Quantitative in vitro to in vivo Extrapolation (IVIVE) of Genotoxicity Data Provides Protective Estimates of in vivo Dose

Marc Beal, Marc Audebert, Tara Barton-Maclaren, Hannah Battaion, Jeffrey Bemis, Xuefei Cao, Connie Chen, Stephen Dertinger, Roland Froetschl, Xiaoqing Guo, et al.

Received: 2 October 2022 | Revised: 29 November 2022 | Accepted: 30 November 2022

DOI: 10.1002/em.22521

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

RESEARCH ARTICLE

Quantitative in vitro to in vivo extrapolation of genotoxicity data provides protective estimates of in vivo dose

Marc A. Beal[1] | Marc Audebert[2] | Tara Barton-Maclaren[3] | Hannah Battaion[4] | Jeffrey C. Bemis[5] | Xuefei Cao[6] | Connie Chen[7] | Stephen D. Dertinger[5] | Roland Froetschl[8] | Xiaoqing Guo[6] | George Johnson[9] | Giel Hendriks[10] | Laure Khoury[11] | Alexandra S. Long[3] | Stefan Pfuhler[12] | Raja S. Settivari[13] | Shamika Wickramasuriya[3] | Paul White[4]

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

[2]Toxalim UMR1331, Toulouse University, INRAE, Toulouse, France

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

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

[5]Litron Laboratories, Rochester, New York, USA

[6]Division of Genetic and Molecular Toxicology, National Center for Toxicological Research, U.S. Food and Drug Administration, Jefferson, Arkansas, USA

[7]Health and Environmental Sciences Institute, Washington, District of Columbia, USA

[8]Genetic and Reproductive Toxicology, BfArM, Bonn, Germany

[9]Swansea University, Swansea, UK

[10]Toxys, Oegstgeest, Netherlands

[11]PrediTox, Toulouse, France

[12]Global Product Stewardship, Procter & Gamble, Cincinnati, Ohio, USA

[13]Mammalian Toxicology Center, Corteva Agriscience, Newark, Delaware, USA

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: M. J. LeBaron

Abstract

Genotoxicity assessment is a critical component in the development and evaluation of chemicals. Traditional genotoxicity assays (i.e., mutagenicity, clastogenicity, and aneugenicity) have been limited to dichotomous hazard classification, while other toxicity endpoints are assessed through quantitative determination of points-of-departures (PODs) for setting exposure limits. The more recent higher-throughput in vitro genotoxicity assays, many of which also provide mechanistic information, offer a powerful approach for determining defined PODs for potency ranking and risk assessment. In order to obtain relevant human dose context from the in vitro assays, in vitro to in vivo

Abbreviations: AED, administered equivalent dose; BER, bioactivity exposure ratio; BMC, benchmark concentration; BMD, benchmark dose; CA, chromosomal aberration; CEPA, Canadian Environmental Protection Act; Cl_{int}, hepatic clearance; C_{ss}, steady-state concentration in plasma; DBCP, 1,2-dibromo-3-chloropropane; DMNA, dimethylnitrosamine; EMS, ethyl methanesulfonate; ENU, *N*-ethyl-*N*-nitrosourea; ExpoCast, exposure forecasting; FISH, fluorescence in situ hybridization; GTTC, Genetic Toxicology Technical Committee; HC, Health Canada; HESI, Health and Environmental Sciences Institute; HPRT, hypoxanthine phosphoribosyl transferase; HTTK, high-throughput toxicokinetics; IARC, International Agency for Research on Cancer; IVIVE, in vitro to in vivo extrapolation; LOGEL, lowest-observed-genotoxic-effect-level; MMS, methyl methanesulfonate; MOA, mode of action; MOE, margin of exposure; MW, molecular weight; NAM, new approach methodology; NOGEL, no-observed-genotoxic-effect-level; OECD, Organization for Economic Co-operation and Development; PBTK, physiologically-based toxicokinetic; p-H3, phospho-histone H3; PhIP.HCl, 2-amino-1-methyl-6-phenylimidazo[4,5-*b*]pyridine hydrochloride; POD, point-of-departure; SEEM3, systematic empirical evaluation of models 3; SWOT, strengths, weaknesses, opportunities, threats; TGR, transgenic rodent gene mutation assay; TK, thymidine kinase; TRAID, Transgenic Rodent Assay Information Database.

extrapolation (IVIVE) models are required to determine what dose would elicit a concentration in the body demonstrated to be genotoxic using in vitro assays. Previous work has demonstrated that application of IVIVE models to in vitro bioactivity data can provide PODs that are protective of human health, but there has been no evaluation of how these models perform with in vitro genotoxicity data. Thus, the Genetic Toxicology Technical Committee, under the Health and Environmental Sciences Institute, conducted a case study on 31 reference chemicals to evaluate the performance of IVIVE application to genotoxicity data. The results demonstrate that for most chemicals considered here (20/31), the PODs derived from in vitro data and IVIVE are health protective relative to in vivo PODs from animal studies. PODs were also protective by assay target: mutations (8/13 chemicals), micronuclei (9/12), and aneugenicity markers (4/4). It is envisioned that this novel testing strategy could enhance prioritization, rapid screening, and risk assessment of genotoxic chemicals.

KEYWORDS
clastogen, genetic toxicology, in vitro to in vivo extrapolation, mutation, new approach methodologies

1 | INTRODUCTION

A critical and early step in the safety evaluation of a chemical, regardless of application (drugs/therapeutics, agrochemicals, pesticides, food additives, etc.), is the assessment of its genotoxic potential. Consequently, international regulatory agencies are in need of genotoxicity data to support hazard and/or risk assessment of data-poor legacy chemicals that already exist in commerce. Genotoxicity assessment focuses on whether a chemical damages DNA, that is, alters DNA sequence (mutagen), damages chromosomes (clastogen), or alters chromosome number (aneugen). The traditional genotoxicity testing strategies rely on assays designed to detect gene mutations, structural chromosomal aberrations (clastogenicity), and numerical chromosomal aberrations (aneuploidy), and have been in place for decades with only marginal changes in the data interpretation and test systems used (Cimino, 2006). These standard tests have played a significant role in hazard identification, but they are lower throughput and provide limited mechanistic information; thus, there exists an opportunity for innovations and alternative testing strategies to modernize genotoxicity evaluation.

Recently, there has been increased international pressure to develop robust alternatives to animal testing (Kavlock et al., 2018). For example, the U.S. Environmental Protection Agency recently committed to the reduction and eventual elimination of animal toxicity testing (Grimm, 2019), and the Canadian federal government has followed suit with similar targets (Trudeau, 2021; Government of Canada, 2021). These ambitious efforts present an opportunity to modernize risk assessment activities, including priority setting and rapid screening assessment, through the exploration and implementation of new approach methodologies (NAMs). NAMs refer broadly to any technology, method, and/or approach supporting risk assessment without the use of animals. NAMs also pertain to modernized methods for data analysis and interpretation. Several in vitro NAMs that provide some measure of genotoxicity already exist. Some examples of NAMs include indicator assays that measure DNA damage reporter signals, such as ToxTracker® (Hendriks et al., 2012), Prediscreen (Khoury et al., 2013, 2016a), and MultiFlow® (Bryce et al., 2018). There are also NAMs that directly measure mutations, such as the FE1 cell-based in vitro version of the MutaMouse transgenic rodent (TGR) gene mutation assay (Maertens et al., 2017; White et al., 2003). These NAMs and other more traditional in vitro approaches for genotoxicity assessment serve as robust tools for hazard identification and can be used to support a weight of evidence assessment. However, the lack of human dose context from in vitro NAM results makes it difficult to use the data in a quantitative risk assessment application, that is, in vivo genotoxicity doses can be compared to human exposure levels to derive margins of exposure, but in vitro genotoxicity concentrations cannot. Furthermore, in vivo dose context is required to improve the utility of NAM data for deriving margins of exposure, which are important metrics for risk management, as well as for potency ranking and prioritization of genotoxicants.

In vitro to in vivo extrapolation (IVIVE) offers the potential to enhance the utility of quantitative genotoxicity NAM assay data in risk assessment applications by offering human dose context to the in vitro results. Specifically, IVIVE models can translate the bioactive, or in this case genotoxic, concentrations measured in vitro (e.g., μM) into a dose expected to induce genotoxiciy in vivo by taking into account chemical disposition (i.e., absorption, distribution, metabolism, and excretion). These IVIVE-modeled doses are referred to as administered equivalent doses (AEDs; in mg/kg bw/day), and they represent the estimated dose that needs to be administered in vivo to reach a steady-state concentration in the plasma that is equal to the concentration inducing genotoxicity in the in vitro test system. Potency ranking by AEDs provides a different ranking than using in vitro data alone, with the AED ranking more relevant to human health (Rotroff et al., 2010; Wetmore et al., 2012). Thus, IVIVE models have the

potential to play an important role in the development of a framework for the interpretation of concentration-response data in a quantitative risk assessment context when evaluating potentially genotoxic substances.

In order to apply IVIVE to a broad class of chemicals, high-throughput toxicokinetics (HTTK) approaches and models have been developed and validated. These HTTK models are based on pharmacokinetic models originally developed by the pharmaceutical industry, but adaptations to these models have allowed for their application to go beyond pharmaceuticals and include environmental and industrial chemicals (Rotroff et al., 2010; Wetmore et al., 2012, 2015; Wetmore, 2015). Simplification of the model parameters has increased the coverage of application to a broad chemical space with increased confidence of implementation (Cohen Hubal et al., 2019; Wambaugh et al., 2018). Recent case studies have demonstrated that application of these models to in vitro bioactivity data allows for the derivation of AEDs to be used as surrogate points-of-departures (PODs) for risk assessment activities. Specifically, the surrogate PODs (i.e., AEDs derived from in vitro data) can be used to derive bioactivity exposure ratios (BERs). BERs are analogous to margins of exposure (MOEs) and are calculated by dividing the AED by the estimated human exposure level. These surrogate PODs derived from in vitro data were demonstrated to be protective of human health (i.e., lower) relative to PODs derived from traditional animal studies (Paul Friedman et al., 2020; Health Canada, 2021; Beal et al., 2022). This trend was observed to be consistent across hundreds of unique chemicals evaluated using these models. However, most of these previous case studies did not consider genotoxicity endpoints. Only one study has demonstrated the utility of IVIVE for deriving AEDs from in vitro genotoxicity data that are protective of human health for most chemicals (24/33 chemicals), but the analysis was limited to data from the micronucleus assay (Kuo et al., 2022). Therefore, more efforts are needed to demonstrate the ability of these models to identify potential genotoxic hazards or risks across the various genotoxicity endpoints (Health Canada, 2021).

Here we present a case study that critically evaluates the utility of applying IVIVE to concentration-response data from different types of genotoxicity NAMs. This work was conducted under the auspices of the Health and Environmental Sciences Institute (HESI) Genetic Toxicology Technical Committee (GTTC), with regulatory commentary and support provided by participating members of the Accelerating the Pace of Chemical Risk Assessment (APCRA) initiative. The GTTC IVIVE subgroup, as part of the broader in vitro genetic toxicology working group, were the leaders of the case study and consisted of experts in the genetic toxicology field from government agencies, academia, and industry. This case study focused on 31 reference chemicals, consisting of established genotoxicants, which were evaluated by in vivo studies using traditional assay protocols that are standard in genetic toxicology hazard identification, as well as in vitro genetic toxicology NAMs. AEDs from in vitro data were derived by applying quantitative benchmark concentration (BMC) modeling and IVIVE to NAM data. This case study aimed to address three main objectives: (1) determine how these NAM-based AEDs compare to PODs from relevant animal studies; (2) assess the utility of BERs derived from AEDs and human exposure estimates in chemical safety evaluations; and (3) use the lessons learned to inform a SWOT (i.e., strengths, weaknesses, opportunities, threats) analysis. It is envisioned that this approach could modernize genotoxicity data interpretation, testing approaches and strategies, as well as support the reduction of animal use in regulatory decision-making.

2 | MATERIALS AND METHODS

2.1 | New approach methodology and chemical selection

Following multiple consultations, the members of the GTTC IVIVE working group reviewed a comprehensive list of genotoxicity NAMs and identified higher-throughput NAMs that contained sufficient concentration-response data for inclusion in this case study. Assays were required to provide some measure of DNA damage response, mutagenicity, clastogenicity, or aneugenicity. The identified assays included the in vitro TGR assay, ToxTracker® (Bscl2 and Rtkn reporter genes only), PrediScreen (γH2AX and phospho-histone H3 [p-H3]), MultiFlow® (γH2AX, p53, p-H3, and polyploidy), and the in vitro MicroFlow® flow cytometry version of the micronucleus assay (i.e., Organization for Economic Co-operation and Development (OECD) test guideline 487).

