Published in final edited form as:

Mamm Genome. 2018 February ; 29(1-2): 182–189. doi:10.1007/s00335-017-9731-6.

Advancing Chemical Risk Assessment Decision-Making with Population Variability Data: Challenges and Opportunities

Weihsueh A. Chiu¹ and Ivan Rusyn¹

¹Department of Veterinary Integrative Biosciences, College of Veterinary Medicine and Biomedical Sciences, Texas A&M University, College Station, TX 77843, USA

Abstract

Characterizing population variability, including identifying susceptible populations and quantifying their increased susceptibility, is an important aspect of chemical risk assessment, but one that is challenging with traditional experimental models and risk assessment methods. New models and methods to address population variability can be used to advance the human health assessments of chemicals in three key areas. First, with respect to hazard identification, evaluating toxicity using population-based *in vitro* and *in vivo* models can potentially reduce both false positive and false negative signals. Second, with respect to evaluating mechanisms of toxicity, enhanced ability to do genetic mapping using these models allows for the identification of key biological pathways and mechanisms that may be involved in toxicity and/or susceptibility. Third, with respect to dose-response assessment, population-based toxicity data can serve as a surrogate for human variability, and thus be used to quantitatively estimate the degree of human toxicokinetic/toxicodynamic variability and thereby increase confidence in setting health-protective exposure limits. A number of case studies have been published that demonstrate the potential opportunities for improving risk assessment and decision-making, and include studies using Collaborative Cross and Diversity Outbred mice, as well as populations of human cell lines from the 1000 Genomes project. Key challenges include the need to apply more sophisticated computational and statistical models analyzing population-based toxicity data, and the need to integrate these more complex analyses into risk assessments and decision-making.

Keywords

Genetics; toxicokinetics; toxicodynamics; susceptibility

Introduction

Chemical risk assessment is the process of identifying and characterizing the potential adverse consequences of chemical exposures. This process is typically divided into the four components of exposure assessment, hazard identification, dose-response assessment, and

Corresponding author: Weihsueh A. Chiu, PhD, Phone: 979-845-4106, wchiu@cvm.tamu.edu, Address: Department of Veterinary Integrative Biosciences, 4458 TAMU, Texas A&M University, College Station, TX 77843-4458, USA.

Conflict of Interest Statement:

On behalf of all authors, the corresponding author states that there is no conflict of interest.

risk characterization (NAS 1983). A critical component of environmental risk management decision-making is characterizing variability in how humans response to exogenous stressors (Zeise et al. 2013). Indeed, many environmental statutes, such as the Clean Water Act (1972), the Clean Air Act (1963), and the Federal Insecticide, Fungicide, and Rodenticide Act (1910), either implicitly or explicitly outline the need to protect not just the “typical,” but also sensitive individuals or populations.

Although great strides have been made in better characterizing exposure variability (NAS 2012), the ability to characterize human variability in hazard and dose-response is still quite limited. For instance, although epidemiologic data are highly informative as to identifying chemical hazards, they generally have limited power to examine variability and susceptibility across the population (NAS 2017; Zeise et al. 2013). Molecular epidemiology and genome-wide association studies (GWAS), while useful for generating or testing hypotheses as to specific genes or mechanisms that underlie susceptibility, are not directly translatable into meaningful quantitative estimates of population-wide variation. After epidemiology, experimental animal bioassays are the next most common type of data used in risk assessment. However, not only are there uncertainties as to extrapolating from animals to humans, but also the use of homogeneous, usually inbred, experimental animal populations makes it impossible to make inferences about human variation (Rusyn et al. 2010). Following the National Research Council’s 2007 Report “Toxicity Testing in the 21st Century” (NAS 2007), a major effort has ensued to use *in vitro*, human cell-based assays to assess hazard and dose-response. In such studies, the emphasis is usually placed on modeling the diversity of tissue and/or species representation, rather than inter-individual variability. Hence, with a few exceptions, existing data and models have been largely unable to address the need for risk assessment to characterize population variability in hazard and dose-response. Thus, it is questionable the extent to which traditional risk assessments are truly predictive (or protective) for a variable human population.

