Received: 4 June 2023 | Revised: 15 September 2023 | Accepted: 16 December 2023

DOI: 10.1002/em.22582

Environmental and Molecular Mutagenesis | Environmental Mutagenesis and Genomics Society | WILEY

REVIEW

Interpretation of in vitro concentration-response data for risk assessment and regulatory decision-making: Report from the 2022 IWGT quantitative analysis expert working group meeting

Marc A. Beal[1] | Guangchao Chen[2] | Kerry L. Dearfield[3] | Min Gi[4] | Bhaskar Gollapudi[5] | Robert H. Heflich[6] | Katsuyoshi Horibata[7] | Alexandra S. Long[8] | David P. Lovell[9] | Barbara L. Parsons[6] | Stefan Pfuhler[10] | John Wills[11] | Andreas Zeller[12] | George Johnson[13] | Paul A. White[14]

[1]Bureau of Chemical Safety, Health Products and Food Branch, Health Canada, Ottawa, Ontario, Canada

[2]Centre for Nutrition, Prevention and Health Services, National Institute for Public Health and the Environment (RIVM), Utrecht, the Netherlands

[3]Retired from US Environmental Protection Agency and US Department of Agriculture, Washington, DC, USA

[4]Department of Environmental Risk Assessment, Graduate School of Medicine, Osaka Metropolitan University, Osaka, Japan

[5]Toxicology Consultant, Midland, Michigan, USA

[6]National Center for Toxicological Research, US Food and Drug Administration, Jefferson, Arkansas, USA

[7]Division of Genetics and Mutagenesis, National Institute of Health Sciences, Kawasaki, Kanagawa, Japan

Abstract

Quantitative risk assessments of chemicals are routinely performed using in vivo data from rodents; however, there is growing recognition that non-animal approaches can be human-relevant alternatives. There is an urgent need to build confidence in non-animal alternatives given the international support to reduce the use of animals in toxicity testing where possible. In order for scientists and risk assessors to prepare for this paradigm shift in toxicity assessment, standardization and consensus on in vitro testing strategies and data interpretation will need to be established. To address this issue, an Expert Working Group (EWG) of the 8th International Workshop on Genotoxicity Testing (IWGT) evaluated the utility of quantitative in vitro genotoxicity concentration-response data for risk assessment. The EWG first evaluated available in vitro methodologies and then examined the variability and maximal response of in vitro tests to estimate biologically relevant values for the critical effect sizes considered adverse or unacceptable. Next, the EWG reviewed the approaches and computational models employed to

List of Abbreviations: ADME, Absorption, Distribution, Metabolism, Excretion; AED, Administered Equivalent Dose; AUC, Area Under the Curve; ALI, Air-Liquid Interface; BMC, Benchmark Concentration; BMCL, Lower Confidence Limit Benchmark Concentration Level; BMD, Benchmark Dose; BMDL, Lower Confidence Limit Benchmark Dose Level; BMR, Benchmark Response; Cl_{int}, Intrinsic Hepatic Clearance Rate; C_{ss}, Steady-State Concentration in the Plasma; DS, Duplex Sequencing; EWG, Expert Working Group; GTTC, Genetic Toxicology Technical Committee; HBGV, Health-Based Guidance Values; HTTK, High-Throughput Toxicokinetics; Hprt, Hypoxanthine-Guanine Phosphoribosyl Transferase; IVIVE, in vitro to in vivo Extrapolation; IWGT, International Workshops on Genotoxicity Testing; LOGEL, Lowest-Observed-Genotoxic-Effect-Level; MN, Micronucleus/Micronuclei; MNvit, in vitro Micronucleus Assay; MOA, Mode of Action; NAM, New Approach Methodology; NOGEL, No-Observed-Genotoxic-Effect-Level; OECD, Organisation for Economic Co-operation and Development; PA, Pyrrolizidine Alkaloids; PBTK, Physiologically Based Toxicokinetics; PDE, Permitted Daily Exposure; PoD, Point-of-Departure; TG, Test Guideline; TGR, Transgenic Rodent; Tk, Thymidine Kinase; ToxValDB, Toxicity Value Database; Xprt, Xanthine-Guanine Phosphoribosyl Transferase.

Kerry L. Dearfield is now retired.

This manuscript reflects the views of the authors and does not necessarily reflect those of the Food and Drug Administration. Any mention of commercial products is for clarification only and is not intended as approval, endorsement, or recommendation.

[8]Existing Substances Risk Assessment Bureau, Healthy Environments and Consumer Safety Branch, Health Canada, Ottawa, Ontario, Canada

[9]St George's Medical School, University of London, London, UK

[10]Global Product Stewardship - Human Safety, Procter & Gamble, Cincinnati, Ohio, USA

[11]Genetic Toxicology and Photosafety, GSK Research & Development, Stevenage, UK

[12]Pharmaceutical Sciences, pRED Innovation Center Basel, Hoffmann-La Roche Ltd, Basel, Switzerland

[13]Swansea University Medical School, Swansea University, Swansea, UK

[14]Environmental Health Science and Research Bureau, Healthy Environments and Consumer Safety Branch, Health Canada, Ottawa, Ontario, Canada

Correspondence
Marc A. Beal, Bureau of Chemical Safety, Health Products and Food Branch, Health Canada, Ottawa, ON, Canada.
Email: marc.beal@hc-sc.gc.ca

Accepted by: H-J. Martus

provide human-relevant dose context to in vitro data. Lastly, the EWG evaluated risk assessment applications for which in vitro data are ready for use and applications where further work is required. The EWG concluded that in vitro genotoxicity concentration-response data can be interpreted in a risk assessment context. However, prior to routine use in regulatory settings, further research will be required to address the remaining uncertainties and limitations.

KEYWORDS
clastogen, genetic toxicology, mutation, new approach methodologies

1 | INTRODUCTION

A group of experts representing academia, government, and the private sector met in Ottawa, Canada (August 23–26, 2022) for the 8th International Workshop on Genotoxicity Testing (IWGT). One of the IWGT Expert Working Groups (EWGs) focused on Genotoxicity Dose–Response Analyses for Potency Comparisons and Risk Assessment. This EWG was divided into four subgroups collectively addressing several complementary topics: (1) Quantitative Dose–response Analysis & the Benchmark Dose (BMD) Approach; (2) Interpretation of In Vivo Dose–response Data for Risk Assessment and Regulatory Decision-making; (3) Interpretation of In Vitro Concentration-response Data for Risk Assessment and Regulatory Decision-making; and (4) Effect Severity and Interpretation of Genetic Toxicity Dose–response Data in a Human Health Context. This report discusses the outcomes of Subgroup 3 discussions, which focussed on critically evaluating the use of in vitro genotoxicity concentration-response data for risk assessment and regulatory decision-making.

The aim of Subgroup 3 was the discussion of various topics related to the development and use of an in vitro-dominant testing strategy for genotoxicity assessment; moreover, an evaluation regarding the utility of approaches for quantitative interpretation of in vitro concentration-response data for risk assessment and regulatory decision-making. The topics addressed included:

1. Mammalian assays for in vitro genotoxicity assessment.
2. Strategies for quantitative interpretation of in vitro concentration-response data, that is, PoD (point of departure) determination.
3. Toxicokinetic models to support interpretation of in vitro concentration-response data in an in vivo context.
4. Advanced in vitro systems for genotoxicity assessment.
5. Uncertainties and limitations associated with interpretation of in vitro concentration-response data in an in vivo context.
6. The utility of in vitro concentration-response data for human health risk assessment and regulatory decision-making.

After discussing the current state of knowledge, and research priorities for the development of risk assessment strategies based on quantitative interpretation of in vitro genotoxicity concentration-response data, the EWG achieved consensus on five key statements, presented below, regarding the utility of in vitro genotoxicity assessment systems for human health risk assessment.

2 | BACKGROUND

Historically, in vitro assays have played a major role in the identification of genotoxic hazards. To this day, the Salmonella reverse mutation assay (Ames test), first developed in the 1970s (Ames et al., 1973b), is among the most commonly used assays for genotoxicity assessment. However, in vitro genotoxicity assays have been used predominately for hazard identification of a compound as genotoxic or non-genotoxic, rather than for hazard assessment (where a PoD is derived) and/or risk assessment (where a PoD is compared to exposure levels) via quantitative interpretation of concentration-response data (White et al., 2020). Recent technological

advancements are now enabling the development and commercialization of higher-throughput and/or higher-content in vitro mammalian cell assays for genotoxicity assessment (Pfuhler et al., 2020; Wang et al., 2021). These assays are based on novel technologies that can support hazard and/or risk assessment without the use of animals. With the increased throughput, most in vitro assays allow for the testing of a broad concentration range, and the higher-content format provides mechanistic insight into a chemical's mode-of-action (MOA). Moreover, advancements in computational toxicology can be paired with in vitro results to quantitatively determine robust PoD estimates that can be interpreted in a human health context (Beal et al., 2021; Paul Friedman et al., 2020). Given that genotoxicity is increasingly being recognized as a bona fide toxicological endpoint (Heflich et al., 2020), these combined advancements provide a tremendous opportunity to develop an in vitro-dominant (i.e., animal-free) testing strategy for risk assessment of genotoxic chemicals.

In recent years, there has been an international shift away from animal toxicity testing, and as a result, the global scientific community has been working to develop robust non-animal alternative test methods. These alternatives to animal toxicity tests are often referred to as new approach methodologies (NAMs), which are broadly defined as any technology, method, and/or approach that can provide chemical toxicological information without the use of animals (ECHA, 2016; Stucki et al., 2022; US EPA, 2018). NAMs include in vitro toxicity assays as well as in silico prediction models. By this definition, NAMs are not new to genotoxicity assessment considering that several in vitro genotoxicity assays have Organisation for Economic Co-operation and Development (OECD) Test Guidelines (TGs; e.g., OECD TGs 473, 476, 487, and 490) (World Health Organization, 2020). In fact, the manuscripts first describing the Ames/Salmonella mutation assay (OECD TG 471) were published five decades ago (Ames et al., 1972a; Ames et al., 1972b; Ames et al., 1973a; Ames et al., 1973b). Furthermore, there are several quantitative structure–activity relationship models that exist for predicting chemicals with genotoxic potential (Hasselgren et al., 2019). Thus, in vitro assays and other NAMs have been a component of regulatory genotoxicity assessment for decades.

There have been several regulatory triggers for alternatives to animal tests. The earliest trigger came in 2003 when the 7th amendment to the European Union's Cosmetics Directive was adopted (Adler et al., 2011). This amendment set forth an immediate end to animal testing for cosmetic products entering the European Union marketplace. Furthermore, it also stipulated a ban by early 2009 on animal testing for cosmetic ingredients, with few exceptions for more complex toxicological endpoints. More recently, there have been additional efforts in North America to reduce the reliance on animal data for toxicity assessment. The U.S. Environmental Protection Agency has committed to a reduction and eventual elimination of toxicity testing using animals (Grimm, 2019) and the Canadian federal government is pursuing similar targets (Government of Canada, 2023). Furthermore, the U.S. Food and Drug Administration is encouraging the use of scientifically valid non-animal alternative test methods in regulatory submissions (FDA, 2017). Thus, there is a strong desire to design and deploy animal-alternative testing platforms that provide high-quality data across all toxicological endpoints, and there is an urgent need for consensus on how to quantitatively employ animal-free data for human health risk assessments.

3 | MAMMALIAN ASSAYS FOR IN VITRO GENOTOXICITY ASSESSMENT

There are numerous in vitro genotoxicity testing methods, and although the more traditional approaches with OECD TGs are the basis for most genotoxicity testing, researchers and chemical evaluators are exploring the potential of modern assays, which offer higher-throughput analysis while often providing more mechanistic information. Although the more recent in vitro assays for assessing genotoxicity lack OECD TGs, some are functionally related to existing in vivo OECD TGs; therefore, chemical evaluators are already familiar with the endpoints. These include in vitro alternatives to the in vivo TG (e.g., in vitro versions of TG488) or are in vitro TG compliant but offer a higher-throughput way of quantification (i.e., the TG487 compliant MicroFlow assay that employs high-throughput scoring by flow cytometry). Although some of the more modern in vitro assays have no direct relation to existing TGs, several have demonstrated a high level of sensitivity and specificity. This section provides a brief summary of in vitro mammalian cell assays that can be used for quantitative assessment of genotoxicity. The in vitro genotoxicity assays are separated into three categories: Gene Mutation Assays, Chromosomal Damage Assays, and Indicator Assays (Table 1).

3.1 | Gene mutation assays

Most of the in vitro mammalian mutation assays are forward mutation assays, meaning that they detect mutational changes that result in loss of function for a single gene. Forward mutation assays select mutants using a selective agent that enables the detection of a phenotypic change. For instance, in the presence of a selective agent, the thymidine kinase (Tk) gene mutation assay (TG 490) detects mutations at the heterozygous Tk gene (OECD, 2016d). The hypoxanthine-guanine phosphoribosyl transferase (Hprt) assay (TG 476) detects mutations at the hemizygous Hprt gene (OECD, 2016b), and the less frequently used Xprt assay (TG 476) detects mutations in the xanthine-guanine phosphoribosyl transferase (gpt) transgene in cells modified to carry a gpt transgene with a deleted Hprt gene. The cells that have a forward mutation and loss of function survive in the presence of the selective agent while wild-type cells accumulate toxic metabolites and are unable to proliferate. Mutant frequency is determined by quantifying clonal growth in the presence and absence of the selective agent.