The case study was limited to chemicals that were evaluated by both the aforementioned in vitro NAM assays and traditional in vivo assays. First, a list was compiled of over 300 possible chemicals with concentration-response data available across the selected NAMs. This list was cross-referenced with chemicals in the HTTK library (discussed in Section 2.3) and databases containing information on in vivo genotoxicity (discussed in Section 2.4). The expert working group examined the compiled list of chemicals and identified well-established genotoxicants for inclusion in this case study. Chemicals were limited to organic compounds (i.e., no organometallic chemicals were included) and needed to have been tested by at least one NAM as well as have known in vivo data available. NAM data were obtained from the literature or supplied by members of the working group. All experimental data required concurrent negative controls and the study design needed to be amenable to BMC modeling (sufficient number of dose groups). Data where a BMC could not be modeled due to experimental design (dose spacing, insufficient number of replicates given the variability, insufficient top concentration) were excluded.

2.2 | Benchmark concentration and benchmark dose modeling

Benchmark concentration and BMD modeling were conducted on in vitro concentration-response data and in vivo dose–response data, respectively, using a modified version of the PROAST R package (version 70.0) that allowed for batch modeling (https://github.com/

MarcBeal/GTTC/tree/main/IVWG/IVIVE). A benchmark response of 100, which corresponds to the concentration or dose required to elicit a doubling in response, was applied to be consistent with the analysis performed on ToxTracker data by Boisvert (2020) as the threshold for positive classification in reporter assays is typically around a 2-fold response. The BMC or BMD was determined based on the best fit of either model 3 (3-parameter model: includes background response, potency, and steepness parameters) or model 5 (4-parameter model: includes maximum response as fourth parameter) from the nested exponential family of models. If raw data were available in the literature or provided by working group members, they were modeled as replicate data as opposed to summary data. Otherwise, data were modeled as continuous summary data (i.e., mean response and standard deviation or standard error with sample size). As a quality control measure, a BMC_{100} or BMD_{100} value was excluded if the corresponding BMCU/BMCL or BMDU/BMDL (upper and lower confidence interval ratio) was above 100 (White, Zeller, et al., 2019). This filter removes concentration-response or dose–response data where a response is not detected or the response is much lower than the benchmark response. Only BMCs and BMDs from data showing a positive concentration- or dose–response trend were considered. ToxTracker BMC_{100} values were used as previously reported (Boisvert, 2020).

2.3 | High-throughput toxicokinetics modeling

Generic IVIVE modeling to estimate the AEDs in mg/kg bw/day was performed using the HTTK R package (version 2.2.1) (Pearce et al., 2017). Specifically, a three-compartment steady-state ("3compartmentss") model, consisting of the gut, liver, and rest of the body, was used to estimate the steady-state plasma concentration (C_{ss}) reached after simulating a constant dose rate of 1 mg/kg bw/day. The relationship between C_{ss} and dose rate is assumed to be linear, and thus, the AED/BMC ratio is proportional to the dose rate divided by the modeled C_{ss}. At a modeled dose rate of 1 mg/kg bw/day, the calculation of AED simplifies to AED = BMC/C_{ss}. In other words, the AED is the administered dose required to reach a steady-state plasma concentration equal to the compound-specific BMC determined for the selected genotoxicity NAM.

To run the 3compartmentss model, the R function calc_mc_css (which.quantile = c(0.95), model = "3compartmentss", output.units = "uM", species = "Human") was used. The 3compartmentss model requires the input of different in vitro toxicokinetic parameters and physical chemical properties: fraction unbound in the plasma protein, hepatic clearance (Cl_{int}), octanol/water partition coefficient (log *P*), and molecular weight (MW). For six of the case study chemicals, there was sufficient input data for the required input parameters within the HTTK library to model C_{ss} values. An additional eight chemicals had available in vitro toxicokinetics data generated in-house by Health Canada (HC) following the published HC approach (Health Canada, 2021). For chemicals with any missing input toxicokinetics parameters, in silico predictions were obtained using ADMET Predictor (version 10) following previously used methods (Beal et al., 2022). In the HC toxicokinetics data, there were some chemicals with negligible biotransformation. For these chemicals, the Cl_{int} was based on the lower Cl_{int} (conservative estimate) determined by the clearance rate predicted by ADMET Predictor or measured using the in vitro clearance rate where half-life was set equal to the maximum assay experimental time (i.e., half-life equals ~360 min, Cl_{int} = 3.851). The MW and log *P* data for all chemicals outside of the HTTK database came from the CompTox Chemicals Dashboard (https://comptox.epa.gov/dashboard/) (Williams et al., 2017). The add_chemtable(overwrite = F) function was used to add any missing input data to the HTTK database prior to modeling. The HTTK data used to run the 3compartmentss model are provided in Table S1. Most chemicals lack in vivo toxicokinetic data and therefore, measured in vivo C_{ss} values are not known; however, previous work has established that the HTTK model predicts C_{ss} values that are on the same order of in vivo measurements for the majority of chemicals (Wambaugh et al., 2015).

2.4 | Identification of traditional in vivo points of departure

BMD_{100} values were modeled following the procedure from Section 2.2 using data taken from the Transgenic Rodent Assay Information Database (TRAID; data freely available on request), the *Pig-a* in vivo Gene Mutation Assay Database (Shemansky et al., 2019), and from compiled in vivo micronucleus data (Soeteman-Hernández et al., 2015; 2016). For reference chemicals lacking in vivo data from the listed sources, individual searches of the Chemical Effects in Biological Systems database (Lea et al., 2017) and literature were conducted by core members of the GTTC working group to identify PODs. All no-observed-genotoxic-effect-levels (NOGELs) or lowest-observed-genotoxic-effect-levels (LOGELs) that were reported by the identified studies were noted, but only LOGELs, indicative of significant genotoxic responses, were used as an in vivo POD in this case study. The in vivo assays included the micronucleus assay, fluorescence in situ hybridization (FISH) assay, chromosomal aberration (CA) assay, aneuploidy assays, TGR assay, *Pig-a* assay, hypoxanthine phosphoribosyl transferase (*Hprt*) assay, and thymidine kinase (*Tk*) assay.

2.5 | Estimation of bioactivity exposure ratios

The Systematic Empirical Evaluation of Models 3 (SEEM3) Exposure Forecasting (ExpoCast) exposure predictions (Ring et al., 2019) were downloaded from the CompTox Chemicals Dashboard (https://comptox.epa.gov/dashboard/) (Williams et al., 2017) on April 21, 2021. ExpoCast SEEM3 is calibrated to human exposure predictions deduced from human biomonitoring data and depends on production volume as an indicator for environmental release as well as four distinct exposure pathways: consumer products (near-field),

nonpesticidal dietary, far-field industrial, and far-field pesticide. In SEEM3, a consensus approach is applied to a suite of 13 high-throughput exposure prediction models to estimate human exposure levels for each chemical. As exposure sources are unknown for most chemicals, chemical structure and physicochemical properties are input into machine learning models to quantify the relevance of chemicals to the four pathways and provide a rapid estimate of potential human exposure levels. BERs were estimated by dividing the AED of each chemical by the exposure prediction (upper 95th credible interval from the predicted population median intake rate) from ExpoCast SEEM3 (U.S. "Total") where exposure estimates were available.

3 | RESULTS

The list of reference chemicals was narrowed to 31 chemicals that had at least one source of NAM data and had in vivo data (Table 1). The included chemicals were genotoxic through various mechanisms such as DNA alkylation, adduct formation, crosslinking, topoisomerase inhibition, indirect mechanisms (e.g., oxidative damage, metabolite inhibition), or for some chemicals a combination of different modes of action. For these 31 chemicals, an AED (in vitro POD based on BMC_{100} and IVIVE) could be compared against an in vivo POD (i.e., POD based on in vivo BMD_{100} or LOGEL). There were published MicroFlow data for 12 chemicals (Allemang et al., 2021), published MultiFlow data for 13 chemicals with raw data provided by Litron Laboratories (Dertinger et al., 2019; Bryce et al., 2017), published or PrediTox-supplied PrediScreen data for 22 chemicals (Khoury et al., 2013; 2016a; 2016b; Kopp et al., 2018), published TGR data for 18 chemicals identified by White, Luijten, et al. (2019), and published or Toxys-supplied ToxTracker data for 21 chemicals (Allemang et al., 2021; Boisvert, 2020).

Across the 31 reference chemicals, a total of 198 AEDs could be modeled from in vitro data (Table S2). Specifically, BMC_{100} values could be derived from the concentration-response data and IVIVE could be applied to estimate an AED from those BMCs. There were 16 in vitro MicroFlow-based AEDs, 52 MultiFlow-based AEDs (10 24 h γH2AX, 13 24 h p53, 3 24 h p-H3, 4 24 h polyploidy, 10 4 h γH2AX, 9 4 h p53, and 3 4 h p-H3), 42 PrediScreen-based AEDs (38 γH2AX and 4 p-H3), 41 ToxTracker-based AEDs (21 *bscl2* and 20 *rtkn*), and 47 TGR-based AEDs. The AEDs ranged from 1.34×10^{-6} to 1.90×10^{4} mg/kg bw/day, with a median AED of 1.37 mg/kg bw/day and a geometric mean AED of 1.48 mg/k bw/day.

In contrast to in vitro data that used BMC_{100} values exclusively, there were many instances where in vivo BMD_{100} values could not be used for POD derivation due to limitations in the animal study design (i.e., limited dose selection). Thus, LOGELs were also reported for in vivo genetic toxicity data. For the 31 reference chemicals, there were 321 in vivo PODs (Table S3). Specifically, there were 177 BMD_{100} values and 144 LOGELs. There were 22 NOGELs identified but these were not used in this case study. There were 135 PODs related to chromosomal damage or aneuploidy (123 micronucleus assay PODs, 2 FISH assay PODs, 6 CA assay PODs, and 4 PODs from measurements of aneuploidy by cytogenetic analysis of metaphases). There were also 186 PODs from gene mutation assays (88 PODs from the TGR assay, 89 PODs from the *Pig-a* assay, 5 PODs from the *Hprt* assay, and 4 PODs from the *Tk* assay). The in vivo PODs ranged from 2.79×10^{-6} to 1.00×10^{3} mg/kg bw/day, with a median POD of 8.77 mg/kg bw/day and a geometric mean POD of 5.20 mg/kg bw/day.

3.1 | Comparison of administered equivalent doses with traditional points-of-departure

In total, a comparison could be made between 198 AEDs and 321 PODs, from a total of 31 chemicals, obtained from either in vitro AEDs or in vivo data (i.e., BMDs and LOGELs). All of the AEDs and PODs are shown in Figure 1. All AEDs and PODs for individual chemicals are also displayed in Figures S1–S31 with 90% confidence intervals provided for AEDs and PODs modeled using PROAST. Examples of these chemical-specific results are highlighted in Figure 2. 3-Nitrobenzanthrone is an example of a chemical where AEDs are clearly highly protective (i.e., over two orders of magnitude) relative to PODs, chlorambucil shows where data from indicator assays provide highly protective AEDs with some overlapping PODs, 4-nitroquinoline 1-oxide shows where the AEDs are more protective (i.e., less than two orders of magnitude lower than the POD), methyl methanesulfonate (MMS) shows where AEDs are overlapping with PODs, vinblastine shows where AEDs are overlapping or slightly higher than PODs but mechanism of action (i.e., aneugenicity) is concordant between AEDs and PODs, and glycidamide shows an example where AEDs are not protective relative to PODs. Two different POD quantiles (median and fifth percentile) were compared to account for the effects of outliers, limitations in dose/concentration range, and various genotoxic mechanisms (Figure 3). First, chemicals were arranged based on the ratio between the median in vitro AED and median in vivo POD. Most chemicals (20/31) had a median AED that was lower than the median POD. The $\log_{10}$ median POD-$\log_{10}$ median AED difference ranged from −1.9 to 3.2 with an average $\log_{10}$ difference of 0.92, indicating that on average the median AED was around 8.3-fold lower than the POD on the arithmetic scale. Eleven of the 31 median AEDs were within 10-fold of the median PODs (three AEDs lower than PODs, eight AEDS higher than PODs). Fifth percentile PODs were also compared to be consistent with the approach previously applied to in vitro bioactivity data (Paul Friedman et al., 2020; Beal et al., 2022). Approximately the same number of chemicals (21/31) had a fifth percentile AED that was lower than the fifth percentile POD. The $\log_{10}$ POD-$\log_{10}$ AED difference based on the fifth percentile values ranged from −3.5 to 4.7 with an average $\log_{10}$ difference of 1.1, indicating that on average the fifth percentile AED was also around 11.7-fold lower than the POD on the arithmetic scale. Seven of the 31 fifth percentile AEDs were within 10-fold of the fifth percentile PODs (two AEDs lower than PODs, five AEDs higher than PODs). Considering that the differences between comparison approaches (i.e., median or

TABLE 1 Summary of reference chemicals and in vitro data sources.

