models and information and does not introduce new models, but allows the user to apply existing knowledge/models • Numerous theoretical models are available for the prediction of toxicity, differing in their applicability domains (some are meant for a narrowly defined class of substances, *e.g.*, substituted phenols, while others aim at predictions, *e.g.*, "low molecular weight organic chemicals" in general)
VEGA (freeware)	• VEGA uses statistical modeling, but supplies the user with "analogue" data • VEGA is a culmination of models from several European Union Framework projects aimed at development of QSARs for toxicity • Endpoints covered by VEGA models include mutagenicity, carcinogenicity, skin sensitization and developmental toxicity

together with in vitro information, can be sufficiently convincing to replace (waive) in vivo testing [REACH, 2007].

Analogue searching (read-across) is a methodology that relies on identifying similar chemicals to the one of interest, not only in terms of (overall) chemical structure, but specifically also with a similar functionality which is related to the toxicological effect of interest. Other properties that require similarity for a successful read-across are physico-chemical properties and/or bioavailability, and when metabolism could play a role in activation or detoxification, a similar (expected) metabolic profile. If a sufficiently similar substance can be found for which toxicity data are available, one can hypothesize that the substance of interest has similar toxic potential. Read-across can be both qualitative (substance X is a known skin sensitizer, therefore very similar substance Y will also very likely be a skin sensitizer) and quantitative (when the potency or the effect dose is "read-across"). For the latter, one assumes that the substance of interest has, for example, a similar lowest observed adverse effect level (LOAEL) as the laboratory-tested chemical having a closely similar structure and/or functionality. If multiple, sufficiently similar substances with toxicity data can be identified, this will strengthen the accuracy of the read-across (often termed a grouping or category approach). If a trend is visible within the group, the read-across can be refined by taking into account this trend (e.g., increasing toxicity with increasing octanol–water partition coefficient).

(Q)SAR/read-across methods are only as reliable as the biological databases on which they are based. The most extensive genetic toxicity databases are those which are built on libraries of data from bacterial mutagenicity studies. Also, in vitro mammalian cell assay data collected more than a few years ago employed test methods and standards no longer seen as reliable for some assays. A survey of pharmaceutical lead compounds found structural determinants for clastogenicity due to non-covalent DNA binding that interfere with topoisomerases [Snyder et al., 2013], thus emphasizing the importance of including models which extend beyond the domain of compounds that are positive in bacterial mutagenicity assays.

Toxicokinetics

Absorption, distribution, metabolism, and elimination (ADME) of a substance are key determinants of toxic potential, including genomic damage. Absorption or bioavailability determines systemic exposure upon exposure of the organism to the substance under study. Tissue distribution of a substance and/or its metabolite(s) after absorption is a result of multiple factors such as lipophilicity of the substance and rate and quantity of blood flow to key organs and tissues. Metabolism or biotransformation of a substance distributed to tissues with xenobiotic metabolic capacity, such as the liver, may lead to enhanced or decreased genomic damage due to metabolic activation or detoxification, respectively. Elimination of a substance and its metabolites lowers body burden of exposure via various routes including renal and biliary excretion, perspiration, and respiration. Xenobiotic metabolism represents a key elimination pathway (metabolic clearance). Experimental evaluation of plasma pharmacokinetics and metabolite identification, organ distribution, and major route(s) of elimination with in vivo studies allows for the generation of relevant data with respect to systemic and organ exposure. While pharmacokinetic data are routinely developed for pharmaceuticals, they are often not available for chemicals being developed for other purposes. Instead, physiologically based pharmacokinetic (PBPK) modeling can be applied for the prediction of in vivo internal doses. PBPK models are built using limited or extensive ADME data, derived from in vivo, in vitro, and/or in silico studies. They provide an estimation of internal exposure based on physicochemical data coupled with known physiological factors such as route of exposure, blood flow, organ weight, and xenobiotic metabolism pathways. PBPK can be used to model various routes of exposure: dermal, oral, intravenous, and inhalation, etc., coupled with exposure regimen (single dose; multiple doses; continuous exposure) to estimate exposure levels in various organs [Chetty et al., 2014]. In some instances, the estimated exposure levels in each tissue can be used, in combination with in silico/in vitro findings, to predict in vivo genomic damage. Incorporation of xenobiotic metabolism and carrier-mediated uptake and efflux transport in conjunction with known population-based variations in these parameters can provide population specific pharmacokinetic parameters, allowing prediction of population-specific effects.

PBPK models vary highly in complexity and in the requirements for data that are used as input. Consequently, the predictions generated by these models may differ greatly in level of uncertainty. Some circumstances may require a higher level of confidence and thus more detailed ADME data. For such purposes, short-term in vivo studies with small numbers of animals could be performed. An advantage of generating in vivo metabolism data during pharmacokinetic studies is that information on other tissues besides the liver can be collected. Liver is considered the site of initial metabolism for most ingested substances, but the gut microflora and intestinal walls are also metabolically competent. In addition, there are unique metabolic capabilities in a variety of tissues including lung, kidney, gonads, and other hormone-producing organs such as thyroid and adrenal glands; thus ADME studies can be an important source of information regarding the target tissue(s) of a particular substance.

A recognized need for future testing is the incorporation of human-specific metabolism. Xenobiotic metabolism may lead to activation as well as inactivation of chemical substances. The current practice of using induced rat liver S9 for in vitro studies may be limited in certain instances due to species differences in xenobiotic metabolism. Species differences in drug metabolizing enzyme pathways are well established. For example, of the major P450 isoform subfamilies involved in drug metabolism, species differences in isoforms are observed for CYP2B, CYP2C, CYP2D, and CYP3A. Pathway differences can lead to species differences in rate of metabolism, identity of metabolites, and associated toxic effects [Lewis et al., 1998]. Estimation of human genomic damage based on in vitro or in vivo data may result in a misleading conclusion if, for example, human-specific metabolism is a key determinant of genomic damage for the substance in question. Particularly in these cases, use of test systems incorporating (some aspects of) human metabolism should be considered.

Mode of Action Information and Human Relevance

Mechanistic understanding of biological pathways involved in toxicity is essential for a next generation strategy to assess genomic damage and its relevance to humans. Several conceptual frameworks on understanding mode of action (MOA) have been published. The most widely-known is the Mode of Action/Human Relevance Framework (HRF) developed by the International Programme on Chemical Safety (IPCS) from the World Health Organization (WHO) and the ILSI Risk Sciences Institute (ILSI RSI) [Boobis et al., 2006, 2008]. The HRF has been updated recently and now also considers dose-response relationships and species concordance in weight-of-evidence analysis of hypothesized mode(s) of action (MOA(s)) for critical effects and their qualitative and/or quantitative relevance to humans [Meek et al., 2014]. The Quantitative Key Events/Dose-Response Framework (Q-KEDRF) from ILSI HESI's RISK21 project [Simon et al., 2014] builds upon both the HRF and the earlier Key Events/Dose-Response Framework (KEDRF) [Julien et al., 2009]. It provides a structured approach for understanding of the dose-response and temporal relationships between the various key events (KEs) and the adverse outcome as well as between the KEs themselves. For this, two additional concepts are introduced: associative events (AEs) and modulating factors (ModFs) [Simon et al., 2014]. AEs can be considered as biomarkers for KEs, whereas ModFs affect the timing and/or dose-response of KEs. Life stage, disease state, genetics, lifestyle, and other factors underlie inter- and intra-individual variability in the nature and strength of ModFs.