Here we provide an overview of the opportunities afforded by using mammalian models that are genetically diverse, genetically defined and reproducible, such as panels of inbred mouse strains, to address the challenge of characterizing human variability to inform environmental decision-making. Recently, Harrill and McAllister (2017) reviewed comprehensively the available rodent population model systems, so the focus here will be on how the mouse *in vivo* and human *in vitro* resources can directly impact risk assessment. Approaches that can enhance hazard identification, including how they may help identify mechanisms of toxicity and susceptibility, will be discussed first. Then, approaches to improve dose-response assessment are discussed. We conclude that genetically-diverse, population-based new experimental systems have great potential to inform risk assessment and decision-making, but require additional integration with mathematical and statistical models that better account for population variability.

Enhancing Hazard Identification

Hazard identification is “the process of determining whether exposure to an agent can cause an increase in the incidence of a health condition ... [including] characterizing the nature and strength of the evidence of causation” (NAS 1983). This process needs to take into

account the fact that in the population, there are individuals that are more or less susceptible to toxicity. In the extreme, there may be subpopulations that are uniquely sensitive, so that the risk is very small or negligible in the rest of the population. For instance, individuals with xeroderma pigmentosum are exquisitely sensitive to even the briefest exposure to sunlight (Hanawalt 1996). Even in less extreme cases, not accounting for susceptibility differences can result in signals being lost in the sea of unsusceptible individuals.

With respect to experimental systems, one of the key debates over the decades has been the human relevance of both positive and negative results observed in rodents (Beyer et al. 2011). However, this debate has largely occurred in the context of studies using homogeneous, single strains of rodents (Maronpot et al. 2016). Thus, one might hypothesize that for any particular exposure and end point combination, there are likely to be different rodent strains that are better or worse models for humans (Figure 1a), so that the response from any single strain (whether positive or negative) may fall outside the range of human responses. By using a population of strains, the chance of actually “hitting the target” of human relevance should increase, thereby reducing false positives and false negatives.

This hypothesis that using multiple strains increases human relevance has been affirmed a number of times through studies in population-based experimental model systems. For instance, Harrill et al. (Harrill et al. 2009a; Harrill et al. 2009b) administered acetaminophen to genetically diverse population of 36 inbred strains of mice and compared to results in humans. Even for a classic hepatotoxicant, there are some mouse strains that were essentially unresponsive even at a high dose (Figure 1b). Thus, if a new chemical is being tested, one might simply miss the hazard if one happened to be using a resistant strain. The only way to routinely avoid this pitfall is by testing in a population. In another example, it was long thought that the laboratory mouse was a poor model for Ebola because it did not display the classic hemorrhagic syndrome. However, it was demonstrated by Rasmussen et al (2014) that use of a genetically diverse panel of 47 mouse strains actually recapitulated not only the hemorrhagic syndrome amongst a subset of the population, but also showed a diversity of phenotypes that mirrors that in human disease outbreaks (Figure 1c).

These examples open the question of how many other cases there are that data from experimental rodent models may have been erroneously dismissed as “non-relevant to humans” due to interspecies differences, and whether the data from studies in rodents can in essence be “rescued” by using a genetically diverse population. There are several resources for such experiments, particularly in mice, including the Hybrid Mouse Diversity Panel (HMDP) (Lusis et al. 2016), the Collaborative Cross (CC) (Threadgill and Churchill 2012), and the Diversity Outbred (DO) (Churchill et al. 2012). However, this approach of testing populations is not without some challenges. One is that traditional pairwise or even trend-based statistical testing is not appropriate because the study design is more similar to epidemiology or clinical trials, where everyone is an individual. Thus, the appropriate data analysis and statistical modeling approaches need to be applied, such as random effects models or hierarchical Bayesian approaches (Aylor et al. 2011; Maurizio et al. 2017; Oreper et al. 2017). A related challenge is study design that addresses the tradeoff between more replicates of a single strain versus more strains (Kaeppler 1997). However, these are not

insurmountable challenges, and additional experience with these models will undoubtedly help to overcome them.

Investigating Mechanisms of Toxicity and Susceptibility

Toxicity is in general a product of gene-environment interactions, rather than purely related to the toxicant alone. Thus, population-based approaches that incorporate genetic diversity provide a vehicle for deeper investigations into the mechanistic basis for both hazard and susceptibility.