Additional forward mutation assays initially designed for in vivo studies also have been adapted for in vitro analyses. The Transgenic Rodent (TGR) Somatic and Germ Cell Gene Mutation Assays [TG 488 (OECD, 2022b)] use transgenic mice or rats that contain either multiple copies of plasmids or phage shuttle vectors that have been integrated into their chromosomes (Boverhof et al., 2011; Lambert et al., 2005; Thybaud et al., 2003). Recovery of the shuttle vectors and their reporter transgenes in a bacterial host allows for the

TABLE 1 Summary of in vitro mammalian cell genetic toxicity assays.

Category	Full assay name	Endpoint	Cell type	Can be carried out in human cells?	Metabolism[a]	OECD TG	High(er)-throughput option?
Gene Mutation	In vitro mammalian cell gene mutation tests using the *HPRT* gene (i.e., HPRT assay)	*Hprt*/*HPRT* forward mutation	Most commonly CHO, CHL, or V79 (Chinese hamster), L5178Y (mouse lymphoma), or TK6 (human lymphoblastoid)	Yes	Must be tested with and without S9	TG476	No
Gene Mutation	In vitro mammalian cell gene mutation tests using the *gpt* gene (i.e., XPRT assay)	Transgene (*gpt*) forward mutation	AS52 (CHO-derived)	No	Must be tested with and without S9	TG476	No
Gene Mutation	in vitro mammalian FE1 transgene mutation assay (i.e., FE1 TGR assay)	Transgene (*lacZ*) forward mutation	FE1 cells (MutaMouse pulmonary epithelial cells)	No	Metabolically competent with respect to some cytochrome P450 isozymes and can be tested with S9	No[b]	No
Gene Mutation	Error-corrected sequencing of somatic mutations	Mutation in any selected genes/loci	Various	Yes	Can use metabolically competent cells or S9	No	Yes[c]
Gene Mutation	Pig-a gene mutation assay	*Pig-a* forward mutation	Various. Most commonly TK6, AHH-1, MCL-5 (human lymphoblastoid), L5178Y (mouse lymphoma)	Yes	Can use metabolically competent cells or S9	No[d]	Yes
Gene Mutation/ Chromosomal Damage	In vitro mammalian cell gene mutation tests using the thymidine kinase gene (i.e., mouse lymphoma assay, or MLA)	*Tk* forward mutation[e]	L5178Y (mouse lymphoma; primarily used) or TK6 (human lymphoblastoid)	Yes	Must be tested with and without S9	TG490	No
Chromosomal Damage	In vitro mammalian cell micronucleus test (i.e., MN_{vit})	Clastogenicity and aneugenicity	Various cell lines or human or other mammalian peripheral blood lymphocytes. Validated cell lines include CHO, V79, CHL/IU (Chinese hamster), L5178Y (mouse), and TK6 (human)	Yes	Can use metabolically competent cells or S9	TG487	Yes[f]
Chromosomal Damage	In vitro mammalian chromosomal aberration test	Clastogenicity	Various cell lines or primary cultures. Commonly CHO, CHL, V79 (Chinese hamster)	Yes	Can use metabolically competent cells or S9	TG473	No
Indicator	Alkaline Comet Assay (i.e., Comet)	DNA damage	Various	Yes	Can use metabolically competent cells or S9	No[g]	Yes[h]
Indicator	Multiflow™ assay	DNA damage[i]	Any, most commonly TK6 (human lymphoblastoid)	Yes	Can use metabolically competent cells or S9	No	Yes
Indicator	ToxTracker®	DNA damage[j]	6 mouse embryonic stem cell reporter cell lines (GFP)	No	Must be tested with and without S9	No	Yes

TABLE 1 (Continued)

Category	Full assay name	Endpoint	Cell type	Can be carried out in human cells?	Metabolism[a]	OECD TG	High(er)-throughput option?
Indicator	Prediscreen	DNA damage[k]	Various human cell lines—published data in HepG2 (liver), LS-174 T (colon), and ACHN (renal)	Yes	Can use metabolically competent cells or S9	No	Yes
Indicator	Tox21 quantitative high throughput screening assays	DNA damage	HCT-116 for p53 assay, HEK293T cells for ATAD5 assay, and CHO for γH2AX assay	Yes	Limited metabolic capability in cell lines	No	Yes
Indicator	Toxicogenomics biomarker for genotoxicity hazard (i.e., TGx-DDI)	DNA damage	TK6 (human lymphoblastoid), HepaRG (human hepatic)	Yes	Can use metabolically competent cells or S9	No	Yes[l]

[a]For assays that do not require the use of specialized cell lines, metabolic activation of test chemicals can be achieved by using metabolically competent cells or through the use of S9. The use of S9 is required for any assays that use specialized cell lines that lack metabolic competence.
[b]Complementary to OECD TG488—in vivo TGR.
[c]Considered high-throughput due to the ability to automate assay and multiplex samples.
[d]Complementary to OECD TG470—in vivo Pig-a gene mutation assay. Flow cytometry is used for high-throughput analysis.
[e]Can detect both gene mutations (large colonies) and chromosomal events (small colonies). For example, large deletions, chromosome rearrangements, and mitotic recombinations.
[f]Higher throughput versions use flow cytometry-based analysis.
[g]Complementary to OECD TG489—in vivo mammalian alkaline comet assay.
[h]Higher throughput versions available, for example, CometChip®. Various version of Comet Assay available, including some that use enzymes to detect specific damage types.
[i]Uses biomarkers (pH3, p53, γH2AX, cleaved PARP, polyploidy) to classify as aneugen/clastogen/pan-genotoxic/non-genotoxic.
[j]DNA damage: Bscl2, Rtkn; Oxidative stress: Srxn1, Blvrb; Protein damage: Ddit3; Generalized cellular stress: Btg2.
[k]Phosphorylation of H2AX and H3.
[l]Considered high-throughput due to the ability to automate assay and multiplex samples.

detection and scoring of mutant frequency. There are several different rodent models available with different transgenes used for mutant frequency determination (e.g., cII, gpt, lacI, lacZ, and spi). Over 20 in vitro assays have been developed from available TGR systems using cell types derived from a diverse array of tissues (e.g., embryo, liver, lung, mammary gland, skin, tongue); these in vitro versions have been used to assess the mutagenicity of over 150 agents (White et al., 2019). The in vitro TGR assays offer several benefits (White et al., 2019) in that they can inform potency ranking and read-across, they provide a high dynamic range, the diverse available systems can detect a variety of mutation types, metabolically competent cell lines can be used, and mutants can be sequenced (Beal et al., 2015; Besaratinia et al., 2012) to characterize mutation spectra, correct for clonal expansion, or characterize human-relevant mechanisms of carcinogenesis (Beal et al., 2020).

The Pig-a gene mutagenicity assay has in vivo versions with TGs [TG470, (OECD, 2022a)] and also has a complementary in vitro version. Pig-a (phosphatidylinositol glycan class A gene) is an endogenous mammalian gene that can serve as a mutation reporter gene in various species including humans. The Pig-a gene encodes a protein that is necessary for the biosynthesis of glycosylphosphatidylinositol (GPI) cell membrane anchors, and because the Pig-a gene is located on the X-chromosome, an inactivating mutation in the single functional gene copy can result in GPI anchor deficiency. In this assay, mutant frequency is enumerated using fluorochrome-conjugated antibodies targeting GPI-anchored proteins, where non-fluorescent cells indicate mutants and fluorescent cells indicate wild-type. In vitro versions of the Pig-a assay have been developed using mammalian cell lines (e.g., human TK6, AHH-1, MCL-5, and mouse L5178Y cells), and cell exposures to known mutagens yield results that are highly consistent with the expected positive response (Bemis & Heflich, 2019). Further validation efforts in additional cell lines could broaden the scope of the in vitro Pig-a assay as a high-throughput approach for genotoxicity assessment.

Error-corrected Next-Generation Sequencing (ecNGS) approaches are novel technologies that enable the direct quantification of low-frequency mutational events present among heterogeneous populations of DNA molecules; these technologies are gaining prominence for mutagenicity assessment (Marchetti et al., 2023). The application of NGS has historically been limited by the relatively high rates of technical errors (1 in 1000) relative to true de novo mutation frequencies (~1 in 100 million). Modern ecNGS approaches improve sequencing accuracy and mutation scoring by employing methods for error correction, including DNA molecule labeling with a unique molecular barcode to achieve error-corrected consensus sequencing. Duplex sequencing (DS) (Salk et al., 2018; Salk & Kennedy, 2020; Schmitt et al., 2012) is one of the ultra-high-accuracy ecNGS technologies that is emerging as an alternative approach to conventional assays based on detection of a mutant phenotype. In contrast to other mutation assays, ecNGS can be applied to detect and quantify mutations in any cell or tissue, from any species, and in any genomic location. ecNGS is less biased as it does not rely on the use of mutant selection that favors loss-of-function mutations (i.e., frameshifts and premature stop codons). Proof-of-principle studies have already demonstrated the utility of investigating mutagen exposures using DS in vitro (Cho et al., 2023; Wang et al., 2021). Parallel investigations using the TGR assay and DS have demonstrated that DS yields comparable results and detects the same types of mutations (LeBlanc et al., 2022; Valentine III et al., 2020). However, DS and other ecNGS methods are also able to detect mutations that remain undetected by the TGR assay (i.e., silent mutations), and ecNGS methods are able to measure the variability in chemical-induced mutagenesis across different regions of the genomes, thus offering broader coverage for effective mutagenicity assessment. Finally, as ecNGS can be run in parallel to any other toxicity assay, ecNGS technologies have the potential to augment genotoxicity assessments by complementing the standard toxicity test battery.

3.2 | Chromosomal damage assays

There are two in vitro mammalian cell chromosomal damage assays with OECD TGs; namely the in vitro mammalian cell chromosomal aberration test (TG 473; OECD, 2016a) and the in vitro micronucleus test (TG 487; OECD, 2014). The chromosomal aberration test is used to detect aberrations in chromosomes or chromatids resulting from clastogenic events (OECD, 2016a). The assay can be performed in established mammalian cell lines and primary cell cultures, but attention needs to be given to the growth ability, spontaneous chromosomal aberration frequency, p53 status, genetic stability, and DNA repair capacity of the cells (OECD, 2016a). At predetermined intervals after cells have been exposed to a test chemical, the cells are treated with a metaphase-arresting agent, harvested, stained, and cells in metaphase are analyzed using microscopy. The original chromosomal aberration test guideline was adopted in 1983 and thus, this standard assay has nearly 40 years of use in hazard/risk assessment. However, there are no recent publications where the in vitro version of this assay has been used for the quantitative interpretation of concentration-response data.

The in vitro mammalian cell micronucleus test (MN_{vit}) is a chromosomal damage assay used for the detection of micronuclei (MN), that is, small extra-nuclear bodies spatially isolated from the nucleus that contain damaged chromosome fragments and/or whole chromosomes. MN originating from acentric chromosome fragments are considered clastogenic events and MN originating from whole chromosomes are considered aneugenic events. MN represents DNA damage that has been transmitted to daughter cells, in contrast to chromosome aberrations, which are scored in metaphase and may not be transmitted. There are many recent examples in the literature where the MN_{vit} assay has been used for quantitative interpretation of in vitro concentration-response data (e.g., Kuo et al., 2022; Wills et al., 2016; Wills et al., 2021b). The precision of the assay has been bolstered, at least in assays conducted using cell lines, by a high-throughput adaptation that scores MN using flow-cytometry (e.g., the MicroFlow® assay; Bryce et al., 2010).

3.3 | Indicator assays

Genotoxicity assays that demonstrate evidence of adverse interactions with DNA without directly measuring mutations or chromosomal damage have been referred to as indicator assays. These assays can contribute to the weight-of-evidence in genotoxicity assessments and can provide mechanistic data that can be used to characterize the nature of the genotoxic responses that a chemical induces (e.g., distinguishing clastogens from aneugens). However, for hazard/risk assessment purposes, greater weight or preference is traditionally given to results from mutation or chromosomal damage assays as these assays measure permanent DNA changes.

The alkaline single-cell gel electrophoresis assay (i.e., comet assay) is a commonly used indicator assay, and an OECD TG exists for the in vivo version of the assay (OECD, 2016c). In the comet assay, microscopy is used to quantify DNA strand breaks, which following chemical exposure, are separated from intact DNA using single-cell gel electrophoresis. Under alkaline conditions (>pH 13), the comet assay is able to detect both single and double-strand breaks due to direct DNA interactions or resulting from alkali labile sites, as well as DNA strand breaks generated as intermediates during DNA excision repair (OECD, 2016c). The DNA damage detected by the comet assay is often transient or intermediate, in that the damage may be faithfully repaired and therefore not lead to a permanent genetic change (e.g., a mutation). Thus, the comet assay is categorized as an indicator assay as opposed to a chromosomal damage assay. An enhanced version of the comet assay, referred to as the CometChip® assay, has been developed to overcome the reproducibility and throughput issues associated with the traditional version (Ge et al., 2014; Weingeist et al., 2013). The CometChip® approach uses gravity to capture individual cells in a microwell to produce an orderly array of single cells as opposed to randomly dispersing cells in agarose. This approach allows for up to 300 cells to be arrayed into each well of a 96-well plate. The 96-well format of the CometChip® enables parallel processing for increased throughput, making the improved comet assay a more popular assay for quantitative analyses of DNA damage relative to the traditional method (e.g., Boyadzhiev et al., 2022).

There are several novel indicator assays that measure changes in DNA damage biomarkers; they can be employed for high-throughput and high-content genotoxicity assessment. These assays report changes in protein or gene expression levels indicative of DNA interactions, DNA damage, oxidative stress, replication inhibition, or other precursor events that might lead to mutations or chromosomal damage. Some of the more common mammalian indicator assays are described below.