Understanding MOA is closely linked to the growing use of adverse outcome pathways (AOPs); MOA and AOP are similar, but distinct. Both can be described as a series of KEs at different levels of biological organization (e.g., molecular, subcellular, cellular, tissue, organism) that result in a pathological or other disease outcome(s) [Ankley et al., 2010 for review; OECD, 2013]. While the AOP is ideally chemical-agnostic, a MOA is chemical-(or agent-) specific. The MOA includes exposure and metabolism, while an AOP starts with the initial molecular interaction following any necessary metabolism. In addition, the term MOA, however, does not necessarily imply adversity [Meek et al., 2014]. An example of an application of the AOP concept for germ cell genomic damage is provided for DNA alkylation [Yauk et al., 2015b]. The AOP concept is highly valuable when designing a new strategy as it provides structure and terminology for organizing toxicological understanding across different levels of biological organization. It also offers a framework for integrating in silico models and in vitro data and in vivo bioassays for toxicity testing. Understanding of the MOA is then essential to link together the observations made in the various assays.

Germ Cell Testing

Much discussion has centered on the concern for genomic damage to the germ cells and the possible resultant heritable risk for an organism. An IWGT workshop examined this issue and provided some key outcomes that need to be considered when formulating risk management questions and developing a testing strategy that incorporates assessment of potential germ cell damage [Yauk et al., 2015a]. Among the major outcomes, it was highlighted that available data suggest that somatic cell tests can detect most germ cell mutagens, but there are strong concerns that suggest caution in drawing conclusions about potential germ cell susceptibility based solely on data from somatic cells. When questioning whether germ cell testing should be conducted, one consideration suggests that if a substance or its metabolite(s) will not reach target germ cells or gonadal tissue, it is not necessary to conduct germ cell tests, notwithstanding the somatic outcomes. Further, it was recommended that negative somatic cell mutagens with clear evidence for gonadal exposure and evidence of toxicity in germ cells could be considered for germ cell mutagenicity testing. For characterizing risk of somatic mutagens that are known to reach the gonadal compartments and expose germ cells, the substance could be assumed to be a germ cell mutagen without further testing. However, if a quantitative analysis is needed to estimate germ cell risk, germ cell testing would provide the type of data that would be needed.

At the IWGT workshop, the working group recognized that possible insights regarding germ cell risk can be obtained from other toxicity tests. Specifically, standard

reproductive toxicology tests (e.g., one-generation and multi-generational reproductive toxicology tests) and repeat dose toxicity studies may provide signals pertinent to germ cell genotoxicity, and are an important source of information relating to potential germ cell hazards [ICH, 2005]. These assays capture important developmental stages (e.g., in utero exposure, most of spermatogenesis) that are not assessed using standard genetic toxicology approaches and provide a wealth of information on reproductive endpoints that can indicate both delivery of the agent to male and female germ cells and gonadal tissues, as well as cytotoxic effects that may occur following exposure to genotoxicants. In addition, conduct of full pathology on reproductive tissues is considered a very sensitive method to identify effects. For example, evidence of reduced testis weight and sperm count, increased implantation loss and post implantation loss, and fetal developmental abnormalities may be indicative of a possible genotoxic mechanism. A critical question relates to whether adverse reproductive endpoints in humans, which are found in ~4% of live births, are related to genetic (germ cell) damage, epigenetic effects, or non-genetic developmental outcomes. The vast majority of adverse human birth outcomes are inherited or caused by non-disjunction events in aging ova, but most of these can be quite readily removed from consideration when evidence for genomic (and now more recently epigenetic) damage is sought [Elespuru, 2011].

Integration of germ cell tests with routine somatic cell testing was encouraged at the workshop. For instance, the Globally Harmonized System of Classification and Labelling of Chemicals (GHS) provides a hazard categorization for germ cell mutagenicity, thus emphasizing the importance of this endpoint [UNECE, 2015]. OECD established guidelines for a few germ cell tests that can be considered as part of a testing strategy, particularly for transgenic rodent mutation models, aberration analysis in spermatogonia, and the dominant lethal test [OECD, 2015]. Though other germ cell tests do not have internationally harmonized guidelines and currently have limitations, implementation of other sperm assays that assess genomic integrity, such as the TUNEL (terminal deoxynucleotidyl transferase dUTP nick end labeling) and SCSA (sperm chromatin condensation) assays are promising test systems. Whole genome sequencing is particularly emphasized as a powerful approach to determine mutational spectra, to give insight to germ cell specific mechanisms, and to quantify induced heritable mutagenesis. Sequencing can also help determine if a "weak" response in a mutation assay may be biologically relevant or not. It was noted, as described above, that limited types of genomic damage are assessed using current germ cell approaches. It was emphasized that assays to address newly emerging genomic endpoints, such as copy number variants (CNVs), were urgently needed.

Epigenetics

Many in the toxicology field propose expanding consideration of genomic damage to include epigenetic mutations. Epigenetic mutations, also called "epimutations," can be defined as persistent changes in gene activity and expression that occur, unlike genetic mutations, without a change in the nuclear DNA sequence. Epimutations also include persistent changes in chromatin (DNA methylation and/or histone modifications) and non-coding RNAs, including microRNAs. The increasing understanding of a multitude of additional factors within the cell that exert a significant amount of control over the ultimate output of the genome has resulted in the creation of the field of epigenetics. Within the last decade, epigenetics has expanded from a niche within genetics, to an established discipline that plays important roles in basic biology, as well as applied human health. Two aspects connecting epigenetic changes and human health outcomes include evidence that substances cause measurable and reproducible changes in epigenetic patterns [Thomson et al., 2014; Zang and Peng, 2015] and that epigenetic alterations are important factors in tumor development and other human health outcomes [Chi et al., 2010; You and Jones, 2012].

In addition to its position within basic science, there is an emerging role of epigenetics in the fields of risk assessment and genetic toxicology. However, at present there is insufficient knowledge to determine how epigenetics information can contribute to toxicology and risk assessment, and how to integrate it into a structured safety program [Goodman et al., 2010]. The important issue that first must be addressed is the need to understand normal variation and influence of epigenetic markers in an unperturbed system [LeBaron et al., 2010; Rasoulpour et al., 2011]. This highlights the critical need for a well-informed base of comparison if the impact of epigenetic changes due to specific exposures is to be appropriately interpreted. Similar concerns have been raised in terms of the recognized importance of epigenetics in pharmaceutical safety studies. A possible testing paradigm was described as a means to examine the feasibility of incorporation of epigenetic endpoints into a larger toxicology testing strategy [Priestley et al., 2012]. Recently, special issues of (i) Mutation Research/Genetic Toxicology and Environmental Mutagenesis ["Epigenetics and Chemical Safety"; Volumes 764–765, 2014] and (ii) Environmental and Molecular Mutagenesis [Volume 55(3), 2014] focused on the impact of environmental exposures on the epigenome. Among the contributions, advancements in the understanding of epigenetic markers were discussed [Thomson et al., 2014], though this information was described in the context of carcinogenesis, and specifically the potential mechanisms of non-genotoxic carcinogens.

The challenges mentioned above and the continued efforts by many laboratories worldwide to further the

knowledge of epigenetics serves to reinforce the desire to apply what is known to regulatory safety programs and human health risk assessment in a rational and well thought-out manner. Though there are obvious links between the fields of genetic toxicology and epigenetics, the current level of understanding does not yet enable a formally prescribed integration of the two and makes it difficult to address epigenetics in a testing strategy at this time. The stated need for more information on normal variation across the multitudes of available markers only serves to reinforce this perspective. As the current risk assessment strategy improves and as the basic biology of epigenetics is better defined, epigenetics should be kept in mind and relevant data added to the knowledge base where feasible in order to produce more useful and effective outcomes for risk assessment and human safety. As a result, the testing strategy described here will not specifically address epigenetics, though enough flexibility exists such that epigenetics can certainly be considered, again where feasible and where appropriate tests have been devised.