Compound	CASRN	Number of unique NAMs tested[a]	MicroFlow	MultiFlow	Prediscreen	TGR	ToxTracker	TK source	HTTK C_{ss} (μM)
Etoposide	33419-42-0	5	Y	Y	Y	Y	Y	HC/HTTK	3.628
N-Ethyl-*N*-nitrosourea	759-73-9	5	Y	Y	Y	Y	Y	ADMET	4.255
Camptothecin	7689-03-4	4	Y	Y	Y		Y	ADMET	10.35
Cyclophosphamide	50-18-0	4	Y		Y	Y	Y	HC/HTTK	2.278
5-Fluorouracil	51-21-8	4	Y	Y	Y		Y	ADMET/HTTK	3.626
Methyl methanesulfonate	66-27-3	4	Y	Y	Y		Y	ADMET/HC	4.86
Mitomycin C	50-07-7	4		Y	Y	Y	Y	ADMET	2.504
4-Nitroquinoline 1-oxide	56-57-5	4		Y	Y	Y	Y	ADMET	5.191
Acrylamide	79-06-1	3	Y		Y	Y		ADMET/HC	10.64
Aflatoxin B1	1162-65-8	3			Y	Y	Y	HTTK	4.425
Benzo[*a*]pyrene	50-32-8	3			Y	Y	Y	HTTK	0.0695
Chlorambucil	305-03-3	3		Y	Y		Y	ADMET/HTTK	232.1
Colchicine	64-86-8	3	Y	Y			Y	HTTK	5.99
7,12-Dimethylbenz[*a*] anthracene	57-97-6	3			Y	Y	Y	HTTK	0.686
Eugenol	97-53-0	3	Y		Y		Y	HTTK	0.912
Ethyl methanesulfonate	62-50-0	3		Y	Y		Y	ADMET	5.468
Glycidamide	5694-00-8	3		Y	Y	Y		ADMET	5.18
Griseofulvin	126-07-8	3		Y	Y		Y	ADMET	1.767
Hydroquinone	123-31-9	3	Y		Y		Y	ADMET/HC	2.416
Paclitaxel	33069-62-4	3		Y	Y		Y	HC/HTTK	2.736
Vinblastine	865-21-4	3	Y		Y		Y	ADMET/HTTK	0.0802
PhIP.HCl[b]	105650-23-5	2				Y	Y	HC	38.41
Emodin	518-82-1	2			Y		Y	ADMET	21.97
Resorcinol	108-46-3	2	Y		Y			HTTK	2.738
2-Acetylaminofluorene	53-96-3	1				Y		ADMET/HC	35.44
3-Aminobenzanthrone	13456-80-9	1				Y		ADMET	20.14
1,2-Dibromo-3-chloropropane	96-12-8	1				Y		ADMET	2.177
Dimethylnitrosamine	62-75-9	1				Y		ADMET/HC	10.19
1,8-Dinitropyrene	42397-65-9	1				Y		ADMET	24.49
N-Hydroxy-4-acetylaminobiphenyl	4463-22-3	1				Y		ADMET	30.77
3-Nitrobenzanthrone	17117-34-9	1				Y		ADMET	18.5

[a]For each chemical, a Y designates that the chemical was tested by the corresponding new approach methodology (NAM) at least once. Chemicals are arranged by the number of unique NAMs tested.
[b]2-Amino-1-methyl-6-phenylimidazo[4,5-*b*]pyridine hydrochloride abbreviated to PhIP.HCl.

fifth percentile) were marginal, only the median results are discussed below.

Using all the available data, there were 11 chemicals where the median in vitro AED was higher than the median in vivo POD, and therefore, the in vitro data is unlikely to be protective of human health compared to the in vivo data (Figure 3). The difference was small for MMS (1.2-fold higher), resorcinol (1.3-fold), ethyl methanesulfonate (EMS; 3.4-fold), eugenol (3.6-fold), dimethylnitrosamine (DMNA; 4.2-fold), cyclophosphamide (7.0-fold), glycidamide (7.5-fold), and acrylamide (9.0-fold). There were three chemicals where the median in vitro AED was more than one order of magnitude higher than the median in vivo POD. This was the case for vinblastine (19.6-fold), 1,2-dibromo-3-chloropropane (DBCP; 23.0-fold), and *N*-ethyl-*N*-nitrosourea (ENU; 85.5-fold).

Additional comparisons were made that focused on genotoxic mode of action (MOA; i.e., mutagenicity, clastogenicity, and/or

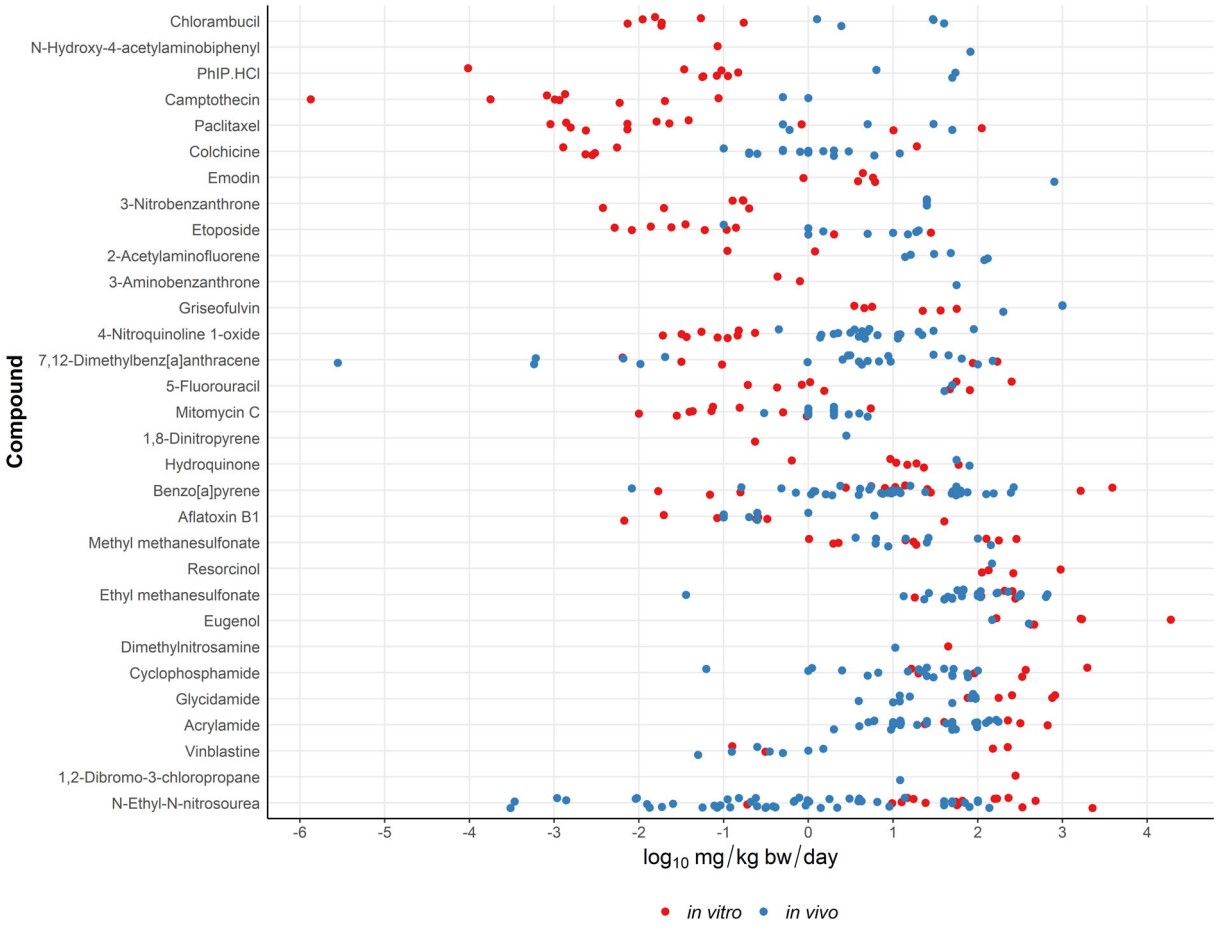

FIGURE 1 Comparison of all administered equivalent doses (AEDs) from in vitro studies and points-of-departures (PODs) from in vivo studies. The in vitro-derived AEDs (red circles) were compared against the in vivo PODs (blue circles). For the chemicals at the top, the median AEDs were orders of magnitude lower than the median PODs. In contrast, the chemicals at the bottom had median AEDs that were higher than median PODs. Confidence intervals for AEDs and PODs are displayed in the Supplementary Figures for the chemical-specific plots.

aneugenicity). There were 13 chemicals that had both in vitro and in vivo gene mutation data from any of the TGR, *Pig-a*, HPRT, and TK assays (Figure 4). The majority of gene mutation AEDs were lower than the corresponding PODs for the same chemical (8/13). The $\log_{10}$ median POD-$\log_{10}$ median AED difference ranged from −1.9 to 2.5 with an average $\log_{10}$ difference of 0.57, indicating that on average the median AED was around 3.7-fold lower than the median POD on the arithmetic scale. Three of the 13 median AEDs were within 10-fold of the median PODs. The chemicals with gene mutation AEDs higher than PODs were etoposide (1.5-fold), cyclophosphamide (6.8-fold), glycidamide (47.9), acrylamide (54.8-fold), and ENU (75.9-fold).

There were 12 chemicals that had both in vitro and in vivo micronucleus data to allow for a direct comparison (Figure 5). All but three of the chemicals (9/12) had a median micronucleus AED that was lower than the corresponding median micronucleus POD. Eugenol, 5-fluorouracil and cyclophosphamide were the chemicals with higher micronucleus AEDs, but the differences were small (1.1-, 1.3-, and 4.3-fold higher, respectively). The $\log_{10}$ median POD-$\log_{10}$ median AED difference ranged from −0.6 to 3.6 with an average $\log_{10}$ difference of 0.86, indicating that on average the median AED was around 7.3-fold lower than the POD on the arithmetic scale. Eight of the 12 median AEDs were within 10-fold of the median PODs.