For example, in the acetaminophen studies previously discussed, the mouse population was treated like an epidemiologic study, specifically a GWAS (Harrill et al. 2009b). The gene *Cd44* was identified as a candidate modulator of susceptibility. The investigators then looked into human epidemiologic data to test the hypothesis – and confirmed that variants of this gene showed differential susceptibility to acetaminophen toxicity. Moreover, this gene is related to the immune response, and it turns out that the key determinant of adverse phenotype (not just whether some liver cells die) is the CD44 status of the macrophages. For some CD44 states, the result is recovery and repair, whereas in other states, the result is apoptosis and inflammation (Negi et al. 2012). In sum, the population mouse study ultimately led to important mechanistic insights that were subsequently confirmed in humans.

More recent studies used mouse CC model to investigate the role of genetic variability in the toxicokinetics and toxicodynamics of environmental chemicals and drugs. A study by Cichocki et al. (2017) quantitatively examined the relationship between perchloroethylene (PERC) toxicokinetics and toxicodynamics at the population level, using 45 CC strains to test whether individuals with increased oxidative metabolism are more sensitive to hepatotoxicity following PERC exposure. Significant variability among strains was observed in toxicokinetics of PERC in every tissue examined. This study demonstrated a complex and highly variable relationship between PERC and oxidative metabolite toxicokinetics and toxicodynamics at the population level by using the CC mouse population model. A subsequent study (Venkatratnam et al. 2017) also used the CC model, evaluating inter-strain variability among 50 strains in trichloroethylene (TCE) metabolism and identifying genetic determinants that are associated with TCE metabolism and effects. The variability in oxidative metabolite levels among strains did not correlate with expression or activity of a number of enzymes known to be involved in TCE oxidation; however, peroxisome proliferator-activated receptor alpha (PPAR α)-responsive genes were found to be associated with strain-specific differences in TCE metabolism. A study of potential genetic determinants of idiosyncratic drug-induced liver injury associated with administration of a drug tolcapone for the treatment of autosomal dominant polycystic kidney disease used 45 CC strains (Mosedale et al. 2017). The authors found that administration of tolcapone in mice resulted in liver oxidative stress, mitochondrial dysfunction, and innate immune response in all strains, but that bile acid homeostasis pathway was most associated with susceptibility to the liver response. Overall, these studies suggest that the CC mouse population is a valuable tool to quantitatively evaluate inter-individual variability in

chemical metabolism and to identify genes and pathways that may underpin population differences and thus be used as biomarkers of susceptibility.

Population-based approaches are not limited to *in vivo* animal studies. Abdo et al. (2015b) used 1,086 lymphoblast cell lines drawn from the 1000 Genomes Project, representing nine populations from five continents, to assess variation in cytotoxic response to 179 chemicals. The data from this study allowed for derivation of chemical-specific estimates of population variability in cytotoxicity (discussed further in the next section on improving dose-response assessment). In addition, important molecular insights into the molecular underpinnings of the variability in toxicity were gained as genetic mapping suggested important roles for variation in membrane and transmembrane genes, with a number of chemicals showing association with SNP rs13120371 in the solute carrier *SLC7A11*, previously implicated in chemo-resistance.

Thus, using these population-based models provides several great opportunities for mechanism-based toxicology. In particular, they can provide complementary data to human epidemiology studies in identifying genetic susceptibilities, providing specific hypothesis that can then be tested in human populations.

Improving Dose-Response Assessment

Although hazard identification and mechanistic investigations are important for risk management decision-making, they are usually not sufficient. Making decisions about whether and how to limit chemical exposures almost always requires assessing the quantitative dose-response relationship and comparing with human exposure estimates. It is important to clearly distinguish between a dose-response for the overall human population and each individual's dose-response curve (NAS 2009) (Figure 2a). Specifically, individuals' dose-response curves are likely to be different, as people differ in their sensitivity to toxicity, and it is important to quantify these differences in order to ensure that public health is protected. Moreover, only after characterizing this variation can a population-level dose-response curve be predicted.