Create Rational Biological Argument

Using the knowledge base and the focus provided by the planning and scoping process, a rational biological argument for deciding the direction of testing can be put forward. The process for building the argument is similar to thinking described in an earlier HESI workgroup product [Dearfield et al., 2011]. The process applies the knowledge base created, and considers the data gaps that need to be addressed, to identify the specific tests and assay targets that are the most appropriate to assess potential genomic damage. For example, MOA/AOP information is used to identify potential key events and the associated assays to be used to detect them. Use of structural analogues and read-across considerations can point to appropriate testing approaches for that type of substance. The purpose of the biological argument then is to identify the tests that are most likely and relevant, to determine any potential for induction of genomic damage by that substance.

When creating the specific biological argument for what testing to pursue, the risk management questions and the extent of the knowledge base will help determine the amount of testing needed. Categorization of exposure scenarios (see Table I) can help inform how to proceed. The biological argument then needs to determine how much concern the substance may present to exposed people. Since the most basic question centers on whether the substance has the capability or not to induce genomic damage, some minimal amount of information/data is necessary to provide some indication of this capability, regardless of the level of concern the substance ultimately may have. Based on the risk management question(s) from planning and scoping and the information from this minimal information/data set, the decision for additional testing or not can be made.

The determination of a level of concern to attach to a substance relies on several factors that will likely be unique to each substance, or class of substances, and perhaps in other instances, mixtures of substances. The extent of potential or projected exposure to humans is a primary factor. As more exposure (as discussed above) is anticipated, the level of concern rises and the argument focuses on what additional testing beyond the minimal set is needed. Associated with the extent of exposure is the question of whether the exposure to the substance is intentional or not. If exposure to a substance is intentional such as for a pharmaceutical, pesticide, or food, then testing beyond the minimal set is likely needed, and even more likely as the projected number of people who will be (or are) exposed increases. For unintentional exposures, such as accidents or environmental contaminants, a different level of concern will be applied depending on the extent of exposure (number of people) and what subgroups of the population may be exposed (e.g., susceptible groups). The amount and types of testing being considered will obviously be different with the different scenarios (e.g., if only concerned about accidental high exposures producing a genomic effect, then testing at a limit dose/concentration would be part of the rationale). Further thinking should extend to consideration of possible genomic damage as it relates to heritable effects in offspring and future generations; this may heighten the level of concern and may call for testing of germ cell related targets. The projected or possible route of exposure will also inform what types of tests will be useful (e.g., dermal exposure versus inhalation versus oral ingestion).

While creating the argument for what tests to conduct, greater weight should be given to consideration of endpoints of genomic damage associated with human disease, especially since human risk is being evaluated [MacGregor et al., 2015b]. Furthermore, the genomic endpoint(s) should be consistent with those identified as key events in an AOP leading from the molecular initiating event to disease [Yauk et al., 2013, 2015b; MacGregor et al., 2015b]. A Quantitative Key Events/Dose-Response Framework (Q-KEDRF) provides a useful structured quantitative approach and guidance for a systematic examination of the dose response and timing of KEs resulting from a dose of a substance that potentially can induce genomic damage [Simon et al., 2014].

Where possible, this approach will be employed in the case studies that will be conducted to further delineate a next generation strategy for genomic damage.

Select Assays and Perform Them

Ideally, the assays selected for testing will specifically address the types of genomic damage of concern, e.g.,

pertinent to the likely mode(s) of action described in the biological argument. The number of assays selected can range from a few tests to several tests. The testing strategy may include other types of toxicity testing (e.g., developmental/reproductive toxicity testing) that could provide insights about possible adverse genetic outcomes. The testing will likely be conducted first with in silico methods and in vitro tests (including high-throughput tests), but can include in vivo animal tests as the need arises [see Thomas et al., 2013 for discussion of such an approach]. An understanding of newer testing approaches can be informative in the selection of the assays to use [Zeiger et al., 2015].

OECD testing guidelines exist for users to follow to ensure consistency, reliability, and reproducibility of results destined to support regulatory purposes [OECD, 2015]. These guidelines describe the minimal protocol and test information needed for an acceptable test. The reliance on OECD testing guidelines has been, and continues to be, a great service to the regulatory testing community. However, the OECD guidelines have been developed only for the tests most widely used and those required by regulatory guidance. The universe of these guidelines is becoming very small compared to the multitude of assays now available to examine all sorts of toxicity pathways and health-related endpoints—assays that can be very useful for assessing genomic damage in its many forms.

It is clear that tests without OECD guidelines will be conducted (and needed) and the test universe will be expanding rapidly (e.g., high-throughput tests, molecular-based tests). Development of OECD guidelines for all these tests will be difficult if not prohibitive. A critical concern related to the use of newly developed, non-guideline assays is their validity, that is, their reliability (reproducibility) and relevance for the particular effect being measured [OECD, 2005]. This is particularly important for methods that have not been extensively used beyond the laboratory that developed the method or for a method that the testing laboratory has no or minimal experience using. OECD recognizes the situation and provides some general guidance on how to describe non-guideline in vitro tests [OECD, 2014a]. This guidance outlines the elements considered relevant for providing a comprehensive description of an in vitro method to facilitate an assessment of the quality of data produced and its potential utility in regulatory applications. Confidence in particular tests will come with increasing and broader use and repetition from competent laboratories, and from peer review.

Minimal Information/Data Set

Regardless of the concern level for any substance, a minimal amount of information is necessary to determine what decisions regarding genomic damage can be made. At the very least, this information should clarify whether there is the potential for genomic damage that requires further testing to address. To collect this minimal information/data set, a starting point would be to use in silico methods (e.g., QSAR, computational methods) and high-throughput assays that can provide a broad coverage of potential toxicity pathways. This approach, wherein high-throughput data from in vitro assays or small organisms is combined with comprehensive and robust computational methods to predict toxicity, has demonstrated its usefulness, e.g., [Liu et al., 2015]. For genomic damage, MOA underlying "traditional" genetic toxicity endpoints as well as MOA that are not associated with direct DNA damage but known or hypothesized to be involved in genomic damage, e.g., oxidative damage, enzyme inhibitors, anti-metabolites, activation of transposable elements, should be covered. These early predictive methods and assays provide initial results, but also insight as to whether the biological argument initially constructed is still feasible, or whether an alternative argument needs to be considered.

The results from a minimal set of information/tests can lead to a presumption of hazard (and possibly an estimation of risk if the test results are amenable) and provide input to the risk management decision among the following options: (i) commit to additional testing if results are positive, (ii) discontinue testing in order to initiate mitigation(s) if positive, or (iii) stop any further work if negative. These decisions also take into account the initial exposure scenario—if minimal and/or low exposure is expected or seen, then the decision to bring the testing to a close is more likely; whereas, if high and/or widespread exposure is expected, additional testing is likely even in the case of an initial negative result (unless circumstances dictate immediate action as in an emergency exposure situation).

Additional Testing

Testing beyond the minimal test set may be necessary under several scenarios. For the most part, results from any additional tests, in concert with the minimal test results, should be sufficient to provide input into the risk estimates needed to address the risk management questions posed. The direction that additional testing would take depends on what the circumstances dictate, such as positive results from the initial test set that need clarification or supporting evidence, or if the projected exposure scenario shows extensive human exposure.