The ToxTracker assay uses different mouse embryonic stem cell reporter cell lines that are preferentially responsive to genotoxic chemicals (Hendriks et al., 2012). Each cell line contains one green fluorescent protein (GFP)-tagged fusion protein for a selected biomarker gene. The detection of GFP fluorescence in the *Bscl2*-GFP and *Rtkn*-GFP cell lines is indicative of DNA replication inhibition and DNA damage, respectively (Hendriks et al., 2016). The ToxTracker assay has been applied to assess over 120 chemicals with different MOAs and quantitative analyses have demonstrated that ToxTracker-informed potency determinations are more informative than simply using the data qualitatively for hazard identification (Boisvert et al., 2023; Wills et al., 2021a). Extended validation studies of the ToxTracker assay are currently being conducted and the information will be used to draft a detailed review paper in support of TG development.[1]

The MultiFlow® assay is a high-throughput flow-cytometry-based indicator assay that uses multiplexed DNA damage response biomarkers to identify and mechanistically classify genotoxic agents. The biomarkers are histone H2AX phosphorylation (γH2AX) to indicate DNA double-strand breaks (clastogen marker), histone H3 phosphorylation (pH 3) to detect mitotic cells (aneugen marker), nuclear p53 as a general marker of DNA damage response, the proportion of 8n DNA content as a marker for polyploidy, and, optionally, cleaved Poly-ADP-ribose polymerase (PARP) as a marker of apoptosis (Avlasevich et al., 2021; Bryce et al., 2016). The multiplexed response data can be analyzed to categorize chemicals exhibiting clastogenic, aneugenic, or pan-genotoxic (both clastogenic and aneugenic) modes of action. Two different analytical approaches have been developed to inform the categorization: the first uses global evaluation factors or cut-off values derived from inter-laboratory investigations, and the second is based on multinomial logistic regression (Bryce et al., 2017). In the cross-laboratory study by Bryce et al. (2017), seven laboratories tested 84 reference chemicals representing the different modes of action to assess the precision and accuracy of the MultiFlow® assay. The aggregate results of 231 experiments demonstrated that assay sensitivity, specificity, and concordance with the expected MOA were 92% (22/24 genotoxicants were correctly identified). In addition to hazard identification, a quantitative approach has been developed to synthesize the multiple marker PoDs for each chemical into an aggregate score for potency ranking (Avlasevich et al., 2021).

PrediScreen is another high-throughput screening assay that relies on histone H2AX and H3 phosphorylation (pH3) markers to discriminate between clastogenic and aneugenic chemicals, respectively (Khoury et al., 2013, 2016a). PrediScreen uses an in-cell Western assay to quantify levels of γH2AX or pH3; genotoxicity is determined by dividing the relative fluorescent units for each marker by their respective controls. A validation study testing the γH2AX marker in HepG2 cells with 61 chemicals showed high sensitivity, with 75% (15/20) of genotoxic chemicals testing positive, and higher specificity, with 90%–100% (37/41; 22/22) of non-genotoxic compounds testing negative (Khoury et al., 2013). PrediScreen has been tested across different cell lines with varying levels of bioactivation (Khoury et al., 2016b), and results show that differential responses across these cell lines (i.e., negative in ACHN while positive in HepG2 or LS-174 T) can support the identification of genotoxic chemicals requiring metabolic activation. The higher-throughput format of PrediScreen allows for several experiments to be conducted simultaneously, enabling multiple replicates and concentrations to be tested in support of precise quantitative analyses (e.g., Khoury et al., 2016b).

The Tox21 program uses five quantitative high-throughput screening assays that measure elements of DNA damage/repair as part of its genotoxicity assessment capability (Hsieh et al., 2019).

Three of these assays use mammalian cell lines. The human HCT-116 cell line is used for a p53 activity assay (β-lactamase reporter assay), human HEK293T cells are used for the ATAD5 activity assay (luciferase reporter assay), and CHO cells are used for the γH2AX activity assay (fluorescent antibody binding). The other two assays use isogenic DT40 chicken lymphoblastoid cell lines that are knockouts for DNA repair proteins KU70/RAD54 or REV3 (differential cytotoxicity assays). These assays are reported to have low sensitivity for detecting genotoxicity, likely due to the preclusion of rat liver S9 mix for metabolic activation (Hsieh et al., 2019). This is supported by the observation that the incorporation of rat liver microsomes with the β-lactamase reporter assay enhances potency and/or efficacy for chemicals requiring metabolic activation (Ooka et al., 2022). Therefore, the assays in their current form cannot alone serve as alternatives to traditional assays without further adaptation to account for metabolic activation.

Transcriptomic biomarkers are proving to be powerful high-content tools for characterizing toxicological endpoints and quantifying the concentration at which toxicity begins to be observed (i.e., PoD determination). Transcriptomic biomarkers have been developed to support genotoxicity assessment with high accuracy for certain modes of action. One such biomarker is the in vitro Toxicogenomic DNA damage-inducing (TGx-DDI) biomarker that presently consists of a panel of 63 genes; the gene expression data can be used to distinguish between DNA damage-inducing (DDI) and non-DDI chemicals (Buick et al., 2020; Li et al., 2015, 2017). For each tested chemical, the relative changes in gene expression of the 63 genes are measured by either microarrays, RNA-seq, or quantitative RT-PCR (Cho et al., 2019); the chemicals are classified by comparing gene expression profiles against the profiles of 28 well-characterized model agents with known diverse (geno)toxic mechanisms. The biomarker was developed using TK6 cells, but it has also been successfully applied to HepaRG™ cell line (Buick et al., 2020). It was noted that the TGx-DDI biomarker alone has a limited ability to classify aneugens such as colchicine that affect microtubule assembly (Buick et al., 2020). The TGx-DDI biomarker is currently being assessed under the Biomarker Qualification Program at the US FDA.[2] Overall, the various indicator assays and their associated data are extremely useful for adding to the weight of evidence in hazard characterization.

4 | STRATEGIES AND CONSIDERATIONS FOR IN VITRO CONCENTRATION-RESPONSE MODELING

Critical analysis of methods and metrics for defining exposure-response relationships from genotoxicity data, with concomitant PoD determination, was the focus of a 2013 IWGT Working Group (MacGregor et al., 2015). The report by the Working Group on Quantitative Approaches to Genetic Toxicology Risk Assessment indicated that quantitative approaches should be well-defined and robust, conservative, transparent, and straightforward (i.e., easily calculated), and results should be linked to undesirable physiological effects (i.e., have interpretable biological meaning). The Working Group experts compared three analytical approaches meeting these criteria: (1) the no-observed-genotoxic-effect-level (NOGEL), that is, the highest dose where there is no statistically significant increase in genotoxic effect relative to vehicle control; (2) the in vivo benchmark dose (BMD) or in vitro benchmark concentration (BMC), that is, the dose/concentration eliciting a predefined change in the response compared to control, whereby the predefined level is referred to as the benchmark response (BMR) or critical effect size (CES); and (3) the breakpoint dose or threshold dose derived using a bi-linear model.

The 2013 Quantitative Approaches Working Group came to the consensus that BMD/BMC modeling is the preferred approach, and the BMD approach subgroup convened for the current IWGT meeting re-affirms support for this consensus (Haber et al., in preparation). The BMD/BMC approach is preferred because (1) analyses can be performed on studies with minimal data; (2) the entire dataset is used in deriving BMDs/BMCs; (3) the size of the effect is predefined; (4) the approach is amenable to covariate analysis; (5) modeling is less affected by experimental design and dose selection; (6) confidence limits of the BMD can be determined, which reflect the quality of the dose–response data; and (7) the benchmark dose lower confidence limit (BMDL) can be used as a conservative in vivo PoD, provided that the CES (or BMR) is sufficiently low. However, most in vivo OECD TGs are designed for hazard identification, and therefore recommend distributing animals among a small number of doses with 5–6 animals per dose in order to optimize statistical testing for pairwise testing; this is a suboptimal experimental design for the application of BMD modeling. For example, OECD TG 488 recommends three experimental doses plus vehicle control, with six animals per dose. Thus, dose–response data resulting from these study designs can be difficult to model if there is not at least one dose resulting in a response close to the predefined CES/BMR. Uncertainty in the shape of the dose–response curve can result in BMDs with large confidence intervals, and most hazard/risk assessments employing such data rely on a PoD metric such as the NOGEL or LOGEL (lowest-observed-genotoxic-effect-level), which do not provide an estimate of the uncertainty associated with the POD. Naturally, in vitro assays do not have the same experimental design constraints as in vivo studies, and higher-throughput in vitro assays can readily test a high number of concentrations in a single experiment (e.g., Avlasevich et al., 2021), generating ideal data sets for concentration-response modeling, and determination of precise BMCs (i.e., with narrow confidence intervals).

4.1 | Benchmark concentration modeling and critical effect size determination

To derive PoD values for chemical hazard/risk assessment, BMC analysis of in vitro concentration-response data must specify a CES. In general, for continuous data, the CES is defined as a fractional change in mean response compared to the mean background. Interestingly,

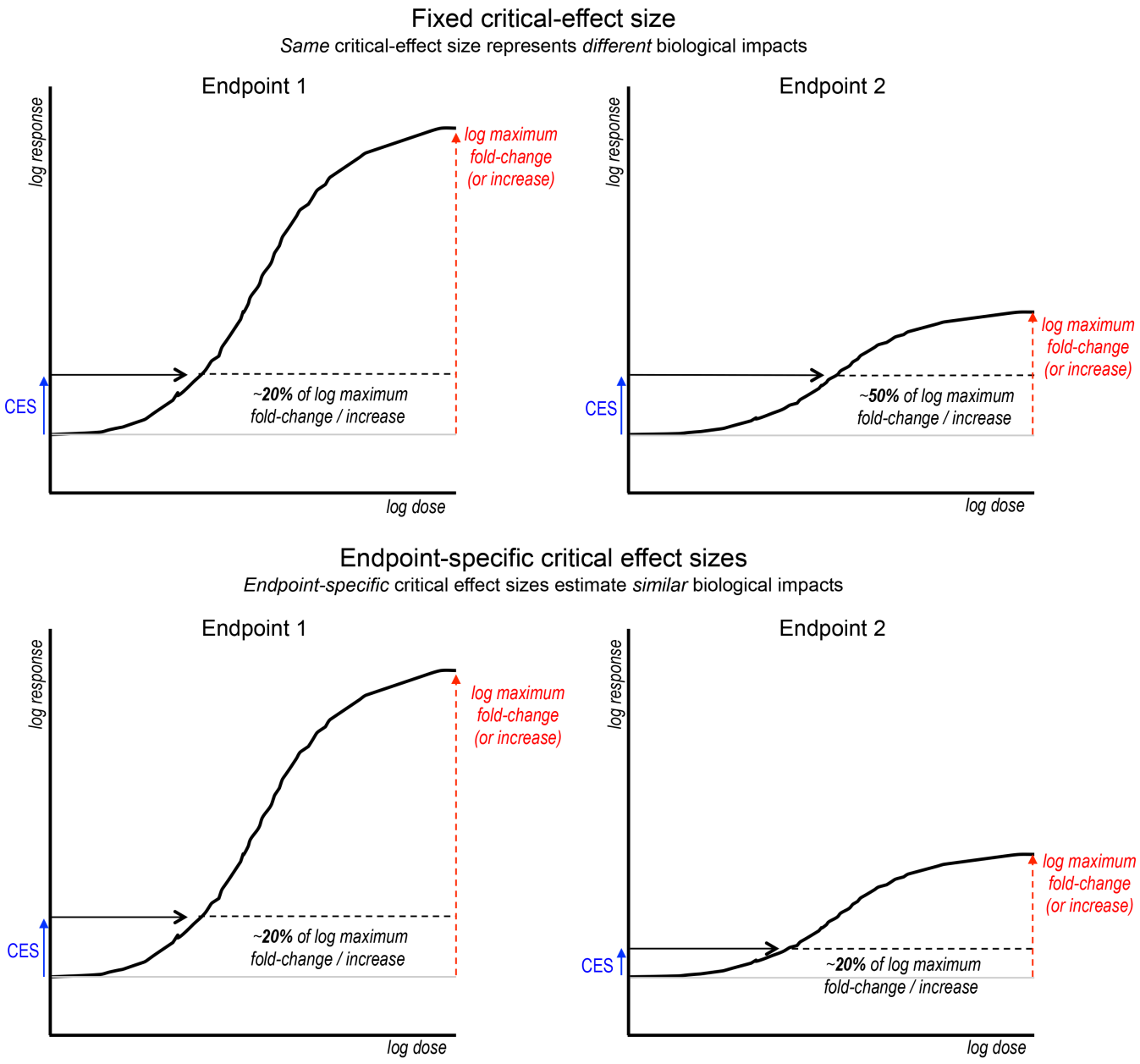

FIGURE 1 Illustration demonstrating the need for endpoint-specific critical effect size (CES) values. (Top) When considering two endpoints with different inducible ranges, the use of the same CES value (i.e., relative increase compared to the background response) leads to the determination of doses that elicit responses, which constitute different proportions of the endpoint-specific maximum. In essence, for an endpoint with a large maximum fold increase in inducible response, the magnitude of the response at the CES constitutes only a small fraction of the maximum for that endpoint. In contrast, for an endpoint with a smaller maximum fold increase in response, the magnitude of the response at the same CES constitutes a much greater proportion of the maximum for that endpoint. (Bottom) By scaling the CES to each endpoint's maximum fold-change, the magnitude of the response at the CES values represents equivalent fractions of the maximum fold-change for those endpoints.