The most common way in which dose-response is characterized is in a "toxicity value," which is usually a single number (but sometimes with a confidence interval) that summarizes the dose at which toxicity is observed or predicted, or at which toxicity is predicted to be absent. As shown in Table 1, for most common toxicity values, population variability is either not addressed, or addressed in a very rudimentary manner. For instance, the Reference Dose is often derived by dividing a NOAEL or BMDL by ten, twice: once for interspecies differences and once for intraspecies differences due to human variability. This assumes that the single strain dose response data is representative of the "average" or "median" individual, and that these default "uncertainty factors" of 10 are adequate to ensure safety across a heterogeneous human population.

As is the case for Hazard Identification, the usual laboratory strains may not be relevant or representative of humans, so using a genetically diverse population will provide better empirical estimates of risk. Additionally, population-based data on toxicokinetic and

toxicodynamic variability can be used to replace default uncertainty factors, and ultimately better characterize the fraction of the population that may be affected by exposure to a toxicant.

One excellent example is from work by French et al (2015), who exposed 600 mice from the DO population to 0, 1, 10, and 100 ppm benzene and then measured micronuclei in blood and bone marrow. Even at baseline, there was a great deal of variability in micronuclei. But there is nonetheless a response at the population level at the lowest concentration of 1 ppm, with an even lower value for the lower confidence limit on benchmark concentration for a 10% increase in micronuclei (Figure 2b). Interestingly, standard B6C3F1 mouse (a cross between female C57BL/6 and male C3H mice, used in much of toxicity testing) is actually less sensitive than the “average” mouse in the DO population. Put another way, the B6C3F1 is on the “resistant” tail of the population distribution for this endpoint, again demonstrating how using a single strain can lead to inaccuracies in risk assessment.

Another idea for which there is proof of principle is that these mouse population models can be used as surrogates for human variability in toxicokinetics. Chiu et al (2014) examined the toxicokinetics of TCE in a panel of 17 different mouse strains that included the B6C3F1. The toxicity of TCE is mediated by metabolism, so we measured two oxidation and two GSH conjugation metabolites (Bradford et al. 2011), and then used a Bayesian population PBPK model to analyze the data. There was substantial variability, particularly in the GSH conjugation metabolites, and it was evident that the B6C3F1 mouse was not typical, and tended to be on one or the other tail of the distribution.

An even more interesting observation from this study came from comparing to a previous Bayesian population PBPK analysis of available human in vivo data, thereby directly examining whether the variability across mouse strains can be used as a surrogate for human variability. In each case, variability was defined quantitatively as ratio between the 95th percentile individual and the median individual. The mouse and human results turned out to be very consistent across multiple dose metrics. For instance, for total oxidation, there was very little variability in both species; for the oxidative metabolite trichloroacetic acid, there was about a 2-fold variation in both species; and for GSH conjugation, both species had about a 7-fold variation.

With respect to toxicodynamics, the proof of principle was demonstrated in the previously mentioned study by Abdo et al. (2015b), which used a panel of 1,086 human lymphoblastoid cell lines to provide quantitative estimates for chemical-specific toxicodynamic variability across the agents tested (Figure 2c), as well as a follow-up study of mixtures (Abdo et al. 2015a). For the median chemical, cytotoxic response in the 1% most “sensitive” individual occurred at concentrations within a factor of $10^{1/2}$ (i.e., approximately the default value of 3) of that in the median individual; however, for many compounds, this factor was >10 , highlighting the need for chemical-specific data on population variability. Like in the case of TCE, it was possible to compare with estimates of toxicodynamic variability from human in vivo studies. Even though the chemicals in each analysis were different, the distributions across chemicals were highly consistent, suggesting that population-based in vitro systems may be able to be used as a surrogate for human toxicodynamic variability. In a follow-up

study, Chiu et al. (2017) demonstrated how these data could be used to develop a “tiered,” Bayesian approach to characterizing toxicodynamic variability, starting with a “default” based on the existing distribution across chemicals, and then proceeding from pilot experiments of ~20 cell lines up to larger experiments of 50 to 100 cell lines if needed to refine variability estimates.