When positive results are found from the minimal test set and additional testing is indicated, for whatever reason, the guidance provided in an earlier ILSI HESI GTTC work product is instructive [Dearfield et al., 2011]. Based on the nature of the substance's structure and the

genomic damage seen, and the information on the potential mode of action, the specific additional tests can be identified (e.g., an assay for a particular event along a toxicity pathway). From the initial tests, if known toxicity pathways are implicated, an understanding could be reached that certain tests can be utilized for evaluating that pathway, thus providing a consistent selection of appropriate additional tests. In a sense, this creates a "network" pattern of test selection for decision flow.

We envisage that additional testing may also comprise whole genome sequencing. The use of whole genome sequencing is rapidly expanding to help determine genomic damage in humans and animals and these approaches are now being incorporated into testing for genomic alterations. It has been shown recently that a typical human genome contains on average about 100 loss-of-function variants, and as many as 20 complete loss-of-function mutations [MacArthur et al., 2012]. However, thus far there are no data on how specific substance exposures may impact the mutational load in the genome. Next generation sequencing technologies are now mature and cost effective enough to be applied to identify intrinsic and extrinsic variables that can affect genomic damage. A recent proof of concept for the application of whole-genome approaches in genetic toxicology was described where array comparative genome hybridization and whole genome sequencing in mice were used to produce a genome-wide survey of induced mutations following paternal exposure to ionizing radiation [Adewoye et al., 2015]. Significant increases in copy number variants indels, and multi-site mutations were found in the descendants of irradiated fathers, demonstrating the practical application of these technologies for assessment of heritable mutations. The study suggests that use of whole genome sequencing methods can be very useful in examining genomic damage and should be considered when delineating a specific strategy for a substance of concern.

When considering additional in vivo studies, the possibility of "piggy-backing" genomic damage endpoints onto existing in vivo protocols (e.g., subchronic toxicity studies, two-generation reproductive toxicity studies, extended one-generation reproductive toxicity studies) should be examined (also ensure that the combination assay does not compromise the genomic damage component). These in vivo assays could be further combined with whole genome sequencing techniques to examine potential genomic damage at the molecular level and provide further detailed information regarding the sequence elements that may be subject to alteration by the specific exposure. Integrated animal testing, for example with multi-tissue transcriptome analysis for pathway perturbations, is also a distinct possibility [Thomas et al., 2013] providing whole-genome assessment of potential pathway perturbations in any tissue. A further possibility is the use of transcriptional signatures to predict toxicities and facilitate chemical screening where, for example, a transcriptomic biomarker, TGx-28.65, was shown be a possible biomarker of genotoxicity [Yauk et al., 2016].

Review Results

After the testing is completed, the results are reviewed to assess whether there is a significant effect in any of the tests for genomic damage. These results are then further analyzed for relevance for human adverse outcomes. It should be noted that cancer is not the only adverse outcome of concern for humans. All possible consequences of genomic damage in humans should be considered. This is important for characterizing the risk (see later).

The review should consider the distinction between statistical significance and biological significance. Assay protocols usually have guidance as to what makes for a statistical increase in effect over controls (or background) and the increases need the usual analysis. But in addition, the biological relevance of any indicated induction of genomic damage should be scrutinized as well and characterized as to being relevant or not for any adverse outcome. It is becoming evident that in some cases assay systems are providing greater sampling power (e.g., flow-cytometric techniques) and are able to distinguish very small increases as statistically significant—it is therefore appropriate to address whether such very small increases are also considered as biologically significant.

Another important conclusion is to ascertain the usefulness of the dose-response function for further quantitative analysis (next section). The study design, developed prior to testing, will determine the number and spread of the concentrations or doses used. This should be planned to provide data that are suited for quantitative dose-response analysis. This will also help with describing any uncertainties and possible variation that might be associated with the data. Increasingly, approaches for extrapolating in vitro concentrations to likely human doses (in vitro to in vivo extrapolation (IVIVE) approaches) are being considered to compare in vivo dosage at probable/possible target(s) in exposed persons [e.g., Embry et al., 2014; Patlewicz et al., 2015]. This may potentially alleviate the need for in vivo testing if the in vitro results can provide meaningful data to extrapolate to predicted human response outcomes for decision makers to consider.

Ultimately, the results need to be reviewed for consistency with the proposed mode of action and whether they are of concern for human adverse outcomes. If the test results do not contribute to this understanding, analysis is needed regarding why not, and whether another biological argument needs to be created and a different testing approach be pursued. The process may require revisiting the original risk management questions and reviewing/revising the original upfront planning and scoping decisions.

Select Appropriate Points of Departure (Dose-Response Modeling)

Genetic toxicity testing has usually been performed at relatively high concentrations for in vitro studies and high (e.g., maximum tolerated dose (MTD)) doses for in vivo studies. This was consistent with the hazard identification approach for addressing whether the substance in question had genomic damaging capability or not. In order to achieve a better understanding of a substance's potential risk to exposed persons, a more detailed dose-response analysis is needed. In some instances, the dose-response curve, with supporting mode of action information, will indicate a linear relationship. In these cases, the estimation of potential adverse effects (i.e., genomic damage) can be "read" directly from the dose-response curve. However, in many other cases a non-linear dose response will be observed and, for the most part, an approach to extrapolate from higher testing doses to lower doses consistent with human exposures will concentrate on selection of a PoD for the extrapolation.

After review of the data, and if the testing and the data are deemed appropriate, a quantitative analysis of the dose-response data can be conducted. Another GTTC work group previously provided guidance for the use of appropriate quantitative approaches for dose-response modeling for well-known genetic toxicity endpoints [Gollapudi et al., 2013; Johnson et al., 2014]. This guidance should be incorporated into the study design and the quantitative analysis. The major outcome of such dose-response modeling is the ability to determine a PoD from the observable effect range of the tested dose-response curve for extrapolation to lower expected and/or anticipated human exposures. This emphasis on the usefulness of the PoD approach was reinforced by expert discussions at the IWGT [MacGregor et al., 2015a].

These collective efforts examined several quantitative analysis approaches in common use for other toxicity endpoints for adaptability to genetic toxicity testing results. These included: (i) the no-observed-genotoxic-effect-level (NOGEL), which is the genotoxicity endpoint equivalent of the no-observed-effect-level (NOEL) used in toxicology; (ii) the benchmark dose (BMD), or benchmark concentration (BMC), which is the dose or concentration where a defined response, or benchmark response (BMR), is observed; and (iii) a statistically defined breakpoint analysis, referred to as the breakpoint dose (BPD), which is determined using a bi-linear dose-response model. For fuller descriptions of these approaches to quantitatively assess the results from genetic toxicity tests and how to apply them, see the publications from these working groups [Gollapudi et al., 2013; Johnson et al., 2014; MacGregor et al., 2015a; also, see Johnson et al., 2015].

Based on the strengths and weaknesses of each of these approaches, it was concluded that the order of preference for deriving the PoD is: BMDL > NOGEL > BPDL. The BMDL was determined to be the most robust and conservative and is recommended for general use as the PoD for genetic toxicity testing analysis [Gollapudi et al., 2013; MacGregor et al., 2015a]. Since BMD has proven its usefulness for other toxicity endpoints as well, we here propose to use the BMDL approach to determine the PoD (or PoDs) for estimating potential human risks associated with genomic damage.