Among the 31 reference chemicals, there were four chemicals with an a priori classification as likely aneugens: colchicine, griseofulvin, paclitaxel, and vinblastine (Kirkland et al., 2016; Oliver et al., 2006; Bryce et al., 2016; Dertinger et al., 2019). To investigate the application of the IVIVE approach to aneugens, we only included the AEDs derived from polyploidy or p-H3 in the comparison with in vivo PODs. It has previously been demonstrated that exposure to spindle poison-type aneugens cause p-H3-positive cells to accumulate, making this a useful biomarker for studying aneugenicity (Muehlbauer and Schuler, 2005). These AEDs were compared with in vivo PODs derived from FISH, micronucleus assay, or aneuploidy assays (Figure 6). The results show that all aneugens tested had a median AED that was lower than the median POD. The $\log_{10}$ median POD-$\log_{10}$ median AED difference ranged from 0.05 to 3.3, and the average $\log_{10}$ difference was 2.0, indicating that on average the median AED was around 90-fold lower than the POD on the arithmetic scale. Only one of the four median AEDs was within 10-fold of

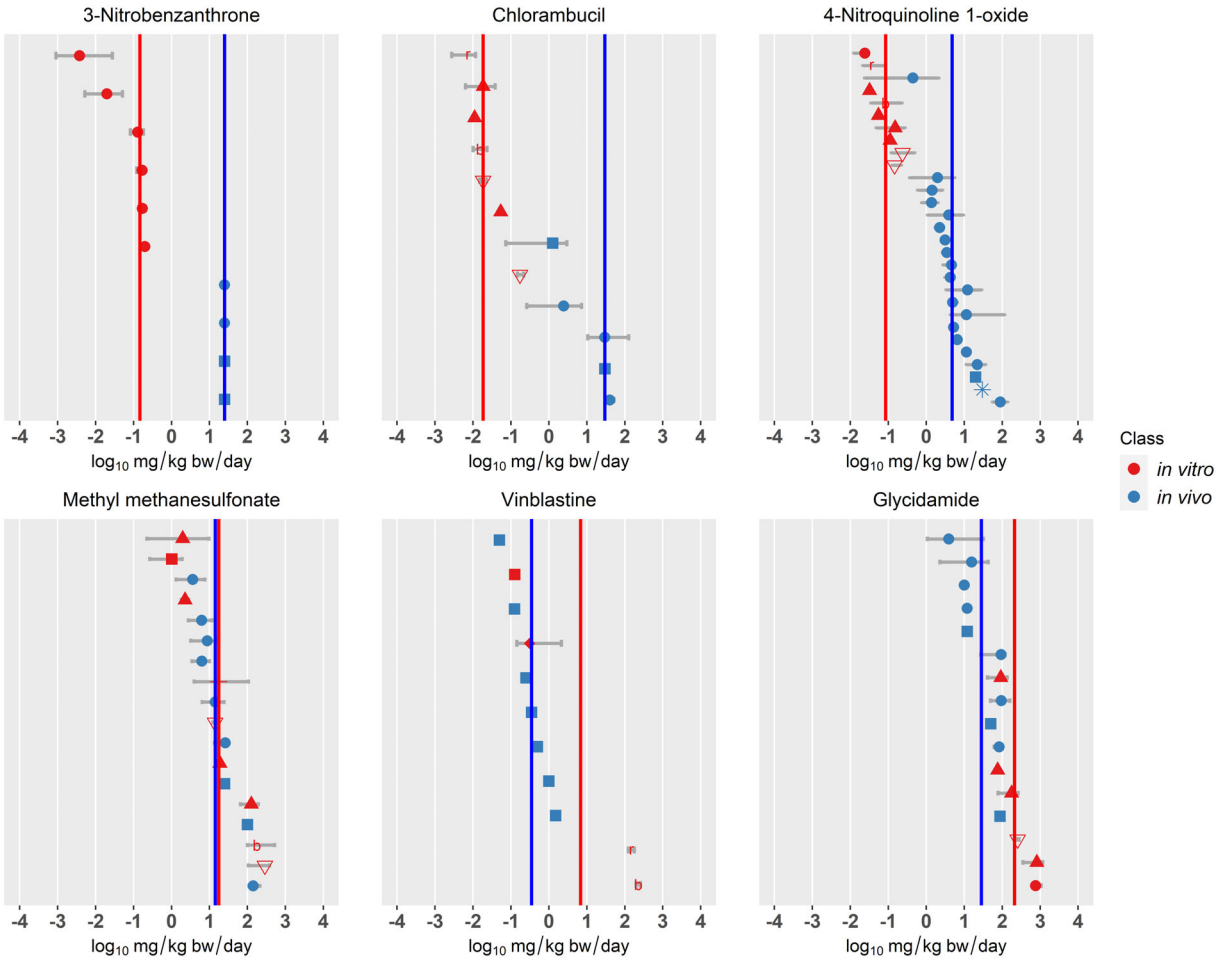

FIGURE 2 Examples of individual chemical administered equivalent dose (AED) and points-of-departures (POD) comparisons with confidence intervals. Example plots are presented for 3-nitrobenzanthrone, chlorambucil, 4-nitroquinoline 1-oxide, methyl methanesulfonate (MMS), vinblastine, and glycidamide. Red data points represent in vitro derived AEDs and blue data points represent in vivo derived PODs. The 90% confidence intervals are shown for data points based on BMD/BMC modeling. The different shapes represent different types of assays: gene mutation (circle), micronuclei (square), bscl2 (b), rtkn (r), γH2AX (solid triangle), p-H3 (diamond), p53 (inverted empty triangle), polyploidy (+), and chromosomal aberration (*). The data are arranged from lowest AED or POD at the top to the highest at the bottom.

the median PODs. Thus, based on the small number of aneugenic chemicals analyzed, the combination of NAM data and IVIVE provides a sensitive approach for deriving surrogate PODs for aneugens that errs on the side of caution.

Finally, a comparison was made between available AEDs derived from measuring changes in DNA damage markers (γH2AX, p53, bscl2, and rtkn) and all available in vivo PODs (Figure 7). The AEDs based on the γH2AX biomarker were compared to PODs for 20 chemicals, and for 13 chemicals the median AED was lower than the median POD. The average difference of $\log_{10}$ medians was 0.57, indicating that on average the median AED based on γH2AX was around 3.7-fold lower on the arithmetic scale. The AEDs based on the p53 marker were compared to PODs for 13 chemicals, and for 9 chemicals the median AED was lower than median POD. The average difference of $\log_{10}$ medians was 0.90, indicating that on average the median AED based on p53 was around 7.9-fold lower on the arithmetic scale. The AEDs based on the bscl2 marker were compared to PODs for 20 chemicals, and for 13 chemicals the median AED was lower than median POD. The average difference of $\log_{10}$ medians was 0.54, indicating that on average the median AED based on bscl2 was around 3.5-fold lower on the arithmetic scale. The AEDs based on the rtkn marker were compared to PODs for 19 chemicals, and for 11 chemicals the median AED was lower than median POD. The average difference of $\log_{10}$ medians was 0.71, indicating that on average the median AED based on rtkn was around 5.1-fold lower on the arithmetic scale. The chemicals where AEDs were not protective were consistent with previous comparisons (i.e., acrylamide, cyclophosphamide, EMS, ENU, eugenol, glycidamide, MMS, resorcinol, and vinblastine) with the additions of benzo[*a*]pyrene, colchicine, and 7,12-dimethylbenz[*a*]anthrene as not protective. In the case of colchicine, the AED based on p53 was protective relative to the in vivo PODs but the AED based on rtkn was not. These results highlight that the biomarker AEDs tend to be closely aligned with in vivo PODs, but that multiple biomarkers may need to be considered in a weight-of-evidence assessment to capture the different forms of genotoxicity.

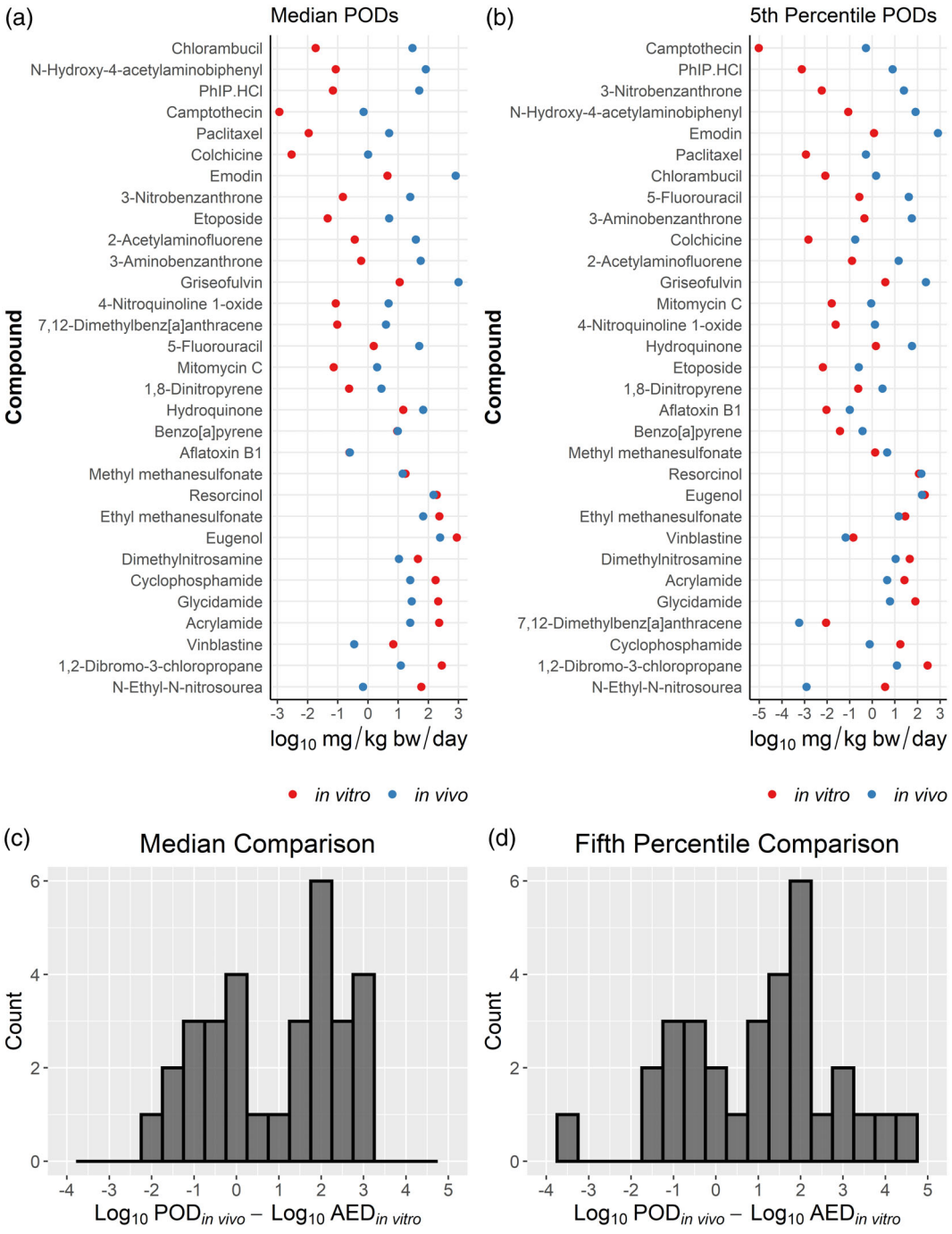

FIGURE 3 Comparison of administered equivalent doses (AEDs) and points-of-departures (PODs) using median or fifth percentile values. In panel A, the median in vitro-derived AEDs (red circles) were compared against the median in vivo PODs (blue circles). For the chemicals at the top, the median AEDs were orders of magnitude lower than the median PODs. In contrast, the chemicals at the bottom had median AEDs that were higher than median PODs. In panel B, the fifth percentile in vitro-derived AEDs (red circles) were compared against the fifth percentile in vivo PODs (blue circles). The bottom histograms display the $\log_{10}$ differences between in vivo PODs and in vitro AEDs (C: median; D: fifth percentile).

3.2 | Derivation of bioactivity exposure ratios

ExpoCast exposure estimates were available for 19 of the 31 reference chemicals. To be as protective as possible, the minimum AED (lowest AED obtained from modeling in vitro data) was compared to the 95th percentile exposure estimate to derive a BER for each of the 19 chemicals (Figure 8; Table S4). BERs were separated into four bins of varying levels of risk potential based on Beal et al. (2022). Specifically, chemicals were assigned to bins of $\log_{10}$BER < 0 (minimum AED below exposure prediction), $\log_{10}$BER 0–2 (minimum AED within 100-fold of