These new data-driven, quantitative, population-based approaches, however, as somewhat challenging to fit into the traditional risk assessment paradigm, and beg for a probabilistic approach to integrating these data. Fortunately, a comprehensive probabilistic framework for dose-response has been developed by WHO (2014), summarized in Chiu and Slob (2015). The idea behind a probabilistic risk assessment framework is to replace the usual point estimates used to derive toxicity values (NOAEL, BMDL, uncertainty factors, etc.) with probability distributions that separately reflect uncertainty and population variability. The underlying concept reflects the distinction in Figure 2a between population and individual dose-response, so that each individual has their own personal dose-response curve. As population exposure increases, some people have a milder effect and some people have a more severe effect, with the result at the population level being the more and more people suffer more and more severe effects. The goal of a protective toxicity value, then, is to protect a large fraction of the population from even small effects. For instance, one could replace the usual RfD, that is derived by dividing a NOAEL by 100, with a quantitatively estimated dose level that protects 99% of the population from a 5% decline in hematocrit with 95% confidence.

In sum, there are numerous opportunities to improve dose-response assessment using population-based models. Having population-level dose-response data helps protect against bias due to individual experimental models being overly sensitive or resistant. They also can serve as surrogates for estimating human variability, with several examples demonstrating good concordance between the variability estimated in experimental systems and the extent of human variability. Some challenges, of course, remain in the integration and communication of these new approaches, such as the need to transition from traditional point-estimate-based toxicity values to a probabilistic expression of dose, severity, incidence, and confidence; and the need for researchers, risk assessors, and risk managers to engage and communicate in order to be able to integrate these data streams together.

Conclusions

Addressing genetic and population variability is an essential component of risk assessment, and new population-based experimental and computational tools available that provide are many opportunities for advancement. For hazard identification, using population-based approaches means that one is more likely to overlap with the range of human responses. The technical challenges in experimental design and data analysis are likely surmountable with experience. For mechanistic research, there are opportunities for identifying the genetic basis for susceptibility using experimental methods, rather than relying on human epidemiology. In fact, by using experimental methods to generate hypotheses, it may be possible to better focus epidemiologic research on testing or confirming those hypothesis. There are also a number of statistical power advantages in the genetic mapping of the

experimental resources that make it a more efficient method. As with any genetics-based investigation, polygenic susceptibilities are more challenging, and these models do not address any of the non-genetic sources of variability that may influence an individual's susceptibility. For dose-response assessment, there are numerous opportunities to better characterize both population and individual dose-response, including toxicokinetic variability, toxicodynamic variability, as well as the overall distribution of dose-response curves across the population. The challenges here are more in making a paradigm shift in terms of how we conduct dose-response assessment away from fixed point estimates to probabilistic analysis, and the richer characterization of individual risk, population incidence, and statistical confidence. Finally, in order to make an impact on risk management decision-making, all these scientific and technical advances need to be coupled with enhanced engagement and communication across the various communities of scientific researchers, risk assessors, and risk managers. These data are not part of "traditional" risk assessments and educating the stakeholders about the value of this additional information is an opportunity for both science communication and education.

Acknowledgments

This publication was made possible, in part, by NIH grant P42 ES027704. Its contents are solely the responsibility of the grantee and do not necessarily represent the official views of the NIH. Further, the NIH does not endorse the purchase of any commercial products or services mentioned in the publication.