Determine Expected/Anticipated Exposures

The primary purpose of an exposure assessment is to estimate dose to exposed persons, which is combined with chemical-specific dose-response data (usually from animal studies) in order to estimate risk [USEPA, 1992]. During the planning and scoping stage, the expected and/or anticipated human exposures to the substance(s) of interest have been identified. These initial findings help determine whether there are any possible exposures, projected routes, and sources of exposure, and whether the potential exposures provide enough concern to move forward with genetic toxicity testing. Once a decision is made to perform such testing, these initial findings no longer suffice. At this stage, an exposure assessment should also be planned.

While the testing of the substance in question is being conducted, an exposure assessment to estimate the likely human exposure can be concurrently developed. Exposure scenarios are modeled with the use of expected or default exposure factors [e.g., USEPA, 2011a; CDC, 2014]. These estimates are important for helping determine the extent of the risk to the population or subgroup of the population identified in the planning and scoping stage. Depending on the purpose of the risk assessment, the exposure assessment may need to emphasize certain areas in addition to quantification of exposure and dose, for example, the number of people exposed and the duration and frequency of exposure(s). Within the ILSI HESI RISK21 project, a four-stage, tiered approach to estimating exposure was developed [Embry et al., 2014]. In this approach, exposure estimates are constructed from information that is increasingly data-rich, ranging from limited information provided by physicochemical properties and route of exposure information to data from human biomonitoring studies [Embry et al., 2014]. The methodology and guidance for exposure assessment is outside the scope of this discussion. The reader is referred to more authoritative sources for the conduct of an exposure assessment [e.g., USEPA, 1992; Embry et al., 2014].

Estimate Candidate Regulatory Levels for Endpoints of Most Concern/Relevance

The PoDs from the testing data and the exposure estimates are combined to ascertain the possible risk from

any observed genomic damage. The two methods that have been explored to provide an indication of risk from potential genomic damage include the margin of exposure (MOE) approach and the reference dose (RfD) approach [Johnson et al., 2014; MacGregor et al., 2015b]. If a linear dose-response relationship is determined from the data, then the estimated risk can be "read" from the curve and associated with the exposure assessment. However, the MOE and RfD approaches will likely be utilized in most instances to estimate a potential regulatory level.

The MOE approach is one that can easily be used when examining genotoxicity data [EFSA, 2005]. Although originally proposed for genotoxic carcinogens, the MOE approach is easily adapted to characterizing any toxicity risk. The method, in which the PoD is compared to the actual or predicted human exposure detailed from the exposure assessment, is relatively straightforward. It may be the preferred approach as it directly incorporates estimated or actual human exposure information in the calculation (i.e., it is a ratio of the PoD to human exposure). The MOE does not actually provide a quantification of risk, but does provide perspective on a level of possible concern for decision making based on the magnitude of the ratio (see next section) [EFSA, 2005]. The MOE approach for genome-damaging substances is demonstrated in associated publications [Johnson et al., 2014, 2015].

The RfD (and reference concentration (RfC)) approach can provide quantitative information for use in risk assessments [USEPA, 2002]. The RfD (expressed in units of mg of substance/kg body weight per day) is defined as an estimate (with uncertainty spanning perhaps an order of magnitude) of a daily exposure to the human population (including sensitive subgroups) that is likely to be without an appreciable risk of deleterious effects during a lifetime [USEPA, 2015]. An RfD is determined by applying uncertainty factors (UFs) to the selected PoD to reflect and address limitations of the data, usually with linear extrapolation to lower doses. The concatenated UFs account for various sources of uncertainty. Typically, for most non-cancer endpoints, a composite or total UF of 100 is often derived by multiplying a factor of 10 for animal to human extrapolation and a factor of 10 for variability within human populations. These UFs are often default values, but can be adjusted if additional information is available. Such information may include incorporation of actual exposure and pharmacokinetic/metabolism characteristics, studies in multiple or more relevant animal species indicating an adjustment is appropriate, or the identification of vulnerable human population subgroups that need increased protection [for fuller discussion see USEPA, 2014b; MacGregor et al., 2015b]. Furthermore, chemical-specific adjustment factors (CSAFs; also known as data-derived extrapolation factors (DDEFs) for interspecies and intraspecies extrapolation) are sometimes available for interspecies differences and human variability and can be used in place of corresponding default UFs [WHO/IPCS, 2005; USEPA, 2014b]. If the dose-response analyses rely on animal-based studies versus human-based studies, then an allometric scaling (e.g., based on body weight$^{3/4}$) needs to be calculated as a basis for scaling toxicity data and values from animal models to human equivalents for human health risk assessments [see USEPA, 2011b for fuller discussion].

Risk Characterization

Risk characterization is a vital component of the risk assessment process [NRC, 1983]. To determine risk, the dose-response relationship for an identified hazard and human exposure information are combined to provide an estimate of potential harm to exposed persons. The characterization describes the findings of the testing and the implications those findings have on adverse health outcomes for exposed people, in particular, the role the genomic damage plays in human disease or adverse outcomes. The estimates for the risk based on the endpoints of most concern or relevance for humans are provided. The qualitative description and the quantitative estimates provide the information needed to address the goals and risk management questions posed during the planning and scoping process [USEPA, 2000].

Among the qualitative discussion of the risk characterization, key findings, results, and decisions are highlighted to support whether there is genotoxic risk to humans or not. Key points to highlight would include: the selection of appropriate genetic endpoints and target tissues, the selection of uncertainty factors and extrapolation methods used, the importance and use of information on MOA, toxicokinetics, metabolism, and predicted exposure information, including exposure biomarkers [taken from MacGregor et al., 2015b].

When considering the MOE approach for regulatory decision making, and any discussion for risk management interventions, the magnitude of the MOE ratio is evaluated and characterized. A larger MOE would be of less concern (e.g., MOE $\geq$ 10,000 may be considered to present minimal risk), though it would not preclude risk managers taking action to further reduce human exposure [EFSA, 2005]. A smaller MOE is likely to be of greater concern (e.g., MOE $<$ 100) and risk managers would want to consider risk management options to increase the ratio (most often by measures to reduce exposure). Other parameters for risk managers to consider when regarding the magnitude of the MOE include the severity of the effect, the MOA, the number of adverse effects observed, whether the observed effect(s) are from animal or human studies, the number of assumptions used in MOE estimations, the size of the affected population, and whether any susceptible subgroups have been identified [taken from Johnson et al., 2014].

A useful presentation for risk managers to consider is shown in Figure 3. Here the PoD is simply divided by

Assay Result	POD	Exposure	MOE
a	x	Ex1	100,000
b	y	Ex1	1000
c	z	Ex1	10
Etc.	…	…	…

Fig. 3. MOE presentation for risk managers.

the exposure estimate to calculate a margin of exposure (MOE) metric. This provides one with an idea of the magnitude of separation between a dose of the substance and when to expect a possible adverse outcome due to the genomic damage.

When considering the RfD approach for regulatory decision making, the decision generally hinges upon what uncertainty factors are applied to the PoD. The general default combination of factors (composite UF) usually is 100 as described above. But there are many considerations that need to be characterized to justify the use of 100, or to justify a different selection of uncertainty factors. These considerations include: species differences and allometric scaling, differences in absorption, distribution, metabolism and pharmacokinetics, differences in duration and/or frequency of exposure, severity of toxicity endpoint, variability among individuals, and uncertainty in the PoD [taken from MacGregor et al., 2015b]. When data are absent that would address a particular factor (e.g., differences in pharmacokinetics between tested animal species and humans), uncertainty factors should be applied to the predicted acceptable exposure level to account for the absence of data [USEPA, 2014b; MacGregor et al., 2015b]. Historically, composite UFs have not exceeded 3,000. Once the RfD is calculated, then risk managers can make decisions regarding the magnitude of genotoxic risk to exposed humans by comparing the expected or calculated exposure to the RfD. Any risk mitigation actions can then be considered.