CES values are the subject of considerable debate. The European Food Safety Authority (EFSA) has specified a default value of 5% for continuous in vivo dose–response data; however, this value can be modified based on biological relevance or statistical considerations (Barlow et al., 2009; EFSA Scientific Committee et al., 2017). Importantly, it is increasingly appreciated that a single CES cannot be employed for all toxicological endpoints (Slob, 2017) and that CES values should appropriately reflect the dynamic range of each endpoint. Figure 1 schematically explains this concept.

In its updated guidance on the use of the BMD approach for hazard/risk assessment, EFSA supports a "biologically relevant" CES for each endpoint, that is, as opposed to defaulting to a single conservative value (e.g. 5%) (EFSA Scientific Committee et al., 2022). However, although a "biologically relevant" CES for each endpoint is preferable

TABLE 2 Summary of Endpoint-Specific Critical Effect Size (CES) Estimates Based on Maximum Response and Within-Group Variation.

	Critical effect sizes			
	In vitro TGR (n = 64)	Flow-cytometry-based MN_{vit} (n = 135)	ToxTracker Bscl2 (n = 189)	ToxTracker Rtkn (n = 269)
Maximum response	60.7% (56.5%–65.4%)[a]	34.0% (32.5%–35.6%)	19.8% (19.5%–20.1%)	37.7% (37.0%–38.4%)
Within-group variation	28.2% (27.1%–29.3%)	21.8% (21.3%–22.5%)	21.0% (20.6%–21.4%)	32.4% (31.9%–33.1%)

[a]Range of calculated CES values based on the lower bound and upper bound of the 90% confidence interval of the maximum response fold-change or the within-group variation.

to a default CES, it is not clear how a biologically relevant CES should be determined. This issue is currently receiving considerable attention from multiple expert working groups (e.g., IWGT, HESI-GTTC, and EFSA). To provide a path forward, and a statistical framework for the determination of endpoint-specific CES values, Slob (2017) developed an "Effect Size theory" based upon analyses of a large number of datasets across numerous vivo endpoints. The theory proposes that the CES for an endpoint should be based on the endpoint's maximum response fold-change (rationale schematically outlined in Figure 1). However, in practice, the maximum fold-change of the response may be difficult to determine. This may be due to, for example, the fact that the dose groups utilized in many studies are not high enough to elicit a response that effectively corresponds to the endpoint maximum. For this reason, Slob (2017) highlighted an empirical relationship that permits estimation of the maximum fold-change for a given endpoint using variation in responses between animals in each dose group (i.e., the within-group variation) (Slob, 2017). This relationship permits the use of within-group variation for the determination of meaningful, endpoint-specific CES values for toxicological endpoints.

To demonstrate the application of the CES determination approach based on the aforementioned Effect Size Theory, endpoint-specific CES values for several exemplar in vitro genotoxicity endpoints were estimated using endpoint-specific maximum fold-change values and endpoint-specific within-group variation values. Effective application of the Effect Size Theory approach requires analysis of a large number of datasets for each endpoint under consideration, and the analyses conducted herein, which are described in Appendix A, were conducted using data made available to the study authors. The conducted analyses yielded endpoint-specific CES values ranging from ~20% to 61% (Table 2). Although these CES values cannot necessarily be considered definitive for the in vitro endpoints examined, the application of the Effect Size Theory provides values that confirm the necessity to determine endpoint-specific values that may deviate from the default value of 5%. Moreover, the analyses provide an effective illustration of how the statistical framework based on the aforementioned Effect Size Theory can be used for the determination of endpoint-specific CES values.

It is important to note that other data analysis approaches have been employed for the determination of endpoint-specific CES values. For example, Zeller et al. (2017) employed an approach that estimates endpoint-specific CES values using the variability exhibited in historical control distributions, that is, standard deviation of historical control values. When the analyses were based on the truncated distribution of historical control values (i.e., the distribution after exclusion of the uppermost 5% as outliers), the endpoint-specific CES values varied quite widely from a low of 34% for the MN assay with flow cytometry scoring to a high of 117% for the Pig-a mutagenicity assay in reticulocytes (Zeller et al., 2017). Similar analyses for the in vitro MN endpoint, which are based on data generated by more than 15 laboratories, yielded an average CES value several fold greater than 5%, generally similar to the in vivo situation (A. Zeller, personal communication and data presented at EMGS 2023). These analyses also confirm the general impression that a default CES value of 5% is indeed overly conservative (i.e., too low) for routine, quantitative interpretation of in vitro genotoxicity concentration-response data. Importantly, regardless of the approach used to determine endpoint-specific CES values, the analyses must employ a large number of concentration-response datasets. To help advance research in this area, and to achieve the ultimate goal of determining robust, endpoint-specific CES values for in vitro genotoxicity endpoints, cooperation from industry, government, and academic laboratories will be required to effectively provide a sufficiently large number of concentration-response datasets.

5 | TOXICOKINETICS MODELS SUPPORTING INTERPRETATION OF IN VITRO CONCENTRATION-RESPONSE DATA

One of the major challenges with employing in vitro data for toxicological risk assessment is the lack of animal/human equivalency. Pharmacokinetics or toxicokinetics models enable the in vitro genotoxic concentration (e.g., the BMC in μM) to be converted into a human-administered equivalent dose (AED) that is biologically appropriate for risk assessment (e.g., mg/kg bw/day). There are various toxicokinetic considerations that need to be considered before applying in vitro data as an alternative to in vivo data. Specifically, there is a need to know how much of the chemical makes it into the cell (in vitro disposition) and the expected absorption, distribution, metabolism, and excretion (ADME) of the chemical in animals (in vivo disposition). There are computational models and new experimental techniques that can be used to contextualize in vitro genotoxicity data, and these models are briefly discussed below.

5.1 | Computational models accounting for in vitro and in vivo disposition

Kinetic modeling is an important analytical tool; it can be used to analyze in vitro concentration-response data and quantitatively predict the dose (i.e., in mg/kg bw/day) that leads to an in vivo biological response. Several recent studies have demonstrated the utility of using kinetic models to derive surrogate PoDs (i.e., AEDs) for hazard and/or risk assessment based on in vitro data across a broad range of endpoints and chemicals (e.g., Paul Friedman et al., 2020). Thus, kinetic modeling already provides an avenue to perform risk assessments and in vivo equivalent potency ranking for chemicals that lack animal toxicity data but possess biologically relevant in vitro toxicity data.

High-throughput toxicokinetic (HTTK) models (Pearce et al., 2017) have been developed that enable in vitro to in vivo extrapolation (IVIVE) to be applied to in vitro bioactivity data across a range of organic chemicals. HTTK models offer simplified versions of physiologically based toxicokinetics (PBTK) models traditionally applied to pharmaceuticals; the lower complexity allows for wider application for assessing industrial and environmental chemicals without substantial chemical-specific toxicokinetic data, but this comes with a higher degree of uncertainty relative to bespoke PBTK models. The simplest HTTK model is the 3compartmentss model, where ss stands for steady state. The 3compartmentss model can be applied to greater numbers of chemicals than other models because it requires the fewest parameters and, therefore, is not limited by missing information for certain chemicals. The 3compartmentss model simulates the steady-state concentration in the plasma (C_{ss}) following infusion dosing, typically at a dose of 1 mg/kg bw/day. Given the assumption of a linear concentration response at steady state conditions, the dose required to reach a C_{ss} that is equal to the BMC for genotoxicity can be extrapolated or back-calculated, and this estimated dose serves as the AED for quantitative hazard/risk assessment.

HTTK models rely on kinetic parameters that can be measured in vitro using higher-throughput methods such as the intrinsic hepatic clearance rate (Cl_{int}) and the chemical fraction unbound in the plasma (Wetmore, 2015; Wetmore et al., 2012). In addition, recent work has established that in silico predictions for HTTK parameters provide suitable alternatives to in vitro data for HTTK applications (Beal et al., 2021; Pradeep et al., 2020). Specifically, modeled C_{ss} values appear to be relatively stable (Pradeep et al., 2020) even though there is a higher degree of uncertainty in the in silico modeled Cl_{int} values (Moreau et al., 2022). The main advantage of HTTK is it provides a high-throughput approach, requiring minimal data input, to derive AEDs from in vitro data.

AED derivation is an important step in the hazard characterization and potency ranking of chemicals using quantitative in vitro data, as hazard potential or potency rank varies greatly depending on whether the modeled human equivalent dose (i.e., AED) or in vitro POD (e.g., a BMC) is used in the assessment (Rotroff et al., 2010). Many of the assumptions used by HTTK are conservative, and AEDs derived from in vitro data are often more conservative (i.e., lower) than rodent in vivo PoDs used in risk characterization (Paul Friedman et al., 2020). Furthermore, HTTK uses a Monte Carlo simulator, known as the Virtual Population Generator for HTTK (HTTK-Pop; Ring et al., 2017), to account for inter-individual variation in the human population and estimate an AED that is on the lower end of the distribution (i.e., conservative AED based on 95th percentile C_{ss}). In the current context of use, AEDs are not meant to predict in vivo rodent PoDs or serve as a direct substitute, considering in vivo PoDs often have limitations related to dose spacing and insufficient number of doses to accurately inform BMD modeling. Rather, AEDs are useful for estimating an in vivo equivalent hazard level where key events leading to adverse outcomes are expected, and these levels are often less than traditional in vivo PoDs (Health Canada, 2021). Indeed, the utility of HTTK in providing conservative lower bound estimates of in vivo toxicity has been demonstrated using quantitative in vitro data covering various toxicological endpoints, including developmental neurotoxicity (Carstens et al., 2022), reproductive and developmental toxicity (Rajkumar et al., 2021a; Rajkumar et al., 2021b; Wang et al., 2022), and more recently, genotoxicity (Beal et al., 2022; Kuo et al., 2022).

The first study to critically examine the ability to apply HTTK and IVIVE to in vitro genotoxicity data examined MN_{vit} data for 292 chemicals (Kuo et al., 2022). A decision-tree-based pipeline was used to identify potential clastogens and aneugens based on increases in micronuclei and hypodiploidy, respectively. BMCs for micronuclei and hypodiploidy were based on CES values of 30% and 60%, respectively, and these defined effect sizes corresponded to approximately one standard deviation above background for each type of genotoxic event. HTTK modeling was used to convert the BMC_{30} and BMC_{60} values into AEDs for 137 clastogens and 14 aneugens, respectively. Only 33 clastogens had in vivo carcinogenicity or genotoxicity data from the EPA's Toxicity Value Database (ToxValDB) for comparison. There were no identified in vivo PoDs for the aneugens. Comparisons of the AEDs for clastogens with the lowest in vivo PoDs from ToxValDB revealed promising results for the application of HTTK to MN_{vit} data. Specifically, AEDs were lower than the corresponding lowest cancer PoDs for most chemicals (26/31; 83.9%); the AEDs were on average 14.2-fold lower than in vivo cancer PoDs. Acrylamide was the only chemical where the AED was two orders of magnitude higher than the in vivo cancer PoD, and thus not protective of human health. In this specific case, the available evidence indicates that the MOA underlying the cancer PoD was unrelated to genotoxicity (Kuo et al., 2022). AEDs were also shown to be lower than the in vivo genotoxicity PoDs for most chemicals (8/12; 66.7%); they were on average 2.6-fold lower than in vivo PoDs. The conclusion of this analysis was that HTTK applications to MN_{vit} data alone yielded AEDs that were one order of magnitude lower than in vivo cancer PoDs, and these AEDs were on the same order of magnitude as in vivo genotoxicity PoDs. The authors acknowledged that a single assay may not be sufficient to capture the different types of genotoxicity reported by the in vivo data, and that further work is needed to explore HTTK application to data from other genotoxicity assays.

In an effort to more thoroughly investigate HTTK applications to the various types of in vitro genotoxicity data, the Health and

Environmental Sciences Institute (HESI) Genetic Toxicology Technical Committee (GTTC) recently conducted a case study focused on established genotoxicants tested by in vitro assays capturing different MOAs (Beal et al., 2022). Specifically, the GTTC examined 31 reference chemicals that had data from the in vitro MicroFlow micronucleus (MN) assay, in vitro TGR (transgenic rodent) mutagenicity assay, and/or in vitro indicator assays (i.e., MultiFlow, PrediScreen, TOXTracker). BMC_{100} values were determined for each chemical, and HTTK was applied to derive AEDs. In total, 198 AEDs were derived for the 31 chemicals, and the AEDs were compared to in vivo genotoxicity PoDs. For this study, a total of 321 in vivo PoDs measured by various assays (i.e., micronucleus, fluorescence in situ hybridization, chromosomal aberration, aneuploidy, TGR, Pig-a, Hprt, Tk) were used in the comparison. In contrast to the previous case study where in vivo PoDs from ToxValDB were predominately NOGELs, the PoDs in the GTTC case study were predominately BMD_{100} values derived from dose–response data using the same CES and model for BMC_{100} derivation. Thus, capturing the same types of genotoxicity both in vitro and in vivo as well as modeling BMCs/BMDs using the same approach allowed for more appropriate comparisons.