References

- Federal Insecticide, Fungicide, and Rodenticide Act, as amended 2012. 1910 In U.S.C. 7 section 136 et seq.
- An Act to improve, strengthen, and accelerate programs for the prevention and abatement of air pollution (Clean Air Act), as amended 1990. 1963 In U.S.C. 42 section 7401 et seq.
- Federal Water Pollution Control Act (Clean Water Act), as amended 2002. 1972 In U.S.C. 33 section 1251 et seq.
- Abdo N, Wetmore BA, Chappell GA, Shea D, Wright FA, Rusyn I. In vitro screening for population variability in toxicity of pesticide-containing mixtures. *Environ Int.* 2015a; 85:147–155. [PubMed: 26386728]
- Abdo N, Xia M, Brown CC, Kosyk O, Huang R, Sakamuru S, Zhou YH, Jack JR, Gallins P, Xia K, Li Y, Chiu WA, Motsinger-Reif AA, Austin CP, Tice RR, Rusyn I, Wright FA. Population-based in vitro hazard and concentration-response assessment of chemicals: the 1000 genomes high-throughput screening study. *Environ Health Perspect.* 2015b; 123:458–466. [PubMed: 25622337]
- ATSDR. Minimal Risk Levels (MRLs). Atlanta, GA: U.S. Department of Health and Human Services, Agency for Toxic Substances and Disease Registry; 2013.
- Aylor DL, Valdar W, Foulds-Mathes W, Buus RJ, Verdugo RA, Baric RS, Ferris MT, Frelinger JA, Heise M, Frieman MB, Gralinski LE, Bell TA, Didion JD, Hua K, Nehrenberg DL, Powell CL, Steigerwalt J, Xie Y, Kelada SN, Collins FS, Yang IV, Schwartz DA, Branstetter LA, Chesler EJ, Miller DR, Spence J, Liu EY, McMillan L, Sarkar A, Wang J, Wang W, Zhang Q, Broman KW, Korstanje R, Durrant C, Mott R, Iraqi FA, Pomp D, Threadgill D, Pardo-Manuel DV, Churchill GA. Genetic analysis of complex traits in the emerging collaborative cross. *Genome Res.* 2011; 21:1213–1222. [PubMed: 21406540]
- Beyer LA, Beck BD, Lewandowski TA. Historical perspective on the use of animal bioassays to predict carcinogenicity: evolution in design and recognition of utility. *Crit Rev Toxicol.* 2011; 41:321–338. [PubMed: 21438739]
- Bradford BU, Lock EF, Kosyk O, Kim S, Uehara T, Harbourt D, DeSimone M, Threadgill DW, Tryndyak V, Pogribny IP, Bleye L, Koop DR, Rusyn I. Interstrain differences in the liver effects of

trichloroethylene in a multistrain panel of inbred mice. *Toxicol Sci.* 2011; 120:206–217. [PubMed: 21135412]