Ultimately, the genotoxicity testing results and risk characterization should be evaluated against the goals and risk management questions developed during the planning and scoping. This will help focus the regulatory decision making squarely on the human risk from genomic damage and provide clear actions for reduction or elimination of any risk that may have been identified.

AUTHOR CONTRIBUTIONS

K.L.D, B.B.G., and M.L. provided leadership to the GTTC "Clean Sheet" workgroup discussions and led the preparation and writing of the manuscript. J.C.B., R.D.B., G.R.D., R.K.E, G.E.J., D.J.K., M.J.L., A.P.L., F.M., L.H.P., E.R., V.T., J.v.B., C.L.Y., and E.Z. all provided written contributions pertaining to their expertise and participated in the preparation and editing of the manuscript. J.Y.T. provided logistical, organizational, and editorial support for the project. All authors contributed to addressing the comments from the internal reviewers.

ACKNOWLEDGEMENTS

Authors thank all members of the "Clean Sheet" work group of the Genetic Toxicology Technical Committee (GTTC) of ILSI HESI for their valuable input during the various discussions on the flexible approach. They also thank Michael Cimino for his thorough review and valuable comments.

REFERENCES

Adewoye AB, Lindsay SJ, Dubrova YE, Hurles ME. 2015. The genome-wide effects of ionizing radiation on mutation induction in the mammalian germline. Nat Commun 6:6684.

Ankley GT, Bennett RS, Erickson RJ, Hoff DJ, Hornung MW, Johnson RD, Mount DR, Nichols JW, Russom CL, Schmieder PK, et al. 2010. Adverse outcome pathways: A conceptual framework to support ecotoxicology research and risk assessment. Environ Toxicol Chem 29:730–741.

Arnot JA, Mackay D, Webster E, Southwood JM. 2006. Screening level risk assessment model for chemical fate and effects in the environment. Environ Sci Technol 40:2316–2323.

Boobis AR, Cohen SM, Dellarco V, McGregor D, Meek ME, Vickers C, Willcocks D, Farland W. 2006. IPCS framework for analyzing the relevance of a cancer mode of action for humans. Crit Rev Toxicol 36:781–792.

Boobis AR, Doe JE, Heinrich-Hirsch B, Meek ME, Munn S, Ruchirawat M, Schlatter J, Seed J, Vickers C. 2008. IPCS framework for analyzing the relevance of a noncancer mode of action for humans. Crit Rev Toxicol 38:87–96.

CDC (Centers for Disease Control and Prevention, National Center for Health Statistics). 2014. The National Health and Nutrition Examination Survey (NHANES). URL: http://www.cdc.gov/nchs/nhanes.htm. Last accessed May 2016.

Chetty M, Rose RH, Abduljalil K, Patel N, Lu G, Cain T, Jamei M, Rostami-Hodjegan A. 2014. Applications of linking PBPK and PD models to predict the impact of genotypic variability, formulation differences, differences in target binding capacity and target site drug concentrations on drug responses and variability. Front Pharmacol 5:1–14. http://doi.org/10.3389/fphar.2014.00258. Last accessed May 2016.

Chi P, Allis CD, Wang GG. 2010. Covalent histone modifications—Miswritten, misinterpreted and mis-erased in human cancers. Nat Rev Cancer 10:457–469.

Cimino MC. 2006. Comparative overview of current international strategies and guidelines for genetic toxicology testing for regulatory purposes. Environ Mol Mutagen 47:362–390.

Dycaico MJ, Stuart GR, Tobal GM, de Boer JG, Glickman BW, Provost GS. 1996. Species-specific differences in hepatic mutant frequency and mutational spectrum among lambda/lacI transgenic rats and mice following exposure to aflatoxin B1. Carcinogenesis 17:2347–2356.

Dearfield KL, Auletta AE, Cimino MC, Moore MM. 1991. Considerations in the U.S. Environmental Protection Agency's testing approach for mutagenicity. Mutat Res 258:259–283.

Dearfield KL, Thybaud V, Cimino MC, Custer L, Czich A, Harvey JS, Hester S, Kim JH, Kirkland D, Levy DD, et al. 2011. Follow-up

actions from positive results of in vitro genetic toxicity testing. Environ Mol Mutagen 52:177–204.

Eastmond DA, Hartwig A, Anderson A, Anwar WA, Cimino MC, Dobrev I, Douglas GR, Nohmi T, Phillips DH, Vickers C. 2009. Mutagenicity testing for chemical risk assessment: Update of the WHO/IPCS Harmonized Scheme. Mutagenesis 24:341–349.

EFSA (European Food Safety Authority). 2005. Opinion of the Scientific Committee on a request from EFSA related to a harmonized approach for risk assessment of substances which are both genotoxic and carcinogenic. EFSA J 282:1–31.

Elespuru RK. 2011. Assessment of heritable genetic effects using new genetic tools and sentinels in an era of personalized medicine. Environ Mol Mutagen 52:253–263.

Embry MR, Bachman AN, Bell DR, Boobis AR, Cohen SM, Dellarco M, Dewhurst IC, Doerrer NG, Hines RN, Moretto A, et al. 2014. Risk assessment in the 21st century: Roadmap and matrix. Crit Rev Toxicol 44:6–16.

Gollapudi BB, Johnson GE, Hernandez LG, Pottenger LH, Dearfield KL, Jeffrey AM, Julien E, Kim JH, Lovell DP, MacGregor JT, et al. 2013. Quantitative approaches for assessing dose-response relationships in genetic toxicology studies. Environ Mol Mutagen 54:8–18.

Goodman JI, Augustine KA, Cunnningham ML, Dixon D, Dragan YP, Falls JG, Rasoulpour RJ, Sills RC, Storer RD, Wolf DC, et al. 2010. What do we need to know prior to thinking about incorporating an epigenetic evaluation into safety assessments? Toxicol Sci 116:375–381.

HPB (Health Protection Branch, Department of National Health and Welfare, Canada). 1993. The assessment of mutagenicity: Health Protection Branch mutagenicity guidelines. Environ Mol Mutagen 21:15–37.

ICCA (International Council of Chemical Associations). 2011. Guidance on Chemical Risk Assessment, 2nd ed. URL: http://www.icca-chem.org/ICCADocs/ICCA_GPS%20July2011_LowResWEB.pdf. Last accessed May 2016.

ICH (International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use). 2005. Detection of Toxicity to Reproduction for Medicinal Products & Toxicity to Male Fertility. Addendum dated 9 November 2000 incorporated in November 2005. ICH Harmonised Tripartite Guideline S5(R2). Geneva, Switzerland. URL: http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Safety/S5/Step4/S5_R2__Guideline.pdf. Last accessed May 2016.

ICH (International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use). 2011. Guidance on Genotoxicity Testing and Data Interpretation for Pharmaceuticals Intended for Human Use, S2(R1). Geneva, Switzerland. URL: http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Safety/S2_R1/Step4/S2R1_Step4.pdf. Last accessed May 2016.

ICH (International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use). 2014. Assessment and Control of DNA Reactive (Mutagenic) Impurities in Human Pharmaceuticals to Limit Potential Carcinogenic Risk. ICH Harmonised Tripartite Guideline M7. Geneva, Switzerland. URL: http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Multidisciplinary/M7/M7_Step_4.pdf. Last accessed May 2016.