The different comparisons by MOA in the GTTC case study revealed that the AEDs based on in vitro genotoxicity data were conservative relative to in vivo PoDs for most chemicals. AEDs based on in vitro TGR data from 13 chemicals showed that most chemicals (8/13; 62%) had lower AEDs relative to PoDs from in vivo gene mutation assays (3.7-fold lower on average). AEDs based on MicroFlow data were on average 7.3-fold lower than in vivo micronucleus PoDs, and overall, most chemicals had lower AEDs than in vivo PoDs (9/12; 75%). The in vitro MicroFlow results were more conservative than the results from the previous case study by Kuo et al. (2022) that did not distinguish in vivo endpoints and did not use BMDs as in vivo PoDs. AEDs based on aneuploidy markers (polyploidy or pH 3) derived for four known aneugens revealed that the AEDs were lower than in vivo PoDs, as measured by relevant assays, for all the aneugens tested (90-fold lower on average). AEDs derived from DNA damage indicator assays (γH2AX, p53, Bscl2, and Rtkn) were closely aligned with in vivo PoDs from all genotoxicity endpoints. For most chemicals (58%–69%), the AEDs from each indicator assay were slightly lower than the in vivo PoDs. On average the AEDs based on the DNA damage biomarkers were 3.5- to 7.9-fold lower than in vivo PoDs. Collectively, the results show that the combined approach of applying HTTK and BMC modeling to in vitro genotoxicity data provides conservative surrogate PoDs for most chemicals that could be applied in risk assessment activities where 2- to 100-fold conservatism is acceptable.

There were some chemicals in the GTTC case study that had AEDs consistently higher than in vivo PoDs regardless of MOA. N-ethyl-N-nitrosurea (ENU) had the highest AED/PoD ratio in the case study. The authors noted that ENU is highly unstable at a neutral pH, and therefore, degradation processes in the assay system were unaccounted for before HTTK application. Likewise, it was recognized that the AED for 1,2-dibromo-3-chloropropane (DBCP) may be overestimated as DBCP is a volatile chemical, and vaporization during in vitro exposures may have limited the amount of chemical making it into the cells. These results highlight that the HTTK approach by itself may not be appropriate for some chemicals. In certain cases, more complex models or other considerations in data interpretation may be required when performing IVIVE.

As highlighted by the GTTC case study, a major challenge hindering the application of IVIVE and the use of in vitro data for potency ranking and risk assessment is the fact that exposure concentrations are rarely verified. It is commonly assumed that the nominal concentration is an adequate estimation of cellular exposure and that chemical concentration remains stable throughout the duration of the in vitro exposure (Schirmer, 2006; Stadnicka-Michalak et al., 2021). However, there are several properties of the in vitro test system that impact the in vitro disposition of the chemical and can reduce the free chemical concentration to which cells are exposed. Specifically, the type of plastic coating the vessel walls, type of adhesive covering plate wells, plate/well size, media volume and type (including presence/concentration of serum and other macromolecules in media), headspace, and cell seeding density are some examples of test system properties affecting the free chemical concentration (Armitage et al., 2014; Stadnicka-Michalak et al., 2021). In addition, physicochemical properties such as volatility, solubility, and ionization state have a large impact on free chemical concentration (Armitage et al., 2021). Thus, validated computational models that account for these issues, and consider in vitro disposition prior to IVIVE, should increase the accuracy of IVIVE implementation.

There are multiple computational models that can be used to predict free concentration based on physicochemical properties and study-specific parameters dependent on the in vitro conditions. For example, the Armitage mass-balance model originally published in 2014 (Armitage et al., 2014) and recently updated (Armitage et al., 2021) estimates the mass distribution of chemicals in different in vitro test systems. The benefits of applying these models to in vitro bioactivity data have been demonstrated previously. For example, a study by Casey et al. (2018) compared in vitro estrogen receptor activity to in vivo estrogenic effects based on uterotrophic studies in rodents. The researchers found that the predictive performance of IVIVE improved when the Armitage model was used to adjust the in vitro bioactivity PoD used in the extrapolation. Further work is needed to determine how the Armitage model or other in vitro disposition models can enhance the application of IVIVE to the quantitative interpretation of genotoxicity dose–response data.

6 | ADVANCED IN VITRO SYSTEMS FOR GENOTOXICITY ASSESSMENT

Most in vitro genotoxicity approaches use two-dimensional (2D) cellular cultures that present several limitations or technical challenges (e.g., uncertain metabolism, lack of cell-to-cell interactions). More complex in vitro testing systems considered to be "in vivo-like" hold the potential of reducing the reliance on animal testing by replicating human physiology in toxicity testing. The extensive progress in 3D culture models representing major routes of exposure including systemic, dermal, and inhalation has been described in a previous IWGT report by the "Use of 3D Tissues in Genotoxicity Testing"

Working Group (Pfuhler et al., 2020). The Working Group concluded that 3D tissue models complement classical 2D cell culture models by offering more in vivo-like behavior for parameters including cell viability, proliferation, differentiation, morphology, and gene/protein expression. Currently, 3D models are more technically challenging, more expensive, lower throughput, few labs have experience working with 3D tissue cultures, and there is no consensus on exposure or culture protocols. Thus, the 3D tissue-based genotoxicity assays may serve as second tier assays to evaluate positive results detected using standard 2D assays. The aforementioned 3D tissue Working Group also recommended continued development of robust protocols for 3D models and confidence building using a set of validation chemicals. This section describes recent progress in the application of 3D models in genotoxicity testing and the potential of these complex in vitro systems for quantitative assessment of genotoxicity.

Many genotoxic compounds require metabolic activation, and thus, the liver's role in metabolism/bioactivation makes it a suitable tissue for genotoxicity testing. Indeed, two-dimensional hepatocyte cultures have been useful models for genotoxicity testing. However, the advancements in 3D liver spheroids have increased the ability to simulate in vivo conditions of in vitro test conditions, and they more appropriately reflect hepatocyte differentiation, longevity, gene expression profiles, metabolic competency, and overall functionality relative to 2D cultures. Recent studies with chemicals requiring metabolic activation have compared the level of induced genotoxicity in 2D and 3D HepaRG cultures using the MN (Seo et al., 2023; Shah et al., 2018) and CometChip® assays (Barranger & Le Hégarat, 2022; Seo et al., 2022). Shah et al. (2018) demonstrated that the levels of CYP1A1 expression following benzo[a]pyrene (BaP) exposure, and CYP1A2 expression following 2-amino-1-methyl-6-phenylimidazo [4,5-b]pyridine (PhIP) treatment, were higher in 3D HepaRG spheroids compared to 2D cultures. Thus, these treatments resulted in greater MN induction in 3D HepaRG cultures relative to 2D, likely due to higher levels of genotoxic metabolites. Similarly, Barranger and Le Hégarat (2022) assessed seven chemicals for genotoxicity in 3D and 2D HepaRG cultures using CometChip®, and five of the chemicals tested required metabolic activation (i.e., BaP, cyclophosphamide [CPA], 7,12-dimethylbenz[a]anthracene [DMBA], 2-acetylaminofluorene [2-AAF], and acrylamide [ACR]). BMC modeling was performed using a CES of 5%, 10%, or 20%; relative potency orders (i.e., potency ranking) of the chemicals were the same across CES. In the chemicals requiring metabolic activation, there was a tendency for lower BMCs (i.e., higher potency) for treatments conducted with the 3D spheroids compared to 2D hepatocytes. Moreover, 2-AAF was only positive in 3D HepaRG cells, possibly due to higher levels of CYP1A2 leading to increased 2-AAF bioactivation. A larger study by Seo et al. (2022) evaluated 34 test compounds in 2D and 3D HepaRG cells using CometChip®. These chemicals included 8 direct-acting agents, 11 genotoxic agents that require metabolic activation, and 15 chemicals that show differing responses in vitro and in vivo. There were 11 chemicals that tested positive in both 2D and 3D cells. BMC_{50} modeling determined that six chemicals had overlapping confidence intervals between 2D and 3D cultures (ENU, etoposide, MMS, ACR, BaP, 3-chloro-1,2-propanediol); the other five chemicals had lower BMC_{50} values in 3D cultures (4-nitroquinoline 1-oxide, DMBA, cisplatin, CPA, and dimethylnitrosamine). A more recent evaluation of the same set of chemicals using the MN assay in 2D and 3D HepaRG demonstrated that MN also has higher sensitivity in 3D models for detecting genotoxicity for the 11 agents requiring metabolic activation (Seo et al., 2023). Overall, the improved sensitivity for 3D liver spheroids to detect genotoxicity induced by chemicals requiring metabolic activation highlights the need to pursue the use of 3D liver models further in quantitative applications.

Human 3D airway or lung models have been developed to simulate inhalation exposures. These models consist of fully differentiated and functional epithelium and enable exposure to air through culturing at an air-liquid interface (ALI). Organotypic ALI airway cultures have been evaluated for their ability to detect genotoxic responses. For example, an ALI tissue model derived from normal human primary bronchial epithelial (NHBE) cells was exposed to 4-(methylnitrosamine)-1-(3-pyridyl)-1-butanone (NNK) and evaluated using the alkaline comet assay (Qin et al., 2019). The ALI cultures showed increased cytochrome P450 expression and higher bioactivation of NNK when compared to undifferentiated NHBE cells in 2D culture. Furthermore, the ALI cultures showed an increase in DNA damage at both concentrations tested (100 and 500 μM), whereas undifferentiated NHBE cells in 2D culture required the presence of exogenous liver S9 to detect a genotoxic response. In a more recent study, ALI cultures were used to investigate the genotoxicity of a 28-day treatment with ethyl methanesulfonate (EMS) using CometChip® and DS assays (Wang et al., 2021). The CometChip® results indicated that all concentrations of EMS-induced DNA damage and the damage manifestation were both time- and concentration-dependent. Furthermore, a robust concentration-response curve for mutation was detected using DS, and the mutation spectra revealed signatures typical of DNA alkylating agents. Together these studies provide evidence that ALI cultures can be used as robust culture systems for investigating the genotoxic effects of inhaled chemicals; further exploratory studies will help to determine how data from ALI cultures can be used quantitatively in hazard evaluations.

Reconstructed human skin (RS) models have been studied much more extensively than other organotypic models and are in a more advanced state of validation with over a decade of research. In the aforementioned IWGT report on 3D models (Pfuhler et al., 2020), the expert group concluded that 3D skin models were sufficiently validated for use with comet and micronucleus assays and that the development of individual OECD TGs should be pursued. Validation studies using 56 coded chemicals tested by 3D skin comet and RS MN demonstrated that these approaches are highly predictive of expected genotoxicity based on in vivo results, and these results were transferable across many laboratories. As demonstrated using other 3D tissue models, studies show that 3D RS exposures produce different results than 2D-equivalent exposures. For example, Wills et al. (2015) investigated exposures to SiO2 nanoparticles and found significant increases in genotoxicity using 2D models. In contrast, the 3D models showed no penetration or cell uptake of the nanoparticles; thus, there was no impact on DNA damage or cell viability. These results highlight the

protective nature of the RS 3D cellular microarchitecture and the importance of using 3D models in interpreting toxicological potential using skin exposures. It is believed that the development of TGs for RS models will increase the justification for applying these models quantitatively to assess genotoxicity and efforts for TG development are currently ongoing.

The emergence of more complex and transformative microphysiological systems (MPS), consisting of interacting organs-on-a-chip (OOC) using microfluidics and human cells, is further bridging the gap between in vitro and in vivo toxicity assessment (Ingber, 2022; Ronaldson-Bouchard & Vunjak-Novakovic, 2018; Wikswo, 2014). These models can recapitulate human physiology by mimicking several types of interactions involving chemicals, cells, matrix materials, tissues, and organoids. Although more technically challenging, MPS systems can offer a more realistic test environment equipped with heterogeneous cells growing in 3D extracellular matrices, in vivo-like tissue perfusion, microfluidic connections (static, unidirectional, or recirculated media) across integrated OOCs, and the appropriate biomimetic (i.e., mechanical and electrical) stimulation. Through the integration of multiple OOCs, MPS can simulate both the target organ and the means by which the chemical enters the system to ultimately reach the target site.

Although MPS systems are relatively new with respect to genotoxicity assessment, there is evidence that laboratories are starting to validate MPS for genotoxicity screening. For example, in the book of abstracts from the 52nd annual meeting of the Environmental Mutagenesis and Genomics Society (EMGS, 2021), Zanoni et al. reported a study that investigated the application of a metabolically competent liver-on-a-chip (LOC) system (endothelial cells plus HepaRG) co-cultured with TK6 cells suspended in transwells. Their preliminary data highlighted the system's ability to detect genotoxicity induced by direct and metabolically activated genotoxicants by assaying MN in TK6 cells, DNA damage by Comet assay in the LOC, and mutations by DS in the LOC. These preliminary findings are an early indication of the promise that MPS holds for genotoxicity assessment; the lack of published data using MPS to assess genotoxicity highlights the need for more focus in this area. It is anticipated that as 3D cultures and MPS technologies advance, there will be more chemicals evaluated using these approaches, and safety evaluations will be relying more on these types of data (Baran et al., 2022). As researchers build confidence in the application of genotoxicity assessments in in vitro models that are more in vivo-like, we anticipate that there will be opportunities to apply the data quantitatively in chemical safety evaluations.