- Chiu WA, Campbell JL, Clewell HJ, Zhou YH, Wright FA, Guyton KZ, Rusyn I. Physiologically-Based Pharmacokinetic (PBPK) Modeling of Inter-strain Variability in Trichloroethylene Metabolism in the Mouse. *Environ Health Perspect.* 2014; 122:456–463. [PubMed: 24518055]
- Chiu, WA., Guyton, KZ., Martin, MT., Reif, DM., Rusyn, I. Use of high-throughput in vitro toxicity screening data in cancer hazard evaluations by IARC Monograph Working Groups. *ALTEX*; 2017.
- Chiu WA, Slob W. A Unified Probabilistic Framework for Dose-Response Assessment of Human Health Effects. *Environ Health Perspect.* 2015; 123:1241–1254. [PubMed: 26006063]
- Churchill GA, Gatti DM, Munger SC, Svenson KL. The Diversity Outbred mouse population. *Mamm Genome.* 2012; 23:713–718. [PubMed: 22892839]
- Cichocki JA, Furuya S, Venkatratnam A, McDonald TJ, Knap AH, Wade T, Sweet S, Chiu WA, Threadgill DW, Rusyn I. Characterization of Variability in Toxicokinetics and Toxicodynamics of Tetrachloroethylene Using the Collaborative Cross Mouse Population. *Environ Health Perspect.* 2017; 125:057006. [PubMed: 28572074]
- French JE, Gatti DM, Morgan DL, Kissling GE, Shockley KR, Knudsen GA, Shepard KG, Price HC, King D, Witt KL, Pedersen LC, Munger SC, Svenson KL, Churchill GA. Diversity Outbred Mice Identify Population-Based Exposure Thresholds and Genetic Factors that Influence Benzene-Induced Genotoxicity. *Environ Health Perspect.* 2015; 123:237–245. [PubMed: 25376053]
- Hanawalt PC. Role of transcription-coupled DNA repair in susceptibility to environmental carcinogenesis. *Environ Health Perspect.* 1996; 104(Suppl 3):547–551. [PubMed: 8781381]
- Harrill AH, McAllister KA. New Rodent Population Models May Inform Human Health Risk Assessment and Identification of Genetic Susceptibility to Environmental Exposures. *Environ Health Perspect.* 2017; 125:086002. [PubMed: 28886592]
- Harrill AH, Ross PK, Gatti DM, Threadgill DW, Rusyn I. Population-based discovery of toxicogenomics biomarkers for hepatotoxicity using a laboratory strain diversity panel. *Toxicol Sci.* 2009a; 110:235–243. [PubMed: 19420014]
- Harrill AH, Watkins PB, Su S, Ross PK, Harbourt DE, Stylianou IM, Boorman GA, Russo MW, Sackler RS, Harris SC, Smith PC, Tennant R, Bogue M, Paigen K, Harris C, Contractor T, Wiltshire T, Rusyn I, Threadgill DW. Mouse population-guided resequencing reveals that variants in CD44 contribute to acetaminophen-induced liver injury in humans. *Genome Res.* 2009b; 19:1507–1515. [PubMed: 19416960]
- Kaepler SM. Quantitative trait locus mapping using sets of near-isogenic lines: relative power comparisons and technical considerations. *Theor Appl Genet.* 1997; 95:384–392.
- Lusis AJ, Seldin MM, Allayee H, Bennett BJ, Civelek M, Davis RC, Eskin E, Farber CR, Hui S, Mehrabian M, Norheim F, Pan C, Parks B, Rau CD, Smith DJ, Vallim T, Wang Y, Wang J. The Hybrid Mouse Diversity Panel: a resource for systems genetics analyses of metabolic and cardiovascular traits. *J Lipid Res.* 2016; 57:925–942. [PubMed: 27099397]
- Maronpot RR, Nyska A, Foreman JE, Ramot Y. The legacy of the F344 rat as a cancer bioassay model (a retrospective summary of three common F344 rat neoplasms). *Crit Rev Toxicol.* 2016; 46:641–675. [PubMed: 27278595]
- Maurizio PL, Ferris MT, Keele GR, Miller DR, Shaw GD, Whitmore AC, West A, Morrison CR, Noll KE, Plante KS, Cockrell AS, Threadgill DW, Pardo-Manuel de Villena F, Baric RS, Heise MT, Valdar W. Bayesian Diallel Analysis Reveals Mx1-Dependent and Mx1-Independent Effects on Response to Influenza A Virus in Mice. *G3 (Bethesda).* 2017
- Mosedale M, Kim Y, Brock WJ, Roth SE, Wiltshire T, Eaddy JS, Keele GR, Corty RW, Xie Y, Valdar W, Watkins PB. Candidate Risk Factors and Mechanisms for Tolvaptan-Induced Liver Injury Are Identified Using a Collaborative Cross Approach. *Toxicol Sci.* 2017; 156:438–454. [PubMed: 28115652]
- NAS. Risk assessment in the federal government: Managing the process. Washington, D.C.: National Academies Press; 1983.
- NAS. Toxicity testing in the 21st century: A vision and a strategy. Washington, DC: The National Academies Press; 2007.