Johnson GE, Soeteman-Hernandez LG, Gollapudi BB, Bodger OG, Dearfield KL, Heflich RH, Hixon JG, Lovell DP, MacGregor JT, Pottenger LH, et al. 2014. Derivation of point of departure (PoD) estimates in genetic toxicology studies and their potential applications in risk assessment. Environ Mol Mutagen 55:609–623.

Johnson GE, Slob W, Doak SH, Fellows MD, Gollapudi BB, Heflich RH, Rees BJ, Soeteman-Hernández LG, Verma JR, Wills JW, et al. 2015. New approaches to advance the use of genetic toxicology analyses for human health risk assessment. Toxicol Res 4:667–676.

Julien E, Boobis AR, Olin SS; ILSI Research Foundation Threshold Working Group. 2009. The Key Events Dose-Response Framework: A cross-disciplinary mode-of-action based approach to examining dose-response and thresholds. Crit Rev Food Sci Nutr 49:682–689.

LeBaron MJ, Rasoulpour RJ, Klapacz J, Ellis-Hutchings RG, Hollnagel HM, Gollapudi BB. 2010. Epigenetics and chemical safety assessment. Mutat Res 705:83–95.

Lewis DF, Ioannides C, Parke DV. 1998. Cytochromes P450 and species differences in xenobiotic metabolism and activation of carcinogen. Environ Health Perspect 106:633–641.

Liu J, Mansouri K, Judson RS, Martin MT, Hong H, Chen M, Xu X, Thomas RS, Shah I. 2015. Predicting hepatotoxicity using ToxCast in vitro bioactivity and chemical structure. Chem Res Toxicol 28:738–751.

MacArthur DG, Balasubramanian S, Frankish A, Huang N, Morris J, Walter K, Jostins L, Habegger L, Pickrell JK, Montgomery SB, et al.; 1000 Genomes Project Consortium. 2012. A systematic survey of loss-of-function variants in human protein-coding genes. Science 335:823–828.

MacGregor JT, Frötschl R, White PA, Crump KS, Eastmond DA, Fukushima S, Guérard M, Hayashi M, Soeteman-Hernández LG, Kasamatsu T, et al. 2015a. IWGT report on quantitative approaches to genotoxicity risk assessment. I. Methods and metrics for defining exposure–response relationships and points of departure (PoDs). Mutat Res 783:55–65.

MacGregor JT, Frötschl R, White PA, Crump KS, Eastmond DA, Fukushima S, Guérard M, Hayashi M, Soeteman-Hernández LG, Johnson GE, et al. 2015b. IWGT report on quantitative approaches to genotoxicity risk assessment. II. Use of point-of-departure (PoD) metrics in defining acceptable exposure limits and assessing human risk. Mutat Res 783:66–78.

Meek ME, Boobis A, Cote I, Dellarco V, Fotakis G, Munn S, Seed J, Vickers C. 2014. New developments in the evolution and application of the WHO/IPCS framework on mode of action/species concordance analysis. J Appl Toxicol 34:1–18.

Milic M, Frustaci A, Del Bufalo A, Sánchez-Alarcón J, Valencia-Quintana R, Russo P, Bonassi S. 2015. DNA damage in non-communicable diseases: A clinical and epidemiological perspective. Mutat Res 776:118–127.

Moskalev AA, Shaposhnikov MV, Plyusnina EN, Zhavoronkov A, Budovsky A, Yanai H, Fraifeld VE. 2012. The role of DNA damage and repair in aging through the prism of Koch-like criteria. Ageing Res Rev 12:661–684.

Müller L, Kikuchi Y, Probst P, Schechtman L, Sofuni T, Tweats D. 1999. ICH-harmonised guidances on genotoxicity testing of pharmaceuticals: Evaluation, reasoning and impact. Mutat Res 436: 195–225.

NRC (National Research Council). 1983. Risk Assessment in the Federal Government: Managing the Process. Washington, DC: The National Academies Press.

NRC (National Research Council). 2007. Toxicity Testing in the 21st Century: A Vision and a Strategy. Washington, DC: The National Academies Press.

NRC (National Research Council). 2009. Science and Decisions: Advancing Risk Assessment. Washington, DC: The National Academies Press.

OECD (Organisation for Economic Co-Operation and Development). 2005. Guidance Document on the Validation and International Acceptance of New or Updated Test Methods for Hazard Assessment. No. 34. ENV/JM/MONO(2005)14. Paris: OECD Publishing.

OECD (Organisation for Economic Co-Operation and Development). 2013. Guidance Document on Developing and Assessing Adverse Outcome Pathways. Series on Testing and Assessment, No. 184. ENV/JM/MONO(2013)6. Paris: OECD Publishing.

OECD (Organisation for Economic Co-Operation and Development). 2014a. Guidance Document for Describing Non-Guideline In Vitro Test Methods. Series on Testing and Assessment, No. 211. ENV/JM/MONO(2014)35. Paris: OECD Publishing.

OECD (Organisation for Economic Co-Operation and Development). 2014b. Document Supporting the WNT Decision to Implement Revised Criteria for the Selection of the Top Concentration in the In Vitro Mammalian Cell Assays on Genotoxicity (Test Guidelines 473, 476 and 487), ENV/JM/TG/RD(2014)20. Paris: OECD Publishing.

OECD (Organisation for Economic Co-Operation and Development). 2015. OECD Guidelines for the Testing of Chemicals, Section 4: Health Effects. Paris: OECD Publishing. URL: http://www.oecd-ilibrary.org/environment/oecd-guidelines-for-the-testing-of-chemicals-section-4-health-effects_20745788. Last accessed May 2016.

Patlewicz G, Simon TW, Rowlands JC, Budinsky RA, Becker RA. 2015. Proposing a scientific confidence framework to help support the application of adverse outcome pathways for regulatory purposes. Reg Toxicol Pharmacol 71:463–477.

Pottenger LH, Andrews LS, Bachman AN, Boogaard PJ, Cadet J, Embry MR, Farmer PB, Himmelstein MW, Jarabek AM, Martin EA, et al. 2014. An organizational approach for the assessment of DNA adduct data in risk assessment: Case studies for aflatoxin B1, tamoxifen, and vinyl chloride. Crit Rev Toxicol 44:248–291.

Priestley CC, Anderton M, Doherty AT, Duffy P, Mellor HR, Powell H, Roberts R. 2012. Epigenetics—Relevance to drug safety science. Toxicol Res 1:23–31.

Rasoulpour RJ, LeBaron MJ, Ellis-Hutchings RG, Klapacz J, Gollapudi BB. 2011. Epigenetic screening in product safety assessment: Are we there yet? Toxicol Mech Methods 21:298–311.

REACH (Registration, Evaluation, Authorisation and Restriction of Chemicals). 2007. European Parliament and Council Regulation (EC) No 1907/2006 (Corrigendum 29 May 2007) and Directive 2006/121/EC (Corrigendum 29 May 2007). Official Journal of the European Union L 396 of 30 December 2006.

Rosenbaum RK, Bachmann TM, Gold LS, Huijbregts MAJ, Jolliet O, Juraske R, Koehler A, Larsen HF, MacLeod M, Margni M, et al. 2008. USEtox—The UNEP-SETAC toxicity model: Recommended characterisation factors for human toxicity and freshwater ecotoxicity in life cycle impact assessment. Int J Life Cycle Assess 13:532–546.