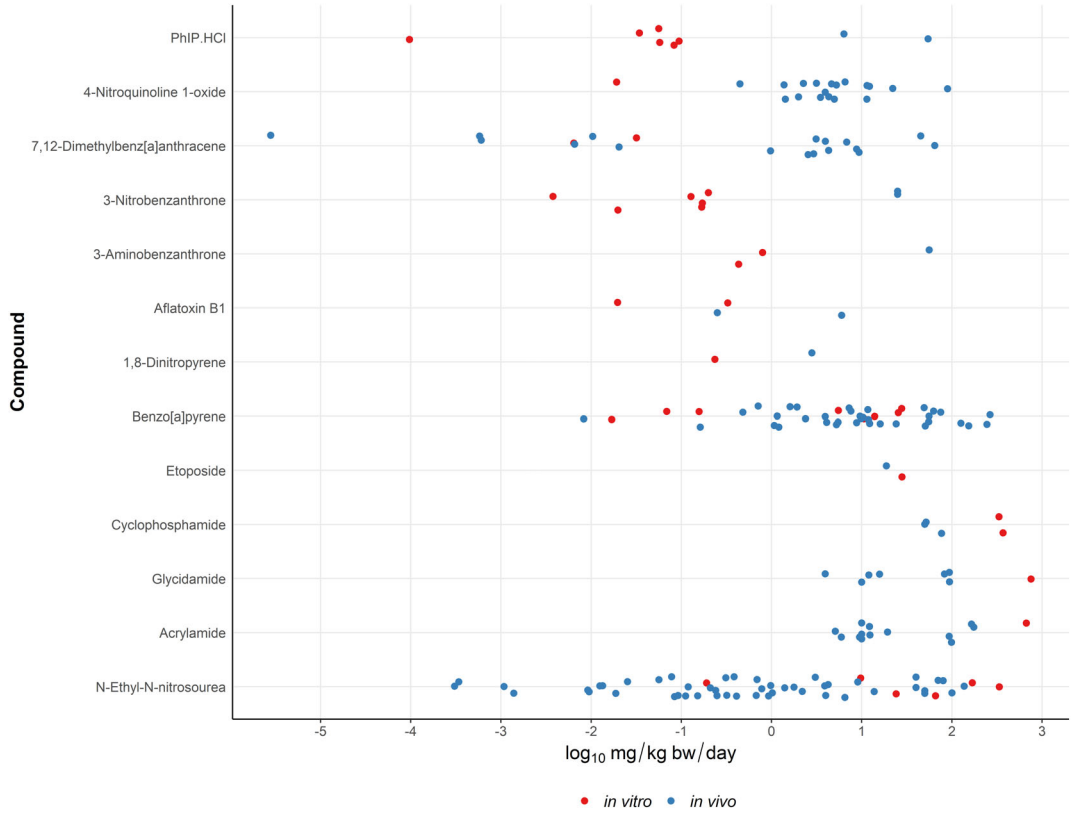

FIGURE 4 Comparison of gene mutation administered equivalent doses (AEDs) and points-of-departures (PODs). The in vitro-derived AEDs from gene mutation data (red circles) were compared against the in vivo PODs from gene mutation data (blue circles).

exposure prediction), $\log_{10}$BER 2–3 (minimum AED above 100 but within 1000-fold of exposure prediction), and $\log_{10}$BER > 3 (minimum AED greater than 1000-fold of exposure prediction). Previous work based on ToxCast bioactivity data demonstrated that the first two bins ($\log_{10}$BER < 0 and $\log_{10}$BER 0–2) were able to identify chemicals concluded to be toxic to human health or the environment based on risk assessments under Section 64 of the Canadian Environmental Protection Act (CEPA), 1999 (Health Canada, 2021). However, this approach was limited to nongenotoxic effects. Applying the same bins to 19 genotoxic chemicals, we determined that one chemical was categorized in the $\log_{10}$BER <0 bin as having the highest potential of concern (etoposide). In addition, there were seven chemicals with a $\log_{10}$BER between 0–2 (benzo[*a*]pyrene, colchicine, 7,12-dimethylbenz[*a*]anthracene, hydroquinone, acrylamide, MMS, and ENU). The remaining chemicals had lower potential for concern with a $\log_{10}$BER between 2 and 3 (nitroquinoline 1-oxide, EMS, eugenol, and emodin), and a $\log_{10}$BER above 3 (5-fluorouracil, resorcinol, 1,8-dinitropyrene, griseofulvin, DMNA, DBCP, and glycidamide).

4 | DISCUSSION

Here we present a collaborative retrospective analysis comparing AEDs, derived from in vitro concentration-response genotoxicity data and IVIVE modeling, with in vivo PODs, derived from traditional animal genotoxicity studies. This work builds on earlier studies that applied a similar approach to quantitative high-throughput screening data, but lacked a thorough assessment of the various mechanisms for genotoxicity (Paul Friedman et al., 2020; Beal et al., 2022; Health Canada, 2021). There were three main objectives for this work: (1) determine how in vitro derived AEDs compare to in vivo PODs for the same chemicals; (2) determine whether the derivation of BERs helps to support the ranking of chemicals based on potential for concern; and (3) use the lessons learned from the case study to develop a SWOT analysis.

The results of this case study demonstrated that the application of IVIVE to in vitro genotoxicity data yielded AEDs that are lower or equal to PODs from animal studies for most (65%) of the chemicals tested (Figure 1). Ignoring genotoxic MOA, the median AEDs were lower than median PODs for 20 chemicals, slightly higher for eight chemicals (i.e., 9-fold and under), and one order of magnitude higher for three chemicals (Figure 3). When the analysis was based on MOA (i.e., mutagenicity, clastogenicity, and aneugenicity), the proportion of chemicals with lower AEDs than in vivo PODs increased for most MOAs. Specifically, there were 8/13, 9/12, and 4/4 chemicals with lower AEDs than PODs for gene mutation (Figure 4), clastogenicity measured by micronuclei (Figure 5), and aneugen markers/aneuploidy (Figure 6), respectively. Analyzing in vitro DNA damage response biomarkers (γH2AX, p53, bscl2, and rtkn) alone also provided highly convergent AEDs relative to in vivo PODs (Figure 7). On average, the

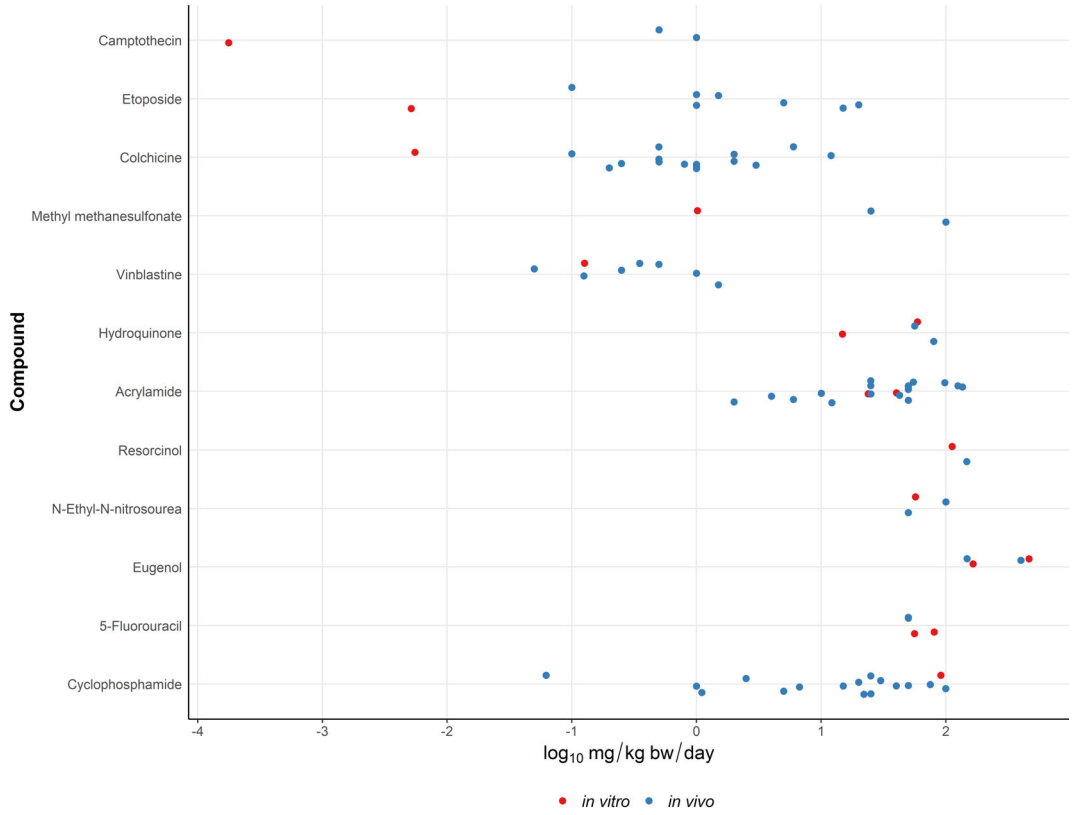

FIGURE 5 Comparison of micronucleus administered equivalent doses (AEDs) and points-of-departures (PODs). The in vitro-derived AEDs from micronucleus data (red circles) were compared against the in vivo PODs from micronucleus data (blue circles).

AEDs based on DNA damage response biomarkers were on the same order of magnitude as in vivo PODs (between 3.5- and 7.9-fold lower on average for the four markers). Overall, the results show that for the majority of the chemicals, the application of IVIVE to in vitro genotoxicity concentration-response data can provide a reasonable approximation of a lower bound estimate of in vivo effect levels.

There were some chemicals assessed in this study that had higher AEDs than in vivo PODs across analyses. In some cases, the AEDs were only slightly higher than in vivo PODs. For example, eugenol (Figure S20) and resorcinol (Figure S30) had median AEDs that were 3.6-fold and 1.3-fold higher than in vivo PODs, respectively. During compilation of in vivo PODs, it was determined that most studies on eugenol and resorcinol reported a NOGEL (6/8 and 6/7 of PODs, respectively; NOGEL results not shown nor used in analysis). Thus, in most cases at guideline-required maximum doses, no genotoxic effects were seen in vivo. These chemicals have been previously categorized as "misleading positives" as they often produce negative results in vivo and positive results in vitro (Kirkland et al., 2016). Thus, the high AEDs for these chemicals (165–18,967 and 112–951 mg/kg bw/day for eugenol and resorcinol, respectively) are confounded by the lower or equivocal genotoxic potential of these chemicals.

Vinblastine (Figure S31) had a median AED 19.6-fold higher than the median in vivo POD when MOA was not taken into account. However, vinblastine is an aneugen and when the analysis focused on markers of aneugenicity (i.e., p-H3), the AED was highly correlated with the in vivo PODs (Figure 2). DMNA (Figure S16) and DBCP (Figure S1) had AEDs that were 4.2-fold and 23.0-fold higher than in vivo PODs, respectively. However, the comparison was limited to one in vitro gene mutation study and one in vivo micronucleus study each. Thus, the discrepancies between AEDs and PODs may be related to the differences in endpoints evaluated for these chemicals. These findings highlight the importance of considering MOA in the assessment of both genotoxicity and potency.

Cyclophosphamide (Figure S15) had an AED that was 7.0-fold higher than the in vivo POD. For the in vitro data, weak responses were observed using the TGR assay in cryopreserved (Luijten et al., 2016) and fresh (Zwart et al., 2012) hepatocytes. A weak response was also observed using the PrediScreen assay in HepG2 cells (Khoury et al., 2013). The ToxTracker system provided AEDs most closely aligned with in vivo PODs, but the addition of S9 was required for metabolic activation (Boisvert, 2020; Wills et al., 2021). Cyclophosphamide is a pro-mutagen that requires activation by cytochrome P450 enzymes, such as CYP2B6 (Kirkland et al., 2016). It is possible that enzyme expression was not sufficiently present in the cells that yielded a weak response. In a recent study by Seo et al. (2022), it was revealed that three-dimensional spheroids from HepaRG cells had CYP2B6 activity that was 6.7- to 25.0-fold higher than two-dimensional cell cultures, indicating that three-dimensional models might be the preferable to studying the genotoxicity of genotoxic chemicals requiring metabolic activation. Furthermore, the