7 | UNCERTAINTIES AND LIMITATIONS OF IN VITRO GENOTOXICITY CONCENTRATION-RESPONSE DATA AND INTERPRETATION

The Working Group discussed the uncertainties and limitations associated with in vitro genotoxicity data that need to be addressed to increase the confidence in quantitative in vitro testing strategies for applications beyond hazard assessment, that is, risk assessment. One of the main uncertainties is that it is difficult to establish the *exposure scenario equivalency* of the in vitro situation relative to the in vivo situation. Further work is required to determine if a short-term in vitro study (hours to days) can adequately capture the biology of a chronic in vivo study (weeks to years). Likewise, there are many uncertainties related to *metabolism* in the in vitro test systems. In vitro toxicity experiments do not routinely characterize the metabolic processes and metabolites formed during in vivo chemical assessment. Therefore, there is uncertainty as to whether the test systems are able to reproduce the same metabolites and/or metabolic processes occurring following human exposures, or if the levels are consistent with in vivo expectations. The uncertainties associated with metabolism also impact the uncertainties surrounding *toxicokinetics*. The toxicokinetic models used as part of IVIVE are sometimes limited in complexity, and the generic models used to date may be missing important organ compartments, are using overly conservative assumptions (e.g., complete absorption), and could be missing key kinetic processes such as extrahepatic metabolic elimination, renal reabsorption, and/or transporter-mediated kinetics. Additionally, high-throughput toxicokinetic models only model the parent compound, and therefore, more complex physiologically-based kinetic (PBK) models that model metabolite concentrations may be necessary to carry out or improve IVIVE for compounds requiring bioactivation. Lastly, there are *idiosyncratic or unique cell system-specific responses* that result from differences in cell growth rate, karyotype, genetic stability, p53 status, and DNA repair capacity. Carefully planned case studies that simultaneously look at in vitro and in vivo effects will be required to quantify the different uncertainties.

It is agreed that additional studies that compare in vitro genotoxicity results with in vivo data will allow for a better understanding and potentially reduce the uncertainties associated with in vitro studies. However, a major limitation associated with these comparisons is that the *available in vivo data have notable data gaps* that make confidence-building or validation challenging. For instance, the studies applying IVIVE to genotoxicity data (Beal et al., 2022; Kuo et al., 2022) highlighted that there was limited overlap in chemicals and/or endpoints tested both in vitro and in vivo. Furthermore, the in vivo studies used a more limited dose range, negatively affecting the precision of BMD modeling as demonstrated by the broad range of PoDs for each chemical that spanned orders of magnitude in some cases. These challenges are not limited to genotoxicity data as demonstrated in the study by Pham et al. (2020), which determined there is a high level of variability associated with in vivo data across 2724 toxicology studies involving 563 chemicals (Pham et al., 2020). Thus, there is a need for refined expectations when evaluating whether an in vitro derived PoD is concordant with in vivo expectations. Based on their analyses, Pham et al. estimated that a reasonable prediction for an in vivo PoD of 1 mg/kg using in vitro data would be 0.08–10 mg/kg (i.e., within approximately 1 order of magnitude). Another limitation for conducting in vitro studies and using the data in a regulatory setting is the lack of *standardization*. The development of guidelines is often outpaced by the rapid evolution of in vitro assays, making guideline development a challenge for novel approaches. Overall, when designing

in vitro genotoxicity studies, researchers will need to think incrementally about assay selection, experimental design and conduct to avoid bias, data generation, and concentration-response modeling, all while working closely with evaluators to ensure optimal use of data in quantitative risk assessments.

8 | POTENTIAL APPLICATIONS OF IN VITRO GENOTOXICITY DATA

The Working Group evaluated the potential hazard and risk assessment applications for which the quantitative interpretation of in vitro genotoxicity data could be used. These potential applications include potency ranking based on BMC confidence intervals, screening and prioritization of data-poor chemicals for further evaluation, and deriving health-based guidance values (HBGVs) in quantitative risk assessments. Each application was assessed on its readiness for routine use and whether further work is required before in vitro concentration-response data, on its own, could be used in the application.

There are already several notable and recent examples where BMC modeling has been applied in order to assess the potency of groups of chemicals. For example, Wills et al. (2016) used BMC modeling to compare potencies of chemicals across multiple factors. Specifically, a computational BMC covariate approach was used to permit the combined analysis of multiple concentration-response data sets that were differentiated by covariates including compound, cell type, or exposure regimen. The covariate approach was applied to the separate analyses of chemicals with different modes of action including in vitro micronucleus data for ionizing radiation, a set of aneugens, two mutagenic azo dyes, and one topoisomerase II inhibitor with varying exposure durations. The covariate method demonstrated increased BMC precision that enabled effective potency ranking of genotoxic agents based on BMC confidence intervals. Thus, this approach is more informative in hazard assessment than just concluding whether a given compound has a genotoxic potential or not. In the recent study by Fortin et al. (2023), an integrated in vitro-based approach was employed to evaluate the potency of 10 data-poor chemicals prioritized by Health Canada's New Substances Assessment and Control Bureau using structural alerts. The data-poor chemicals were evaluated using the TGx DDI, MicroFlow, and MultiFlow assays, and BMC modeling was applied to the concentration-response data to rank the chemicals by genotoxic potency. The BMC analysis identified clusters of chemicals with differing potencies. Overall, the EWG concluded that potency ranking using in vitro data is sufficiently developed; it is already routinely used in hazard evaluation as demonstrated by the highlighted examples.

Another risk assessment application where in vitro genotoxicity data could be used is the screening and prioritization of chemicals for further evaluation. Similar to potency ranking, BMC modeling is applied to concentration-response data, but the analyses are extended to factor in the toxicokinetics of the chemicals and the known exposure levels. There are a few recent examples where this approach has been applied. The study by Kuo et al. (2022) compared MN_{vit}-derived AEDs to exposure estimates to calculate bioactivity exposure ratios (BERs), an in vitro-based approach analogous to the margin of exposure approach. Chemicals with low BERs have a higher potential for concern in that the exposure level is approaching the estimated dose at which bioactivity (e.g., genotoxicity) induces a biologically relevant CES (i.e., AED derived from a BMC). In the study by Kuo et al., there were exposure estimates for 130 chemicals (122 classified as clastogenic, 8 classified as aneugenic). Using BER to screen chemicals, they identified several candidates as priorities for further scoping and risk evaluations. Similarly, the study by Beal et al. (2022) derived BERs for 19 chemicals that had available exposure estimates. Interestingly, the high BER group consisted mainly of chemicals used in therapeutics, and the BERs for these chemicals seem to reflect the chemicals' lower exposure levels and relatively low genotoxic potentials (i.e., relatively high AEDs). In contrast, the low BER chemicals included chemicals previously identified as probable carcinogens (etoposide, a chemotherapy drug) or chemicals that are ubiquitous in the environment and are also highly mutagenic (BaP, a polycyclic aromatic hydrocarbon). Lastly, a study by Schrenk et al. (2022) ranked the genotoxic potency of DNA reactive pyrrolizidine alkaloids (PAs) and this information was used to modulate regulatory action via the European Medicals Agency. Specifically, Schrenk et al. used a provisional PBTK model to derive PA concentration-time profiles and area under the curve (AUC) values in liver. The liver AUC values were multiplied by previously derived scaling factors for DNA adduct formation to estimate the in vivo levels of DNA adduct formation in human liver tissue. This approach provided an in vivo equivalent potency ranking to screen for the PAs possessing higher hazard potential. With these three supporting examples, the EWG concluded that screening and prioritization using in vitro data are sufficiently developed for more routine use in fit-for-purpose applications such as screening of in-commerce legacy chemicals or natural toxins. Approaches, however, may differ across chemical safety jurisdictions.

The Working Group agreed that the goal for in vitro genotoxicity data would involve its use in quantitative risk assessments used to set HBGVs. Progress has been made in the application of BMD modeling to derive HBGVs from in vivo animal data. For example, a study investigating *N*-nitrosamines (*N*-nitrosodiethylamine and *N*-nitrosodimethylamine) used in vivo TGR data and BMD modeling to derive permitted daily exposure (PDE) limits (Johnson et al., 2021). Each PDE was calculated by multiplying the $BMDL_{50}$ based on mutation data by 50 kg/person and dividing by a composite uncertainty factor of 5000 (interspecies (5), intraspecies (10), exposure duration (10), and severity (10) factors). The authors also derived a PDE based on carcinogenicity data using a $BMDL_{10}$ and a composite uncertainty factor of 500 (interspecies (5), intraspecies (10), exposure duration (1), and severity (10) factors). Interestingly, the PDEs based on mutation data were lower than the PDEs derived from cancer data. The higher sensitivity for mutation data is consistent with expectations because mutations are predicted to occur at lower doses and earlier time points than cancer as mutations are an early key event in the adverse outcome pathway for cancer initiation (Moore et al., 2008). Thus, there is utility in basing HBGV on

in vivo mutagenicity data. However, there is no evidence where HBGVs have been derived using in vitro genotoxicity data.

The studies that calculated AEDs using IVIVE indicate that there is some progress toward the development of in vitro approaches for HBGV derivation. But there is still a need for more work in this area. The Working Group agreed that HBGV derivation using in vitro data is an emerging opportunity in the early stages of development and indicated that there is a need for validated computation workflows, standard operating procedures, guidance documents, and prospective case studies to demonstrate proof-of-principle of using in vitro data for HBGV derivation.

9 | CONCLUDING REMARKS AND CONSENSUS STATEMENTS

The workshop discussions highlighted the modernization and advancements in in vitro genotoxicity assays and the recent progress made in the interpretation of concentration-response data for quantitative applications. Given the recent shift away from animal use in toxicity testing, there is an urgent need to continue building confidence in the quantitative use of in vitro genotoxicity concentration-response data. The in vitro testing strategies discussed in this report are the future of mainstream toxicity testing and will become the focus of risk assessments. Therefore, the EWG developed consensus statements as suggestions for building up the capabilities and usefulness of in vitro tests and the use of resultant data for regulatory decisions. The EWG achieved consensus, captured in five statements, about the current state of knowledge and future directions regarding the quantitative use of in vitro genotoxicity data for risk assessment and regulatory decision-making.

Consensus Statement 1: The commonly used mammalian in vitro genotoxicity endpoints (e.g., gene mutation and chromosomal damage assays) have the potential for use in quantitative applications such as the derivation of bioactivity exposure ratios (BERs).

Consensus Statement 2: Initial analysis indicates that CES may need to be endpoint (assay) specific. Additional data should be analyzed to give a more robust recommendation on appropriate CES for various assays and endpoints.

Consensus Statement 3: Because it is anticipated that in vivo genotoxicity data will become less available in the near future, it is critical to address the uncertainties associated with in vitro data and IVIVE to support the interpretation of concentration-response data in a human health context.

Consensus Statement 4: As researchers build confidence in the application of more human-relevant models reflecting higher complexity, there will be opportunities to apply the data quantitatively in chemical safety evaluations. It is anticipated that higher complexity models will improve the ability to interpret in vitro genotoxicity data in the human health context.

Consensus Statement 5: Quantitative interpretation of in vitro genotoxicity data has already demonstrated utility in potency ranking, and this strategy is sufficiently developed to support screening and prioritization of data-poor substances for further evaluation. Further case studies and standardization efforts are needed before PoDs derived from in vitro data can be employed for quantitative risk assessments.

AUTHOR CONTRIBUTIONS

All authors participated in the discussions that resulted in consensus statements described in the report. All authors contributed to the drafting of the report. Marc A. Beal served as Chair of the in vitro EWG and played a major role in leading group discussions and drafting of the report. The roles of Co-Chairs and Rapporteur of the broader Quantitative Analysis working group were held by Paul A. White, George Johnson, and Andreas Zeller. Table 1 was created by Alexandra S. Long and this information formed the basis for the section on in vitro methodologies. The CES calculations were done by Guangchao Chen with support from John Wills.

ACKNOWLEDGMENTS

The authors would like to acknowledge the contributions of other members of the Genotoxicity Dose–Response Analyses for Potency Comparisons and Risk Assessment working group as well as the other IWGT attendees. The authors are also grateful for helpful comments provided by internal reviewers Bas Bokkers, Xiaoqing Guo, Nan Mei, Nikolai Chepelev, and Deon Williams as well as anonymous peer-reviewers.

ORCID

Marc A. Beal https://orcid.org/0000-0003-2760-4368
Barbara L. Parsons https://orcid.org/0000-0002-3005-2552
Stefan Pfuhler https://orcid.org/0000-0001-8869-5975
Andreas Zeller https://orcid.org/0000-0002-3565-6347
George Johnson https://orcid.org/0000-0001-5643-9942

ENDNOTES

[1] https://www.oecd.org/env/ehs/testing/work-plan-test-guidelines-programme-july-2021.pdf.

[2] https://fda.force.com/ddt/s/ddt-project?ddtprojectid=77 (DDT Project Number DDT-BMQ-000008).

REFERENCES

Adler, S., Basketter, D., Creton, S., Pelkonen, O., van Benthem, J., Zuang, V. et al. (2011) Alternative (non-animal) methods for cosmetics testing: current status and future prospects—2010. *Archives of Toxicology*, 85, 367–485.

Ames, B.N., Gurney, E.G., Miller, J.A. & Bartsch, H. (1972a) Carcinogens as frameshift mutagens: metabolites and derivatives of 2-acetylaminofluorene and other aromatic amine carcinogens. *Proceedings of the National Academy of Sciences of the United States of America*, 69, 3128–3132.

Ames, B.N., Lee, F.D. & Durston, W.E. (1973a) An improved bacterial test system for the detection and classification of mutagens and carcinogens. *Proceedings of the National Academy of Sciences of the United States of America*, 70, 782–786.

Ames, B.N., Sims, P. & Grover, P.L. (1972b) Epoxides of carcinogenic polycyclic hydrocarbons are frameshift mutagens. *Science*, 176, 47–49.

Ames, B.N., Durston, W.E., Yamasaki, E. & Lee, F.D. (1973b) Carcinogens are mutagens: a simple test system combining liver homogenates for activation and bacteria for detection. *Proceedings of the National Academy of Sciences of the United States of America*, 70, 2281–2285.

Armitage, J.M., Sangion, A., Parmar, R., Looky, A.B. & Arnot, J.A. (2021) Update and evaluation of a high-throughput in vitro mass balance distribution model: IV-MBM EQP v2.0. *Toxics*, 9, 315.