Rosenbaum RK, Huijbregts MAJ, Henderson AD, Margni M, McKone TE, van de Meent D, Hauschild MZ, Shaked S, Li DS, Gold LS, et al. 2011. USEtox human exposure and toxicity factors for comparative assessment of toxic emissions in life cycle analysis: Sensitivity to key chemical properties. Int J Life Cycle Assess 16:710–727.

Simon TW, Simons SS Jr, Preston RJ, Boobis AR, Cohen SM, Doerrer NG, Fenner-Crisp PA, McMullin TS, McQueen CA, Rowlands JC. 2014. The use of mode of action information in risk assessment: Quantitative key events/dose–response framework for modeling the dose–response for key events. Crit Rev Toxicol 44: 17–43.

Snyder RD, Holt PA, Maguire JM, Trent JO. 2013. Prediction of noncovalent drug/DNA interaction using computational docking models: Studies with over 1350 launched drugs. Environ Mol Mutagen 54:668–681.

Thomas RS, Philbert MA, Auerbach SS, Wetmore BA, Devito MJ, Cote I, Rowlands JC, Whelan MP, Hays SM, Andersen ME, et al. 2013. Incorporating new technologies into toxicity testing and risk assessment: Moving from 21st century vision to a data-driven framework. Toxicol Sci 136:4–18.

Thomson JP, Moggs JG, Wolf CR, Meehan RR. 2014. Epigenetic profiles as defined signatures of xenobiotic exposure. Mutat Res 764–765:3–9.

UNECE (United Nations Economic Commission for Europe). 2015. Globally Harmonized System of Classification and Labelling of Chemicals (GHS, Rev.6). URL: http://www.unece.org/index.php?id=40514. Last accessed May 2016.

Uryga A, Gray K, Bennett M. 2016. DNA damage and repair in vascular disease. Annu Rev Physiol 78:45–66.

USEPA (U.S. Environmental Protection Agency). 1992. Guidelines for Exposure Assessment. Washington, DC. EPA/600/Z-92/001. URL: http://www.epa.gov/raf/publications/pdfs/GUIDELINES_EXPOSURE_ASSESSMENT.PDF. Last accessed May 2016.

USEPA (U.S. Environmental Protection Agency). 2000. Risk Characterization Handbook. Washington, DC. EPA 100-B-00-002. URL: http://www.epa.gov/spc/pdfs/rchandbk.pdf. Last accessed May 2016.

USEPA (U.S. Environmental Protection Agency). 2002. A Review of the Reference Dose and Reference Concentration Processes. Risk Assessment Forum. Washington, DC. EPA/630/P-02/002F. URL: http://www.epa.gov/raf/publications/review-reference-dose.htm. Last accessed May 2016.

USEPA (U.S. Environmental Protection Agency). 2011a. Exposure Factors Handbook. Washington, DC. EPA/600/R-09/052F. URL: http://www.epa.gov/ncea/efh/pdfs/efh-complete.pdf. Last accessed May 2016.

USEPA (U.S. Environmental Protection Agency). 2011b. Recommended Use of Body Weight3/4 as the Default Method in Derivation of the Oral Reference Dose. Final. Washington, DC: Risk Assessment Forum. EPA/100/R11/0001. URL: http://www.epa.gov/raf/publications/pdfs/recommended-use-of-bw34.pdf. Last accessed May 2016.

USEPA (U.S. Environmental Protection Agency). 2014a. Framework for human health risk assessment to inform decision making. Final. Washington, DC.: Risk Assessment Forum EPA/601/D-12/001. URL: http://www.epa.gov/raf/files/hhra-framework-final-2014.pdfwok. Last accessed May 2016.

USEPA (Environmental Protection Agency). 2014b. Guidance for applying quantitative data to develop data-derived extrapolation factors for interspecies and intraspecies extrapolation. Washington, DC: Risk Assessment Forum, Office of the Science Advisor. EPA/100/R-14/002. URL: http://www.epa.gov/raf/DDEF/pdf/ddef-final.pdf. Last accessed May 2016.

USEPA (U.S. Environmental Protection Agency). 2015. Integrated Risk Information System. Washington, DC. URL: http://www.epa.gov/iris/help_ques.htm#whatiris. Last accessed May 2016.

USEPA (U.S. Environmental Protection Agency). 2016. The Frank R. Lautenberg Chemical Safety for the 21st Century Act. Washington, DC. URL: https://www.epa.gov/assessing-and-managing-chemicals-under-tsca/frank-r-lautenberg-chemical-safety-21st-century-act. Last accessed May 2016.

Villeneuve DL, Crump D, Garcia-Reyero N, Hecker M, Hutchinson TH, LaLone CA, Landesmann B, Lettieri T, Munn S, Nepelska M, et al. 2014. Adverse outcome pathway (AOP) development. I. strategies and principles. Toxicol Sci 142:312–320.

Wambaugh JF, Setzer RW, Reif DM, Gangwal S, Mitchell-Blackwood J, Arnot JA, Joliet O, Frame A, Rabinowitz J, Knudsen TB, et al. 2013. High-throughput models for exposure-based chemical prioritization in the ExpoCast project. Environ Sci Technol 47: 8479–8488.

WHO/IPCS (World Health Organization/International Programme on Chemical Safety). 2005. Chemical-specific adjustment factors for interspecies differences and human variability: Guidance document

for use of data in dose/concentration-response assessment. (Harmonization Project Document No. 2). Geneva, Switzerland. URL: http://www.inchem.org/documents/harmproj/harmproj/harmproj2.pdf. Last accessed May 2016.

Wolters S, Schumacher B. 2013. Genome maintenance and transcription integrity in aging and disease. Front Genet 4:19.

Yauk CL, Bishop J, Dearfield KL, Douglas GR, Hales BF, Luijten M, O'Brien JM, Robaire B, Sram R, van Benthem J, et al. 2013. The development of adverse outcome pathways for mutagenic effects for the Organization for Economic Co-operation and Development. Environ Mol Mutagen 54:79–81.

Yauk CL, Aardema M, van Benthem J, Bishop J, Dearfield KL, DeMarini DM, Dubrova YE, Honma M, Lupski JR, Marchetti F, et al. 2015a. Approaches for identifying germ cell mutagens: Report of the 2013 IWGT workshop on germ cell assays. Mutat Res 783:36–54.

Yauk CL, Lambert IB, Meek ME, Douglas GR, Marchetti F. 2015b. Development of the adverse outcome pathway "Alkylation of DNA in male premeiotic germ cells leading to heritable mutations" using the OECD's users' handbook supplement. Environ Mol Mutagen 56:724–750.

Yauk CL, Buick JK, Williams A, Swartz CD, Recio L, Li HH, Fornace AJ Jr, Thomson EM, Aubrecht J. 2016. Application of the TGx-28.65 transcriptomic biomarker to classify genotoxic and non-genotoxic chemicals in human TK6 cells in the presence of rat liver S9. Environ Mol Mutagen 57:243–260.

You JS, Jones PA. 2012. Cancer genetics and epigenetics: Two sides of the same coin? Cancer Cell 22:9–20.

Zang C, Peng G. 2015. Non-coding RNAs: An emerging player in DNA damage response. Mutat Res 763:202–211.

Zeiger E, Gollapudi B, Aardema MJ, Auerbach S, Boverhof D, Custer L, Dedon P, Honma M, Ishida S, Kasinski AL, et al. 2015. Opportunities to integrate new approaches in genetic toxicology: An ILSI-HESI workshop report. Environ Mol Mutagen 56:277–285.