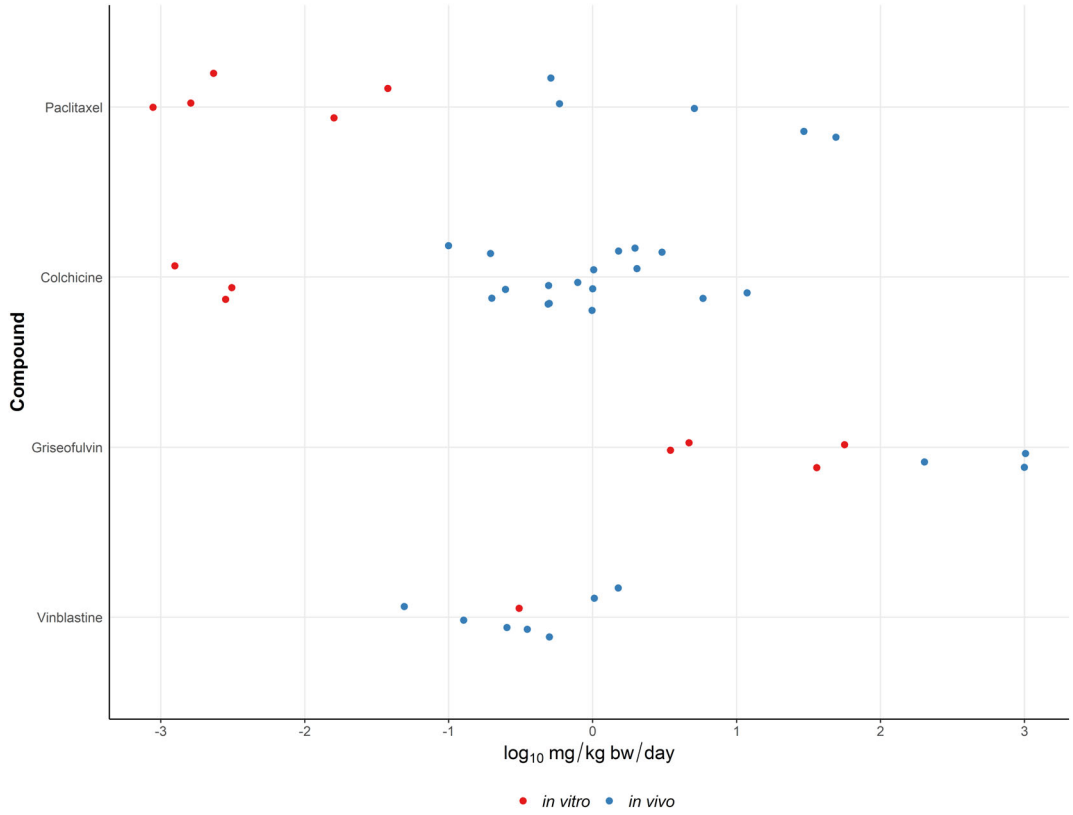

FIGURE 6 Comparison of aneugen administered equivalent doses (AEDs) and points-of-departures (PODs). The in vitro-derived AEDs (red circles) were compared against the in vivo PODs (blue circles). Only data from assays or markers measuring aneuploidy or polyploidy (i.e., in vitro p-H3 and polyploidy; in vivo FISH, MN, and aneuploidy) were included in the comparison.

cyclophosphamide results highlight the challenge associated with using a generic model in IVIVE application. Specifically, the IVIVE was based on the disposition of the parent compound (cyclophosphamide) as a surrogate for the disposition of the active metabolite(s). Thus, interpretation of AEDs for genotoxic substances requiring metabolic activation needs careful consideration during quantitative assessment of hazard or risk, and there will be a need to prioritize the use of chemical-specific models for certain chemicals.

There were three chemicals that consistently had higher AEDs than in vivo PODs regardless of MOA considerations. Acrylamide (Figure S9) and its metabolite glycidamide (Figure S21) had median AEDs that were 9.0-fold and 7.5-fold higher than median in vivo PODs, respectively. In addition, ENU (Figure S26) is a potent mutagen and is often used as a positive control. The median AED for ENU was 85.5-fold higher than the median in vivo POD. It is possible that there are considerations related to disposition of these chemicals that are unaccounted for by the generic HTTK model. For example, ENU is highly unstable at a neutral pH and is hydrolyzed rapidly (Arcos et al., 2013), and previous measurements have estimated that the half-life of ENU in phosphate buffered saline could be as low as 10 min at neutral pH (Tosato et al., 1987). Thus, the nominal concentration for ENU in the aqueous phase that is used in the derivation of AEDs may be drastically higher than the true concentration that the cells are exposed to. Other sources of toxicokinetics data or other computational models, such as mass balance models (Armitage et al., 2014) or higher-tier physiologically-based toxicokinetic (PBTK) models, may help minimize discrepancies between AEDs and PODs for certain chemicals. For example, this could help lower the AED for volatile chemicals such as DBCP and is worth exploring for both acrylamide and glycidamide. However, unstable chemicals like ENU may not be appropriate for these types of analyses. Further work is required to establish the domains of applicability and exclusion criteria for the application of toxicokinetic models to genotoxicants. Without established exclusion criteria and application guidelines, it will be difficult to determine the chemicals where IVIVE application can be considered reliable.

Traditionally, application of genotoxicity data has been limited to hazard identification and potency ranking of genotoxic potential is rarely used (White et al., 2020). A detailed case for why the assumptions perpetuating the screen-and-bin paradigm for genotoxic chemicals are false was presented in White et al. (2020). Specifically, the three false assumptions are that (1) genotoxic chemicals are rare and exposures are avoidable; (2) genotoxicity dose–response relationships do not contain a low-dose region where the response is indistinguishable from background (i.e., dose region mechanistically characterized by zero-order kinetics); and (3) genotoxicity cannot be regarded as a bona fide toxicological endpoint. The acknowledgement that these assumptions are false warrants a shift towards the use of quantitative

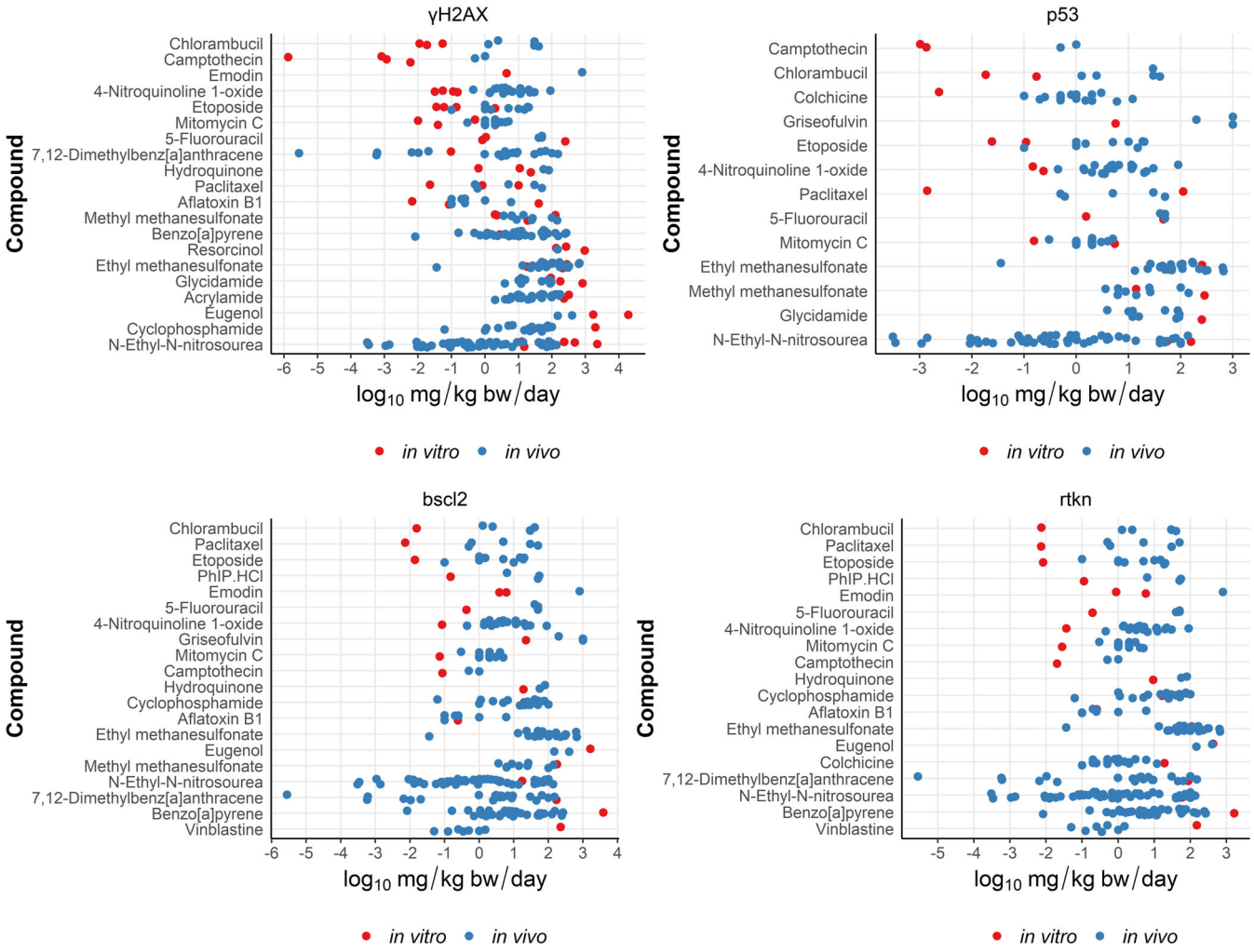

FIGURE 7 Comparison of administered equivalent doses (AEDs) based on DNA damage indicator assays and all points-of-departures (PODs). PODs were based on all available in vivo data and AEDs were based on biomarkers from DNA Damage indicator assays: γH2AX (MultiFlow and PrediScreen), p53 (MultiFlow), bscl2 (ToxTracker), and rtkn (ToxTracker).

analyses in the assessment of genotoxic chemicals. For example, recent work conducted under the GTTC demonstrated that derivation of MOEs based on mutagenicity-based BMDs yielded similar regulatory decisions as carcinogenicity-derived MOEs for most chemicals (Chepelev et al., 2022). Thus, the case has been made that MOEs based on in vivo genotoxicity data can support risk assessment. However, there has been no detailed investigation on the application of genotoxicity-based BERs, which are analogous to MOEs based on in vivo data for assessing risk.

In this case study, we derived BERs for 19 chemicals with available in vitro genotoxicity data and modeled exposure estimates (Figure 8). The minimum AED was used in BER derivation to be conservative and demonstrate how the BER approach could be used in screening or prioritization. Other risk assessment activities may require more careful evaluation of the AEDs and exposure values used in BER derivation. The chemicals with exposure estimates were ranked by BER and were categorized into separate bins capturing varying levels of potential for concern. Etoposide was the only chemical in the $\log_{10}$BER < 0 bin and identified as having a high potential for concern. Etoposide is widely used during chemotherapy due to its ability to inhibit the topoisomerase II enzyme, thereby damaging DNA in rapidly dividing cancer cells (Attia et al., 2003). The International Agency for Research on Cancer (IARC) classified Etoposide as a carcinogen by itself (group 2A: probably carcinogenic) and in combination with other drugs (group 1: carcinogenic) (IARC, 2000). Thus, the BER approach, using ExpoCast to predict the potential for exposure for nontherapeutic uses, was effective in identifying etoposide as a chemical with a higher potential for concern. However, evaluation of the uses and exposures that are not related to use as an ingredient in drugs in the Government of Canada screening assessment report of etoposide, concluded that the lowest therapeutic dose was several orders of magnitude higher than what humans are exposed to through environmental media. As such, risks from indirect exposure to etoposide are likely to be negligible (Environment and Climate Change Canada, 2014a). In addition to etoposide, benzo[*a*]pyrene and 7,12-dimethylbenz[*a*]anthrene also had very low BERs