Armitage, J.M., Wania, F. & Arnot, J.A. (2014) Application of mass balance models and the chemical activity concept to facilitate the use of in vitro toxicity data for risk assessment. *Environmental Science & Technology*, 48, 9770–9779.

Avlasevich, S., Pellegrin, T., Godse, M., Bryce, S., Bemis, J., Bajorski, P. et al. (2021) Biomarkers of DNA damage response improve in vitro micronucleus assays by revealing genotoxic mode of action and reducing the occurrence of irrelevant positive results. *Mutagenesis*, 36, 407–418.

Baran, S.W., Brown, P.C., Baudy, A.R., Fitzpatrick, S.C., Frantz, C., Fullerton, A. et al. (2022) Perspectives on the evaluation and adoption of complex in vitro models in drug development: workshop with the FDA and the pharmaceutical industry (IQ MPS Affiliate). *ALTEX-Alternatives to Animal Experimentation*, 39, 297–314.

Barlow, S., Chesson, A., Collins, J.D., Flynn, A., Hardy, A., Jany, K.-D. et al. (2009) Use of the benchmark dose approach in risk assessment Guidance of the Scientific Committee. *EFSA Journal*, 1150, 1–72.

Barranger, A. & Le Hégarat, L. (2022) Towards better prediction of xenobiotic genotoxicity: CometChip technology coupled with a 3D model of HepaRG human liver cells. *Archives of Toxicology*, 96, 2087–2095.

Beal, M.A., Audebert, M., Barton-Maclaren, T., Battaion, H., Bemis, J.C., Cao, X. et al. (2022) Quantitative in vitro to in vivo extrapolation of genotoxicity data provides protective estimates of in vivo dose. *Environmental and Molecular Mutagenesis*, 64(2), 105–122.

Beal, M.A., Gagne, M., Kulkarni, S.A., Patlewicz, G., Thomas, R.S. & Barton-Maclaren, T.S. (2021) Implementing in vitro bioactivity data to modernize priority setting of chemical inventories. *ALTEX—Alternatives to Animal Experimentation*, 39, 123–139.

Beal, M.A., Gagné, R., Williams, A., Marchetti, F. & Yauk, C.L. (2015) Characterizing benzo[a]pyrene-induced lacZ mutation spectrum in transgenic mice using next-generation sequencing. *BMC Genomics*, 16, 812.

Beal, M.A., Meier, M.J., LeBlanc, D.P., Maurice, C., O'Brien, J.M., Yauk, C.L. et al. (2020) Chemically induced mutations in a MutaMouse reporter gene inform mechanisms underlying human cancer mutational signatures. *Communications Biology*, 3, 1–11.

Bemis, J.C. & Heflich, R.H. (2019) In vitro mammalian cell mutation assays based on the pig-a gene: A report of the 7th International Workshop on Genotoxicity Testing (IWGT) Workgroup. *Mutation Research, Genetic Toxicology and Environmental Mutagenesis*, 847, 403028.

Besaratinia, A., Li, H., Yoon, J.I., Zheng, A., Gao, H. & Tommasi, S. (2012) A high-throughput next-generation sequencing-based method for detecting the mutational fingerprint of carcinogens. *Nucleic Acids Research*, 40, e116.

Boisvert, L., Derr, R., Osterlund, T., Hendriks, G. & Brandsma, I. (2023) Quantitative interpretation of ToxTracker dose–response data for potency comparisons and mode-of-action determination. *Environmental and Molecular Mutagenesis*, 64, 132–143.

Boverhof, D.R., Chamberlain, M.P., Elcombe, C.R., Gonzalez, F.J., Heflich, R.H., Hernández, L.G. et al. (2011) Transgenic animal models in toxicology: historical perspectives and future outlook. *Toxicological Sciences*, 121, 207–233.

Boyadzhiev, A., Solorio-Rodriguez, S.A., Wu, D., Avramescu, M.L., Rasmussen, P. & Halappanavar, S. (2022) The high-throughput in vitro CometChip assay for the analysis of metal oxide nanomaterial induced DNA damage. *Nanomaterials*, 12, 1844.

Bryce, S.M., Avlasevich, S.L., Bemis, J.C., Phonethepswath, S. & Dertinger, S.D. (2010) Miniaturized flow cytometric in vitro micronucleus assay represents an efficient tool for comprehensively characterizing genotoxicity dose-response relationships. *Mutation Research*, 703, 191–199.

Bryce, S.M., Bernacki, D.T., Bemis, J.C. & Dertinger, S.D. (2016) Genotoxic mode of action predictions from a multiplexed flow cytometric assay and a machine learning approach. *Environmental and Molecular Mutagenesis*, 57, 171–189.

Bryce, S.M., Bernacki, D.T., Bemis, J.C., Spellman, R.A., Engel, M.E., Schuler, M. et al. (2017) Interlaboratory evaluation of a multiplexed high information content in vitro genotoxicity assay. *Environmental and Molecular Mutagenesis*, 58, 146–161.

Buick, J.K., Williams, A., Gagné, R., Swartz, C.D., Recio, L., Ferguson, S.S. et al. (2020) Flow cytometric micronucleus assay and TGx-DDI transcriptomic biomarker analysis of ten genotoxic and non-genotoxic chemicals in human HepaRG™ cells. *Genes and Environment*, 42, 5.

Carstens, K.E., Carpenter, A.F., Martin, M.M., Harrill, J.A., Shafer, T.J. & Paul Friedman, K. (2022) Integrating data from in vitro new approach methodologies for developmental neurotoxicity. *Toxicological Sciences*, 187, 62–79.

Casey, W.M., Chang, X., Allen, D.G., Ceger, P.C., Choksi, N.Y., Hsieh, J.H. et al. (2018) Evaluation and optimization of pharmacokinetic models for in vitro to in vivo extrapolation of estrogenic activity for environmental chemicals. *Environmental Health Perspectives*, 126, 97001.

Cho, E., Buick, J.K., Williams, A., Chen, R., Li, H.H., Corton, J.C. et al. (2019) Assessment of the performance of the TGx-DDI biomarker to detect DNA damage-inducing agents using quantitative RT-PCR in TK6 cells. *Environmental and Molecular Mutagenesis*, 60, 122–133.

Cho, E., Swartz, C.D., Williams, A., V. Rivas, M., Recio, L., Witt, K.L. et al. (2023) Error-corrected duplex sequencing enables direct detection and quantification of mutations in human TK6 cells with strong interlaboratory consistency. *Mutation Research/Genetic Toxicology and Environmental Mutagenesis*, 889, 503649.

ECHA. (2016) *New approach methodologies in regulatory science. Proceedings of a scientific workshop*. Helsinki: European Chemicals Agency.

EFSA Scientific Committee, Hardy, A., Benford, D., Halldorsson, T., Jeger, M.J., Knutsen, K.H. et al. (2017) Update: use of the benchmark dose approach in risk assessment. *EFSA Journal*, 15, e04658.

EFSA Scientific Committee, More, S.J., Bampidis, V., Benford, D., Bragard, C., Halldorsson, T.I. et al. (2022) Guidance on the use of the benchmark dose approach in risk assessment. *EFSA Journal*, 20, e07584.

EMGS. (2021) Abstracts from the environmental mutagenesis and genomics society 52nd Annual Meeting, September 22-25, 2021, virtual meeting—real science program chair: Jeff Bemis, PhD new investigator Co-chair: Isabelle R. Miousse, PhD. *Environmental and Molecular Mutagenesis*, 62(1), 1–95.

Fortin, A.-M.V., Long, A.S., Williams, A., Meier, M.J., Cox, J., Pinsonnault, C. et al. (2023) Application of a new approach methodology (NAM)-based strategy for genotoxicity assessment of data-poor compounds. *Frontiers in Toxicology*, 5, 1098432.

Ge, J., Prasongtanakij, S., Wood, D.K., Weingeist, D.M., Fessler, J., Navasummrit, P. et al. (2014) CometChip: a high-throughput 96-well platform for measuring DNA damage in microarrayed human cells. *Journal of Visualized Experiments*, 92, 50607.

Government of Canada. (2023) Bill S-5, Strengthening Environmental Protection for a Healthier Canada Act—Summary of Amendments.

Grimm, D. (2019) US EPA to eliminate all mammal testing by 2035. *Science*, 365, 1231.

Hasselgren, C., Ahlberg, E., Akahori, Y., Amberg, A., Anger, L.T., Atienzar, F. et al. (2019) Genetic toxicology in silico protocol. *Regulatory Toxicology and Pharmacology*, 107, 104403.

Health Canada. (2021) Science approach document—Bioactivity exposure ratio: Application in priority setting and risk assessment.

Heflich, R.H., Johnson, G.E., Zeller, A., Marchetti, F., Douglas, G.R., Witt, K.L. et al. (2020) Mutation as a toxicological endpoint for regulatory decision-making. *Environmental and Molecular Mutagenesis*, 61, 34–41.

Hendriks, G., Atallah, M., Morolli, B., Calléja, F., Ras-Verloop, N., Huijskens, I. et al. (2012) The ToxTracker assay: novel GFP reporter

systems that provide mechanistic insight into the genotoxic properties of chemicals. *Toxicological Sciences*, 125, 285–298.

Hendriks, G., Derr, R.S., Misovic, B., Morolli, B., Calléja, F.M.G.R. & Vrieling, H. (2016) The extended ToxTracker assay discriminates between induction of DNA damage, oxidative stress, and protein misfolding. *Toxicological Sciences*, 150, 190–203.

Hsieh, J.-H., Smith-Roe, S.L., Huang, R., Sedykh, A., Shockley, K.R., Auerbach, S.S. et al. (2019) Identifying compounds with genotoxicity potential using Tox21 high-throughput screening assays. *Chemical Research in Toxicology*, 32, 1384–1401.

Ingber, D.E. (2022) Human organs-on-chips for disease modelling, drug development and personalized medicine. *Nature Reviews Genetics*, 23, 467–491.

Johnson, G.E., Dobo, K., Gollapudi, B., Harvey, J., Kenny, J., Kenyon, M. et al. (2021) Permitted daily exposure limits for noteworthy N-nitrosamines. *Environmental and Molecular Mutagenesis*, 62, 293–305.

Khoury, L., Zalko, D. & Audebert, M. (2013) Validation of high-throughput genotoxicity assay screening using γH2AX in-cell western assay on HepG2 cells. *Environmental and Molecular Mutagenesis*, 54, 737–746.

Khoury, L., Zalko, D. & Audebert, M. (2016a) Complementarity of phosphorylated histones H2AX and H3 quantification in different cell lines for genotoxicity screening. *Archives of Toxicology*, 90, 1983–1995.

Khoury, L., Zalko, D. & Audebert, M. (2016b) Evaluation of four human cell lines with distinct biotransformation properties for genotoxic screening. *Mutagenesis*, 31, 83–96.

Kuo, B., Beal, M.A., Wills, J.W., White, P.A., Marchetti, F., Nong, A. et al. (2022) Comprehensive interpretation of in vitro micronucleus test results for 292 chemicals: from hazard identification to risk assessment application. *Archives of Toxicology*, 96, 2067–2085.

Lambert, I.B., Singer, T.M., Boucher, S.E. & Douglas, G.R. (2005) Detailed review of transgenic rodent mutation assays. *Mutation Research*, 590, 1–280.

LeBlanc, D.P., Meier, M., Lo, F.Y., Schmidt, E., Valentine, C., 3rd, Williams, A. et al. (2022) Duplex sequencing identifies genomic features that determine susceptibility to benzo (a) pyrene-induced in vivo mutations. *BMC Genomics*, 23, 1–15.

Li, H.-H., Chen, R., Hyduke, D.R., Williams, A., Frötschl, R., Ellinger-Ziegelbauer, H. et al. (2017) Development and validation of a high-throughput transcriptomic biomarker to address 21st century genetic toxicology needs. *Proceedings of the National Academy of Sciences of the United States of America*, 114, E10881–E10889.

Li, H.-H., Hyduke, D.R., Chen, R., Heard, P., Yauk, C.L., Aubrecht, J. et al. (2015) Development of a toxicogenomics signature for genotoxicity using a dose-optimization and informatics strategy in human cells. *Environmental and Molecular Mutagenesis*, 56, 505–519.

MacGregor, J.T., Frötschl, R., White, P.A., Crump, K.S., Eastmond, D.A., Fukushima, S. et al. (2015) IWGT report on quantitative approaches to genotoxicity risk assessment I. Methods and metrics for defining exposure–response relationships and points of departure (PoDs). *Mutation Research/Genetic Toxicology and Environmental Mutagenesis*, 783, 55–65.

Marchetti, F., Cardoso, R., Chen, C.L., Douglas, G.R., Elloway, J., Escobar, P.A. et al. (2023) Error-corrected next-generation sequencing to advance nonclinical genotoxicity and carcinogenicity testing. *Nature Reviews Drug Discovery*, 22, 165–166.

Moore, M.M., Heflich, R.H., Haber, L.T., Allen, B.C., Shipp, A.M. & Kodell, R.L. (2008) Analysis of in vivo mutation data can inform cancer risk assessment. *Regulatory Toxicology and Pharmacology*, 51, 151–161.

Moreau, M., Mallick, P., Smeltz, M., Haider, S., Nicolas, C.I., Pendse, S.N. et al. (2022) Considerations for improving metabolism predictions for in vitro to in vivo extrapolation. *Frontiers in Toxicology*, 4, 894569.

OECD. (2014) Test No. 487: In vitro mammalian cell micronucleus test OECD publishing.