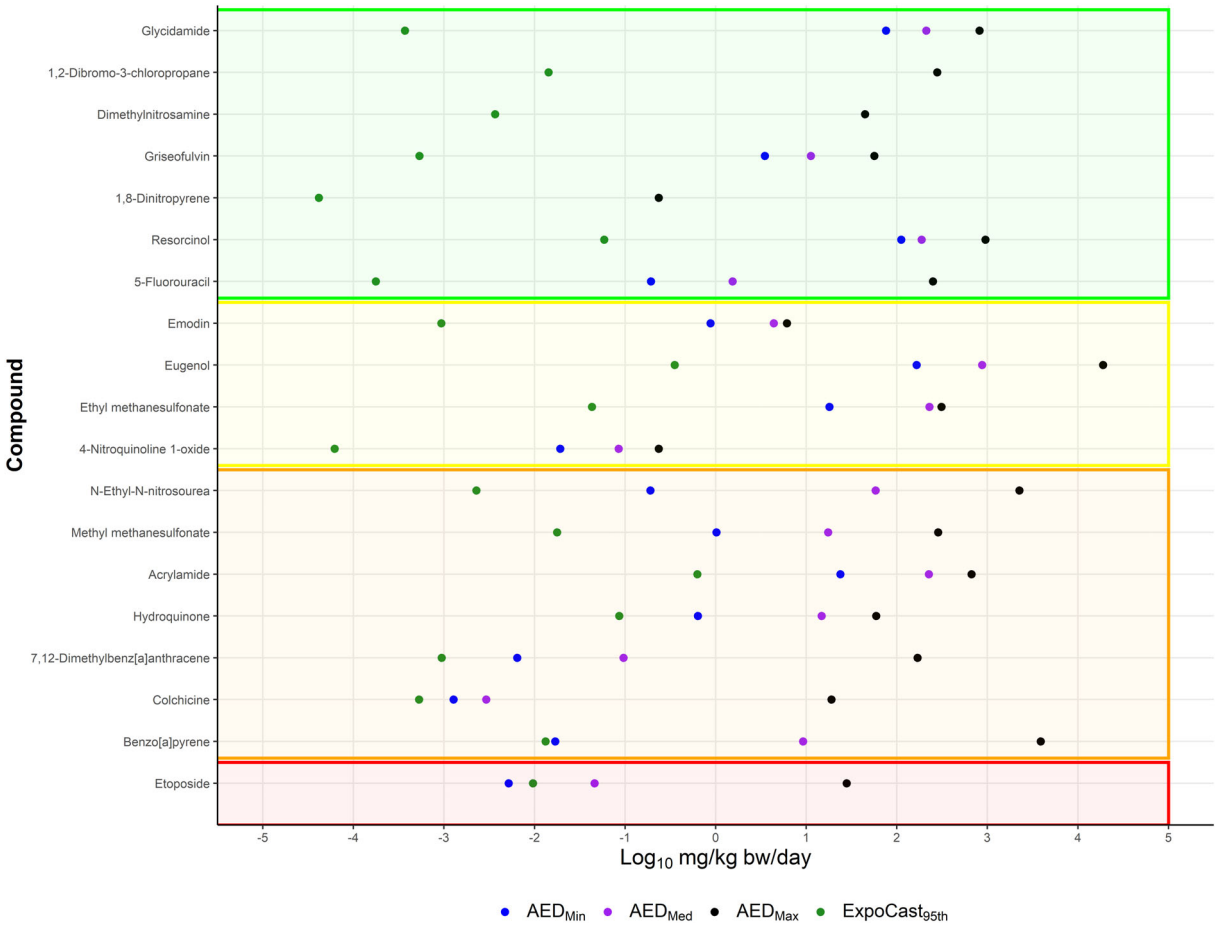

FIGURE 8 Bioactivity exposure ratios. Chemicals were ranked by the bioactivity exposure ratio (BER) calculated by dividing the minimum administered equivalent dose (AED) by the 95th percentile ExpoCast exposure prediction. Red box indicates $\log_{10}$BER < 0, orange box displays $\log_{10}$BER 0–2, yellow box displays $\log_{10}$BER 2–3, and green box indicates $\log_{10}$BER > 3.

($\log_{10}$BERs were 0.11 and 0.83, respectively); these chemical have been identified as potent mutagens that are also ubiquitous in the environment (IARC, 2010). These findings suggest that the BER approach is a useful early tier screening tool capable of identifying genotoxicants based on the certainty of the hazard identification, but further refinement of estimated exposure is required for quantitative risk assessment purposes.

In contrast to the low BER chemicals, the chemicals with the highest BERs (i.e., $\log_{10}$BER > 3) include environmental contaminants with regulated exposure levels (DBCP and DMNA) or very low exposure levels (1,8-dinitropyrene). DBCP and DMNA are contaminants in ground water and chlorinated drinking water, respectively. DBCP was previously used as a nematocide in the United States but was eventually banned from use by the US EPA in the late 1970s (Olsen et al., 1995). DMNA is a by-product of drinking water chlorination, and allowable levels and strict reporting measures have been put in place by regulatory agencies to limit DMNA exposures (Mitch et al., 2003). The high BERs for the environmental contaminants are consistent with the regulations in place to minimize exposures.

The high BER group also includes chemicals that are used as common therapeutics meeting important health needs due to their risk/benefit profile. Specifically, 5-flourouracil is a nitrogenous base analogue used to treat cancers and viral infections, and IARC classified 5-fluorouracil as group 3 due to inadequate or limited carcinogenicity data: "The agent (mixture) is unclassifiable as to carcinogenicity in humans" (IARC, 1987). Similarly, resorcinol, which is used in production of adhesives and is used in acne medication, was classified as group 3 by IARC (IARC, 1987). In contrast, griseofulvin is an antifungal agent that has been classified as a group 2B carcinogen and "is possibly carcinogenic to humans" as a result of "sufficient evidence" in experimental animals but "inadequate evidence" in humans (IARC, 2001). However, griseofulvin is on the World Health Organization's list of essential medicines (WHO, 2021) and a risk assessment concluded that griseofulvin was not entering the environment in a quantity that may constitute a danger to human health (Environment and Climate Change Canada, 2014b). The higher BERs for these chemicals used as therapeutics seem to reflect both a combination of their lower exposure levels and relatively low genotoxic potential.

Glycidamide was identified as the chemical with the highest BER; however, considering that glycidamide is formed through the epoxidation of acrylamide and is the main genotoxic metabolite of acrylamide (National Toxicology Program, 2014; EFSA Panel on Contaminants in

TABLE 2 IVIVE genotoxicity-approach: Strengths, Weaknesses, Opportunities, Threats (SWOT) analysis

IVIVE genotoxicity-approach	
Strengths: • Compliant with 3Rs; in vivo OECD genotoxicity test guidelines require at the very least 20 animals per chemical • Establishes framework for interpretation of in vitro NAM-based concentration-response data in a quantitative risk assessment context • Contributes to modernization of risk assessment approach for data-poor chemicals • Expands on quantitative framework for interpretation of genotoxicity concentration-response data • Approach allows for screening and potency ranking of chemicals with modeled human dose context • Multi-assay assessment provides mechanistic insight into genotoxic mode of action • Genotoxicity data coupled with IVIVE models provide surrogate PODs protective of human health for most chemicals • Open source computational models allow for transparency and transferability of IVIVE approach	*Opportunities*: • Other types of NAM data can be integrated into this approach • Approach could be applied to human cell model to study anticipated effects in target organ (i.e., human HepaRG instead of rodent liver) • Automated computational workflow can be applied to a broad chemical space; such approach can potentially increase the throughput of NAM-based chemical screening • Establishment of a domain of applicability by refining/standardizing in vitro test conditions and understanding the impact of key parameters on the outcomes of IVIVE for increased confidence in model implementation • Mass balance model and higher tier toxicokinetics models may improve the accuracy of IVIVE and minimize the discrepancies between AEDs and PODs • Deriving bioactivity exposure ratios (i.e., margins of exposure) can be used for quantitative risk assessment activities (i.e., chemical prioritization and screening) • Increasing the confidence of models will support transition to in vitro only testing framework • Prospective studies with comparable designs and broader scope of chemicals and NAMs will increase the understanding of the correlation between AEDs and PODs
Weaknesses: • Unclear on critical effect size to use in deriving in vitro BMCs; some weaker responses may have been omitted when applying computational workflow • The in vitro-derived PODs are extremely low for certain chemical classes relative to in vivo PODs • For some chemicals, the in vitro AEDs are higher than the in vivo PODs • For pro-mutagenic compounds requiring metabolic activation, the current approach relies on kinetics of parent compound as a surrogate • The in vitro cell models employed in certain NAMS lack metabolic capabilities; thus, addition of an exogenous metabolism system (i.e., S9) with uncertain human comparability is needed for manifestation of the genotoxicity • Uncertainty of optimal in vitro exposure duration and timing of the assays and standardization across labs • Key toxicokinetic properties for certain test chemicals may not have been accounted for by the generic HTTK model, leading to large discrepancies between AEDs and PODs • Requires technical expertise for concentration-response analysis, computational modeling, and interpretation of the analyses in the context of human exposure • Uncertainty related to variability and reproducibility of in vivo PODs used in comparison with in vitro AEDs (Ly Pham et al., 2020)	*Threats*: • IVIVE modeling parameter assumptions have not been uniformly specified or accepted for particular regulatory applications • Currently no standardization of the testing conditions for the NAMs exist • Regulatory hesitance to adopt nonanimal-based approaches for risk assessment; currently no regulatory guidance or test guidelines

Abbreviations: AED, administered equivalent dose; BMC, benchmark concentration; HTTK, high-throughput toxicokinetics; IVIVE, in vitro to in vivo extrapolation; NAM, new approach methodology; OECD, Organization for Economic Co-operation and Development; PODs, points-of-departures.

the Food Chain (CONTAM), 2015), the BERs of these two chemicals should be evaluated together. Glycidamide and acrylamide have AEDs on the same order of magnitude but very different exposure estimates (Table S4), the latter exposure estimate of acrylamide being more applicable to the existing risk assessment on acrylamide. In the EFSA's Scientific Opinion on acrylamide in food (EFSA Panel on Contaminants in the Food Chain (CONTAM), 2015; European Food Safety Authority (EFSA) et al., 2022), it was concluded that the MOEs for cancer-related effects of acrylamide (425 for adults; 50 for toddlers) were in a range that was indicative of a concern for public health. Interestingly, the BER for acrylamide based on the median and minimum AEDs were 361 and 38 and these BERs were slightly lower than the MOE for adults and toddlers, respectively. Thus, the combined assessment of glycidamide and acrylamide BERs supports the utility of genotoxicity-derived BERs as an effective screening tool in chemical safety evaluations.

An intended outcome of the presented case study was to use the results and lessons learned to inform a SWOT analysis. The specific purpose was to encourage a discussion on the usefulness and feasible risk assessment applications of the IVIVE/genetox NAM approach. The analysis was performed generically and not limited to specific endpoints or assays. The outcomes of the SWOT analysis are given in Table 2. Although the application of IVIVE to genotoxicity NAMs is still in its infancy, apparent by the presence of threats and weaknesses, there are several notable strengths of the approach and opportunities to apply the approach in certain risk assessment activities.

This case study represents an early step in building confidence in the application of IVIVE to in vitro genotoxicity data in support of chemical risk assessment modernization. The results demonstrate that the IVIVE approach can provide a reasonable lower bound estimate of the effect levels for in vivo genotoxicity. Furthermore, BERs derived from modeled AEDs can be used as a tool in screening and prioritization efforts. There were some limitations associated with the retrospective design of the case study. For example, some traditional in vivo PODs were based on study designs with limited dose range or insufficient maximum doses, making it difficult to quantify a precise in vivo effect level for comparison with AEDs. Thus, for some chemicals there were differences in in vivo PODs across studies that varied by orders of magnitude (Figure 1). To account for this, the median AEDs and PODs were compared (Figure 3), and 90% confidence intervals were considered on a chemical-by-chemical basis (Figure 2; Supplementary Figures). There were varying levels of data availability for AEDs and PODs (one to 50 values), and in some instances, the median AED or POD was based on a single value. Another limitation was that there are multiple genotoxic endpoints that can be considered, and the available endpoints for comparison were not always consistent between in vitro and in vivo studies. In an effort to modulate the comparison between in vitro AEDs and in vivo PODs, comparisons were made by MOA where possible. Furthermore, many of the in vivo PODs were based on LOGELs as the study designs were insufficient for BMD modeling; thus, there was a trade-off between comparison reliability (in some cases) and the number of chemicals that could be included in the case study. To build on this work, future prospective case studies comparing in vitro and in vivo studies with similar designs and a broader scope of both the chemicals and methods are needed. For example, these studies should also explore other computational models (e.g., in vitro disposition models, higher-tier PBTK models), other routes of exposures (e.g., systemic or potentially inhalation exposures in organotypic air-liquid-interface airway models (Wang et al., 2021), dermal exposures in 3D skin models), and more complex in vitro test systems (e.g., repeated dosing in liver spheroids, multi-organ-on-a-chip systems, or microphysiological systems). The exploration of these novel in vitro test systems and computational models will presumably improve the extrapolation of in vitro genotoxicity data to human responses and lead to the formation of novel genotoxicity testing strategies that have a lower reliance on the use of animals. These efforts will ultimately support the modernization of risk assessment.

AUTHOR CONTRIBUTIONS

Marc A. Beal lead the study design and prepared the manuscript draft with important intellectual input from all the authors. Data compilation was carried out by Marc A. Beal, Marc Audebert, Hannah Battaion, Jeffrey C. Bemis, Stephen D. Dertinger, Laure Khoury, Alexandra S. Long, Raja S. Settivari, Shamika Wickramasuriya, and Paul White. All authors approved the final manuscript.

ACKNOWLEDGMENTS

We would like to thank all the GTTC participants that provided valuable feedback throughout the duration of this case study. We would also like to thank David Kirkland and the GTTC in vivo working group members for sharing useful data for this project, Sunil Kulkarni for assistance with ADMET Predictor, Stephanie Smith-Roe for assistance with the Chemical Effects in Biological Systems database, and Nikolai Chepelev for supporting TRAID. We appreciate all the helpful comments provided by reviewers at Health Canada: Deon Williams, Carmen Cheung, Matthew Gagne, and Nikolai Chepelev; reviewers at the U.S. Food and Drug Administration: Susan Burks; and anonymous reviewers at HESI. This article reflects the views of the authors and does not necessarily reflect the position of their agencies or institutions. Any mention of commercial products is for clarification only and is not intended as approval, endorsement, or recommendation.

CONFLICT OF INTEREST

Marc Audebert and Laure Khoury are co-founders of Preditox SAS, a company specialized in genotoxicity prediction. Laure Khoury is CEO and Marc Audebert serves as consultant to Preditox SAS. Jeffrey C. Bemis and Stephen D. Dertinger are employed by Litron Laboratories, a company that sells MicroFlow and MultiFlow reagents kits, and offers testing services based on these assays. Giel Hendriks is employed by Toxys, a company that offers ToxTracker kits and services.

ORCID

Marc A. Beal https://orcid.org/0000-0003-2760-4368
Roland Froetschl https://orcid.org/0000-0001-8721-7509
Xiaoqing Guo https://orcid.org/0000-0001-5951-2156
George Johnson https://orcid.org/0000-0001-5643-9942
Stefan Pfuhler https://orcid.org/0000-0001-8869-5975

REFERENCES

Allemang, A., De Abrew, K.N., Shan, Y.K., Krailler, J.M. & Pfuhler, S. (2021) A comparison of classical and 21st century genotoxicity tools: a proof of concept study of 18 chemicals comparing in vitro micronucleus, ToxTracker and genomics-based methods (TGx-DDI, whole genome clustering and connectivity mapping). *Environmental and Molecular Mutagenesis*, 62, 92–107.

Arcos, J.C., Argus, M.F. & Woo, Y.-T. (2013) *Aliphatic carcinogens: structural bases and biological mechanisms*. New York, New York: Academic Press, Inc.

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.

Attia, S.M., Kliesch, U., Schriever-Schwemmer, G., Badary, O.A., Hamada, F.M. & Adler, I.-D. (2003) Etoposide and merbarone are

clastogenic and aneugenic in the mouse bone marrow micronucleus test complemented by fluorescence in situ hybridization with the mouse minor satellite DNA probe. *Environmental and Molecular Mutagenesis*, 41, 99–103.

Beal, M.A., Gagne, M., Kulkarni, S.A., Patlewicz, G., Thomas, R.S. & Barton-Maclaren, T.S. (2022) Implementing in vitro bioactivity data to modernize priority setting of chemical inventories. *ALTEX*, 39, 123–139.

Boisvert, L. (2020) Critical examination of selected aspects of the ToxTracker® In vitro genotoxicity assay: evaluation of S9 metabolic activation protocols and quantitative interpretation of dose-response data. Master's Thesis - University of Ottawa, 245.

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

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

Bryce, S.M., Bernacki, D.T., Smith-Roe, S.L., Witt, K.L., Bemis, J.C. & Dertinger, S.D. (2018) Investigating the generalizability of the MultiFlow® DNA damage assay and several companion machine learning models with a set of 103 diverse test chemicals. *Toxicological Sciences*, 162, 146–166.

Chepelev, N., Long, A.S., Beal, M., Barton-Maclaren, T., Johnson, G., Dearfield, K.L. et al. (2022) Establishing a quantitative framework for regulatory interpretation of genetic toxicity dose-response data: MOE (margin of exposure) case study of 48 compounds with both in vivo mutagenicity and carcinogenicity dose-response data. *Environmental and Molecular Mutagenesis*. https://doi.org/10.1002/em.22517

Cimino, M.C. (2006) Comparative overview of current international strategies and guidelines for genetic toxicology testing for regulatory purposes. *Environmental and Molecular Mutagenesis*, 47, 362–390.

Cohen Hubal, E.A., Wetmore, B.A., Wambaugh, J.F., el-Masri, H., Sobus, J. R. & Bahadori, T. (2019) Advancing internal exposure and physiologically-based toxicokinetic modeling for 21st-century risk assessments. *Journal of Exposure Science & Environmental Epidemiology*, 29, 11–20.

Dertinger, S.D., Kraynak, A.R., Wheeldon, R.P., Bernacki, D.T., Bryce, S.M., Hall, N. et al. (2019) Predictions of genotoxic potential, mode of action, molecular targets, and potency via a tiered multiflow® assay data analysis strategy. *Environmental and Molecular Mutagenesis*, 60, 513–533.

EFSA Panel on Contaminants in the Food Chain (CONTAM). (2015) Scientific opinion on acrylamide in food. *EFSA Journal*, 13, 4104.

Environment and Climate Change Canada. (2014a) Screening assessment on etoposide.

Environment and Climate Change Canada. (2014b) Screening assessment on twenty-three substances on the domestic substances list used primarily as pharmaceuticals.

European Food Safety Authority (EFSA), Benford, D., Bignami, M., Chipman, J.K. & Ramos Bordajandi, L. (2022) Assessment of the genotoxicity of acrylamide. *EFSA Journal*, 20, e07293.

Government of Canada. (2021) 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.

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

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.

IARC. (2001) IARC monographs on the evaluation of the carcinogenic risk of chemicals to humans. *IARC*, 79, 1–763.

IARC. (1987) Overall evaluations of carcinogenicity: and updating of IARC monographs, vol. 1 to 42. IARC monographs on the evaluation of the carcinogenic risk of chemicals to humans. *IARC*, 7, 1–440.

IARC. (2000) Some antiviral and antineoplastic drugs, and other pharmaceutical agents World Health Organization.

IARC. (2010) The IARC monographs, vol: 100: a review and update on occupational carcinogens. Proceedings of the International Agency for Research on Cancer, Lyon, France, 23.

Kavlock, R.J., Bahadori, T., Barton-Maclaren, T.S., Gwinn, M.R., Rasenberg, M. & Thomas, R.S. (2018) Accelerating the pace of chemical risk assessment. *Chemical Research in Toxicology*, 31, 287–290.

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.

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.

Kirkland, D., Kasper, P., Martus, H.J., Müller, L., Van Benthem, J., Madia, F. et al. (2016) Updated recommended lists of genotoxic and non-genotoxic chemicals for assessment of the performance of new or improved genotoxicity tests. *Mutation Research, Genetic Toxicology and Environmental Mutagenesis*, 795, 7–30.

Kopp, B., Vignard, J., Mirey, G., Fessard, V., Zalko, D., Le Hgarat, L. et al. (2018) Genotoxicity and mutagenicity assessment of food contaminant mixtures present in the French diet. *Environmental and Molecular Mutagenesis*, 59, 742–754.

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.

Lea, I.A., Gong, H., Paleja, A., Rashid, A. & Fostel, J. (2017) CEBS: a comprehensive annotated database of toxicological data. *Nucleic Acids Research*, 45, D964–D971.

Luijten, M., Zwart, E.P., Dollé, M.E.T., De Pooter, M., Cox, J.A., White, P.A. et al. (2016) Evaluation of the LacZ reporter assay in cryopreserved primary hepatocytes for in vitro genotoxicity testing. *Environmental and Molecular Mutagenesis*, 57, 643–655.

Ly Pham, L., Watford, S., Pradeep, P., Martin, M.T., Thomas, R., Judson, R. et al. (2020) Variability in in vivo studies: defining the upper limit of performance for predictions of systemic effect levels. *Computational Toxicology*, 15, 1–100126.

Maertens, R.M., Long, A.S. & White, P.A. (2017) Performance of the in vitro transgene mutation assay in MutaMouse FE1 cells: evaluation of nine misleading ("false") positive chemicals. *Environmental and Molecular Mutagenesis*, 58, 582–591.

Mitch, W.A., Sharp, J.O., Trussell, R.R., Valentine, R.L., Alvarez-Cohen, L. & Sedlak, D.L. (2003) N-Nitrosodimethylamine (NDMA) as a drinking water contaminant: a review. *Environmental Engineering Science*, 20, 389–404.

Muehlbauer, P.A. & Schuler, M.J. (2005) Detection of numerical chromosomal aberrations by flow cytometry: a novel process for identifying aneugenic agents. *Mutation Research*, 585, 156–169.

National Toxicology Program. (2014) NTP technical report on the toxicology and carcinogenesis studies of glycidamide (CASRN 5694-00-8) in F344/N Nctr rats and B6C3F1/Nctr mice (drinking water studies).

Oliver, J., Meunier, J.R., Awogi, T., Elhajouji, A., Ouldelhkim, M.C., Bichet, N. et al. (2006) SFTG international collaborative study on in vitro micronucleus test V. using L5178Y cells. *Mutation Research*, 607, 125–152.

Olsen, G.W., Bodner, K.M., Stafford, B.A., Cartmill, J.B. & Gondek, M.R. (1995) Update of the mortality experience of employees with

occupational exposure to 1,2-dibromo-3-chloropropane (DBCP). *American Journal of Industrial Medicine*, 28, 399–410.

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.

Ring, C.L., Arnot, J.A., Bennett, D.H., Egeghy, P.P., Fantke, P., Huang, L. et al. (2019) Consensus modeling of median chemical intake for the U.S. population based on predictions of exposure pathways. *Environmental Science & Technology*, 53, 719–732.

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.

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*, 39, 583–604.

Shemansky, J.M., McDaniel, L.P., Klimas, C., Dertinger, S.D., Dobrovolsky, V.N., Kimoto, T. et al. (2019) Pig-a gene mutation database. *Environmental and Molecular Mutagenesis*, 60, 759–762.

Soeteman-Hernández, L.G., Fellows, M.D., Johnson, G.E. & Slob, W. (2015) Correlation of in vivo versus in vitro benchmark doses (BMDs) derived from micronucleus test data: a proof of concept study. *Toxicological Sciences*, 148, 355–367.

Soeteman-Hernández, L.G., Johnson, G.E. & Slob, W. (2016) Estimating the carcinogenic potency of chemicals from the in vivo micronucleus test. *Mutagenesis*, 31, 347–358.

Tosato, M.L., Terlizzese, M. & Dogliotti, E. (1987) Effects of buffer composition on water stability of alkylating agents. The example of N-ethyl-N-nitrosourea. *Mutation Research*, 179, 123–133.

Trudeau, J. (2021) Minister of Health Mandate Letter.

Wambaugh, J.F., Hughes, M.F., Ring, C.L., MacMillan, D.K., Ford, J., Fennell, T.R. et al. (2018) Evaluating In vitro-In vivo extrapolation of toxicokinetics. *Toxicological Sciences*, 163, 152–169.

Wambaugh, J.F., Wetmore, B.A., Pearce, R., Strope, C., Goldsmith, R., Sluka, J.P. et al. (2015) Toxicokinetic triage for environmental chemicals. *Toxicological Sciences*, 147, 55–67.

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.

Wetmore, B.A., Wambaugh, J.F., Allen, B., Ferguson, S.S., Sochaski, M.A., Setzer, R.W. et al. (2015) Incorporating high-throughput exposure predictions with dosimetry-adjusted In vitro bioactivity to inform chemical toxicity testing. *Toxicological Sciences*, 148, 121–136.

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.

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

White, P.A., Douglas, G.R., Gingerich, J., Parfett, C., Shwed, P., Seligy, V. et al. (2003) Development and characterization of a stable epithelial cell line from Muta™mouse lung. *Environmental and Molecular Mutagenesis*, 42, 166–184.

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.

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., Zeller, A., Pfuhler, S. & Johnson, G.E. (2019) Re: Gi Pfuhler S., Johnson G.E. 2018, In vivo positive mutagenicity of 1,4-dioxane and quantitative analysis of its mutagenicity and carcinogenicity in rats, archives of toxicology 92:3207-3221. *Archives of Toxicology*, 93, 211–212.

WHO. (2021) World Health Organization model list of essential medicines: 22nd list 2021 WHO.

Williams, A.J., Grulke, C.M., Edwards, J., McEachran, A.D., Mansouri, K., Baker, N.C. et al. (2017) The CompTox chemistry dashboard: a community data resource for environmental chemistry. *Journal of Cheminformatics*, 9, 61.

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

Zwart, E.P., Schaap, M.M., Van Den Dungen, M., Braakhuis, H.M., White, P.A., Van Steeg, H. et al. (2012) Proliferating primary hepatocytes from the pUR288 lacZ plasmid mouse are valuable tools for genotoxicity assessment in vitro. *Environmental and Molecular Mutagenesis*, 53, 1–8.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: 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*, 1–18. Available from: https://doi.org/10.1002/em.22521