OECD. (2016a) Test No. 473: In Vitro Mammalian Chromosomal Aberration Test OECD Publishing.

OECD. (2016b) Test No. 476: In Vitro Mammalian Cell Gene Mutation Test OECD Publishing.

OECD. (2016c) Test No. 489: In vivo mammalian alkaline comet assay. OECD Guidelines for the Testing of Chemicals, Section, 4.

OECD. (2016d) Test No. 490: in vitro mammalian cell gene mutation tests using the thymidine kinase gene. OECD Guidelines for the Testing of Chemicals.

OECD. (2022a) Test No. 470: mammalian erythrocyte pig-a gene mutation assay OECD Publishing.

OECD. (2022b) Test No. 488: Transgenic Rodent Somatic and Germ Cell Gene Mutation Assays OECD Publishing.

Ooka, M., Zhao, J., Shah, P., Travers, J., Klumpp-Thomas, C., Xu, X. et al. (2022) Identification of environmental chemicals that activate p53 signaling after in vitro metabolic activation. *Archives of Toxicology*, 96, 1975–1987.

Paul Friedman, K., Gagne, M., Loo, L.H., Karamertzanis, P., Netzeva, T., Sobanski, T. et al. (2020) Utility of in vitro bioactivity as a lower bound estimate of in vivo adverse effect levels and in risk-based prioritization. *Toxicological Sciences*, 173, 202–225.

Pearce, R.G., Setzer, R.W., Strope, C.L., Wambaugh, J.F. & Sipes, N.S. (2017) Httk: R package for high-throughput toxicokinetics. *Journal of Statistical Software*, 79, 1–26.

Pfuhler, S., van Benthem, J., Curren, R., Doak, S.H., Dusinska, M., Hayashi, M. et al. (2020) Use of in vitro 3D tissue models in genotoxicity testing: strategic fit, validation status and way forward. Report of the working group from the 7th International Workshop on Genotoxicity Testing (IWGT). *Mutation Research: Genetic Toxicology and Environmental Mutagenesis*, 850–851, 503135.

Pham, L.L., Watford, S.M., Pradeep, P., Martin, M.T., Thomas, R.S., Judson, R.S. et al. (2020) Variability in in vivo studies: defining the upper limit of performance for predictions of systemic effect levels. *Computational Toxicology*, 15, 100126.

Pradeep, P., Patlewicz, G., Pearce, R., Wambaugh, J., Wetmore, B. & Judson, R. (2020) Using chemical structure information to develop predictive models for in vitro toxicokinetic parameters to inform high-throughput risk-assessment. *Computational Toxicology*, 16, 100136.

Qin, Q., Wu, Q., Wang, Y., Xiong, R., Guo, L., Fu, X. et al. (2019) Effects of cellular differentiation in human primary bronchial epithelial cells: metabolism of 4-(methylnitrosamine)-1-(3-pyridyl)-1-butanone. *Toxicology In Vitro*, 55, 185–194.

Rajkumar, A., Luu, T., Beal, M.A., Barton-Maclaren, T.S., Hales, B.F. & Robaire, B. (2021b) Phthalates and alternative plasticizers differentially affect phenotypic parameters in gonadal somatic and germ cell lines. *Biology Reproduction*, 106, 613–627.

Rajkumar, A., Luu, T., Beal, M.A., Barton-Maclaren, T.S., Robaire, B. & Hales, B.F. (2021a) Elucidation of the effects of bisphenol a and structural analogs on germ and steroidogenic cells using single cell high-content imaging. *Toxicological Sciences*, 180, 224–238.

Ring, C.L., Pearce, R.G., Setzer, R.W., Wetmore, B.A. & Wambaugh, J.F. (2017) Identifying populations sensitive to environmental chemicals by simulating toxicokinetic variability. *Environment International*, 106, 105–118.

Ronaldson-Bouchard, K. & Vunjak-Novakovic, G. (2018) Organs-on-a-chip: a fast track for engineered human tissues in drug development. *Cell Stem Cell*, 22, 310–324.

Rotroff, D.M., Wetmore, B.A., Dix, D.J., Ferguson, S.S., Clewell, H.J., Houck, K.A. et al. (2010) Incorporating human dosimetry and exposure into high-throughput in vitro toxicity screening. *Toxicological Sciences*, 117, 348–358.

Salk, J.J. & Kennedy, S.R. (2020) Next-generation genotoxicology: using modern sequencing technologies to assess somatic mutagenesis and cancer risk. *Environmental and Molecular Mutagenesis*, 61, 135–151.

Salk, J.J., Schmitt, M.W. & Loeb, L.A. (2018) Enhancing the accuracy of next-generation sequencing for detecting rare and subclonal mutations. *Nature Reviews Genetics*, 19, 269–285.

Schirmer, K. (2006) Proposal to improve vertebrate cell cultures to establish them as substitutes for the regulatory testing of chemicals and effluents using fish. *Toxicology*, 224, 163–183.

Schmitt, M.W., Kennedy, S.R., Salk, J.J., Fox, E.J., Hiatt, J.B. & Loeb, L.A. (2012) Detection of ultra-rare mutations by next-generation sequencing. *Proceedings of the National Academy of Sciences*, 109, 14508–14513.

Schrenk, D., Fahrer, J., Allemang, A., Fu, P., Lin, G., Mahony, C. et al. (2022) Novel insights into pyrrolizidine alkaloid toxicity and implications for risk assessment: occurrence, genotoxicity, toxicokinetics, risk assessment-a workshop report. *Planta Medica*, 88, 98–117.

Seo, J.-E., He, X., Muskhelishvili, L., Malhi, P., Mei, N., Manjanatha, M. et al. (2022) Evaluation of an in vitro three-dimensional HepaRG spheroid model for genotoxicity testing using the high-throughput CometChip platform. *ALTEX-Alternatives to Animal Experimentation*, 39(4), 583–604.

Seo, J.-E., Li, X., le, Y., Mei, N., Zhou, T. & Guo, X. (2023) High-throughput micronucleus assay using three-dimensional HepaRG spheroids for in vitro genotoxicity testing. *Archives of Toxicology*, 97, 1–13.

Shah, U.-K., Mallia, J.O., Singh, N., Chapman, K.E., Doak, S.H. & Jenkins, G.J.S. (2018) A three-dimensional in vitro HepG2 cells liver spheroid model for genotoxicity studies. *Mutation Research, Genetic Toxicology and Environmental Mutagenesis*, 825, 51–58.

Slob, W. (2017) A general theory of effect size, and its consequences for defining the benchmark response (BMR) for continuous endpoints. *Critical Reviews in Toxicology*, 47, 342–351.

Stadnicka-Michalak, J., Bramaz, N., Schönenberger, R. & Schirmer, K. (2021) Predicting exposure concentrations of chemicals with a wide range of volatility and hydrophobicity in different multi-well plate setups. *Scientific Reports*, 11, 4680.

Stucki, A.O., Barton-Maclaren, T.S., Bhuller, Y., Henriquez, J.E., Henry, T.R., Hirn, C. et al. (2022) Use of new approach methodologies (NAMs) to meet regulatory requirements for the assessment of industrial chemicals and pesticides for effects on human health. *Frontiers in Toxicology*, 4, 964553.

Thybaud, V., Dean, S., Nohmi, T., de Boer, J., Douglas, G.R., Glickman, B.W. et al. (2003) In vivo transgenic mutation assays. *Mutation Research*, 540, 141–151.

US EPA. (2018) Strategic plan to promote the development and implementation of alternative test methods within the TSCA program.

US FDA. (2017) *FDA's predictive toxicology roadmap*. Silver Spring, MD: US Food and Drug Administration.

Valentine, C.C., III, Young, R.R., Fielden, M.R., Kulkarni, R., Williams, L.N., Li, T. et al. (2020) Direct quantification of in vivo mutagenesis and carcinogenesis using duplex sequencing. *Proceedings of the National Academy of Sciences*, 117, 33414–33425.

Wang, X., Luu, T., Beal, M.A., Barton-Maclaren, T.S., Robaire, B. & Hales, B.F. (2022) The effects of organophosphate esters used as flame retardants and plasticizers on granulosa, Leydig, and spermatogonial cells analyzed using high-content imaging. *Toxicological Sciences*, 186, 269–287.

Wang, Y., Mittelstaedt, R.A., Wynne, R., Chen, Y., Cao, X., Muskhelishvili, L. et al. (2021) Genetic toxicity testing using human in vitro organotypic airway cultures: assessing DNA damage with the CometChip and mutagenesis by duplex sequencing. *Environmental and Molecular Mutagenesis*, 62, 306–318.

Weingeist, D.M., Ge, J., Wood, D.K., Mutamba, J.T., Huang, Q., Rowland, E.A. et al. (2013) Single-cell microarray enables high-throughput evaluation of DNA double-strand breaks and DNA repair inhibitors. *Cell Cycle*, 12, 907–915.

Wetmore, B.A. (2015) Quantitative in vitro-to-in vivo extrapolation in a high-throughput environment. *Toxicology*, 332, 94–101.

Wetmore, B.A., Wambaugh, J.F., Ferguson, S.S., Sochaski, M.A., Rotroff, D.M., Freeman, K. et al. (2012) Integration of dosimetry, exposure, and high-throughput screening data in chemical toxicity assessment. *Toxicological Sciences*, 125, 157–174.

White, P.A., Long, A.S. & Johnson, G.E. (2020) Quantitative interpretation of genetic toxicity dose-response data for risk assessment and regulatory decision-making: current status and emerging priorities. *Environmental and Molecular Mutagenesis*, 61, 66–83.

White, P.A., Luijten, M., Mishima, M., Cox, J.A., Hanna, J.N., Maertens, R.M. et al. (2019) In vitro mammalian cell mutation assays based on transgenic reporters: a report of the International Workshop on Genotoxicity Testing (IWGT). *Mutation Research/Genetic Toxicology and Environmental Mutagenesis*, 847, 403039.

Wikswo, J.P. (2014) The relevance and potential roles of microphysiological systems in biology and medicine. *Experimental Biology and Medicine (Maywood, N.J.)*, 239, 1061–1072.

Wills, J.W., Halkes-Wellstead, E., Summers, H.D., Rees, P. & Johnson, G.E. (2021a) Empirical comparison of genotoxic potency estimations: the in vitro DNA-damage ToxTracker endpoints versus the in vivo micronucleus assay. *Mutagenesis*, 36, 311–320.

Wills, J.W., Hondow, N., Thomas, A.D., Chapman, K.E., Fish, D., Maffeis, T.G. et al. (2015) Genetic toxicity assessment of engineered nanoparticles using a 3D in vitro skin model (EpiDerm™). *Particle and Fibre Toxicology*, 13, 1–21.

Wills, J.W., Johnson, G.E., Doak, S.H., Soeteman-Hernández, L.G., Slob, W. & White, P.A. (2016) Empirical analysis of BMD metrics in genetic toxicology part I: in vitro analyses to provide robust potency rankings and support MOA determinations. *Mutagenesis*, 31, 255–263.

Wills, J.W., Verma, J.R., Rees, B.J., Harte, D.S.G., Haxhiraj, Q., Barnes, C.M. et al. (2021b) Inter-laboratory automation of the in vitro micronucleus assay using imaging flow cytometry and deep learning. *Archives of Toxicology*, 95, 3101–3115.

World Health Organization. (2020) EHC240: principles and methods for the risk assessment of chemicals in food. In: *Hazard identification and characterization: toxicological and human studies*. Stuttgart, Germany: Wissenchaftliche Verlagsgesellschaft mbH.

Zeller, A., Duran-Pacheco, G. & Guérard, M. (2017) An appraisal of critical effect sizes for the benchmark dose approach to assess dose-response relationships in genetic toxicology. *Archives of Toxicology*, 91, 3799–3807.

How to cite this article: Beal, M.A., Chen, G., Dearfield, K.L., Gi, M., Gollapudi, B., Heflich, R.H. et al. (2024) Interpretation of in vitro concentration-response data for risk assessment and regulatory decision-making: Report from the 2022 IWGT quantitative analysis expert working group meeting. *Environmental and Molecular Mutagenesis*, 1–20. Available from: https://doi.org/10.1002/em.22582

APPENDIX A

The Slob (2017) Effect Size Theory approach was applied to compiled concentration-response datasets for (i) the in vitro TGR assay (64 datasets; Beal et al., 2022), (ii) the in vitro MN_{vit} assay (135 datasets; Kuo et al., 2022), and the in vitro ToxTracker assay (189 datasets for *Bscl2* and 269 datasets for *Rtkn*; Boisvert et al., 2023). These endpoints effectively represent the aforementioned categories of in vitro genotoxicity endpoints. The analyses provide estimates of CES values that are appropriately scaled to maximum response and/or within-group variation averaged across the available studies. Although the analyses did not evaluate all the possible types of in vitro genotoxicity assays, the results show that CES values based on maximum response would be 60.7%, 34.0%, 19.8%, and 37.7% for TGR, MN_{vit}, ToxTracker *Bscl2*, and ToxTracker *Rtkn*, respectively (Table 2). The CES values based on within-group variation would be 28.2%, 21.8%, 21.0%, and 32.4%, respectively (Table 2). The range of CES values derived from this exercise suggests that no single CES value will be appropriate for all in vitro genotoxicity endpoints; moreover, the CES values employed for analysis of in vitro genotoxicity concentration data may have to be assay-specific. A comprehensive analysis of available concentration-response data for all aforementioned assays, including all possible versions of each assay in terms of the cell types used, assay methods employed, etc., is beyond the scope of this report. Further work is required to employ the Effect Size Theory approach to define CES values for other assays.