- NAS. Science and Decisions: Advancing Risk Assessment. Washington, DC: National Academies Press; 2009.
- NAS. Exposure Science in the 21st Century: A Vision and a Strategy. Washington, DC: 2012.
- NAS. Using 21st Century Science to Improve Risk-Related Evaluations. Washington, DC: 2017.
- Negi LM, Talegaonkar S, Jaggi M, Ahmad FJ, Iqbal Z, Khar RK. Role of CD44 in tumour progression and strategies for targeting. *J Drug Target*. 2012; 20:561–573. [PubMed: 22758394]
- Oreper D, Cai Y, Tarantino LM, de Villena FP, Valdar W. Inbred Strain Variant Database (ISVdb): A Repository for Probabilistically Informed Sequence Differences Among the Collaborative Cross Strains and Their Founders. *G3 (Bethesda)*. 2017; 7:1623–1630. [PubMed: 28592645]
- Rasmussen AL, Okumura A, Ferris MT, Green R, Feldmann F, Kelly SM, Scott DP, Safronetz D, Haddock E, LaCasse R, Thomas MJ, Sova P, Carter VS, Weiss JM, Miller DR, Shaw GD, Korth MJ, Heise MT, Baric RS, de Villena FP, Feldmann H, Katze MG. Host genetic diversity enables Ebola hemorrhagic fever pathogenesis and resistance. *Science*. 2014; 346:987–991. [PubMed: 25359852]
- Rusyn I, Gatti DM, Wiltshire T, Kleeberger SR, Threadgill DW. Toxicogenetics: population-based testing of drug and chemical safety in mouse models. *Pharmacogenomics*. 2010; 11:1127–1136. [PubMed: 20704464]
- Threadgill DW, Churchill GA. Ten years of the Collaborative Cross. *Genetics*. 2012; 190:291–294. [PubMed: 22345604]
- U.S. EPA. Risk Assessment Guidance for Superfund Volume I Human Health Evaluation Manual (Part A). Washington, DC: U.S. Environmental Protection Agency; 1989. EPA/540/1/90/002
- U.S. EPA. A Review of the Reference Dose and Reference Concentration Processes. Washington, DC: U.S. Environmental Protection Agency; 2002. EPA/630/P-02/002F
- U.S. EPA. Guidelines for Carcinogen Risk Assessment. Washington, DC: U.S. Environmental Protection Agency; 2005. EPA/630/P-03/001F
- U.S. EPA. Benchmark Dose Technical Guidance. Washington, DC: U.S. Environmental Protection Agency; 2012. EPA/100/R-12/001
- Venkatratnam A, Furuya S, Kosyk O, Gold A, Bodnar W, Konganti K, Threadgill DW, Gillespie KM, Aylor DL, Wright FA, Chiu WA, Rusyn I. Collaborative Cross Mouse Population Enables Refinements to Characterization of the Variability in Toxicokinetics of Trichloroethylene and Provides Genetic Evidence for the Role of PPAR Pathway in Its Oxidative Metabolism. *Toxicol Sci*. 2017; 158:48–62. [PubMed: 28369613]
- Wetmore BA. Quantitative in vitro-to-in vivo extrapolation in a high-throughput environment. *Toxicology*. 2015; 332:94–101. [PubMed: 24907440]
- WHO/IPCS. Guidance Document on Evaluating and Expressing Uncertainty in Hazard Characterization. Geneva: World Health Organization International Program on Chemical Safety; 2014.
- Zeise L, Bois FY, Chiu WA, Hattis D, Rusyn I, Guyton KZ. Addressing human variability in next-generation human health risk assessments of environmental chemicals. *Environ Health Perspect*. 2013; 121:23–31. [PubMed: 23086705]

Fig. 1. Illustration of how population models can enhance hazard identification

Panel a: Mouse populations may contain of mix of strains that are “good” or “poor” models of humans, in that they result in responses that are “inside” or “outside” the range observed in humans. Using a single strain, there is a greater likelihood of “missing” human relevance as compared to using a population. **Panel b:** Comparison of liver ALT responses in different strains of mice (open and closed circles) with those in different human individuals (squares), based on data reported in Harrill et al. (2009). Some strains (marked with “X”) are essentially non-responsive, even at 5-fold higher dose than the dose in humans. Even though

some humans are also non-response, using one of the non-responsive strains would result in missing the hazard. **Panel c:** Comparison of mouse and human population distribution of phenotypic outcomes after Ebola infections (HF = hemorrhagic fever), based on data reported in Rasmussen et al. (2014) [mouse] and McElroy et al. (2014) [human]. Even though some mouse strains are resistant, this is also the case for humans. Moreover, the population distribution is similar across phenotypes.

Fig. 2. Illustration of how population models can enhance dose-response assessment
Panel a: Populations contain a mix of individual dose-response curves, which when aggregated, can constitute the population-level dose-response relationship, adapted from the National Academies report *Science and Decisions* (NAS 2009). **Panel b:** Comparison of individual-level concentration-response data and the modeled population-level dose-response curve for DO mice exposed to benzene, based on data reported in French et al. (2015). The dose at which a 10% increase in bone marrow micronuclei occurs at the population level is lower than the dose at which this occurs in B6C3F1 mice, suggesting

using a single strain can be misleading. **Panel c:** Comparison of individual-level concentration-response data and the modeled population-level dose-response curve for cytotoxicity responses across 1086 lymphoblastoid cell lines exposed to zinc pyrithione, based on data reported in Abdo et al. (2015). The population concentration-response is based on fraction of cell lines at each concentration for which a 10% increase in cytotoxic response occurs (i.e., where the individual curves cross the dashed horizontal line in the left panel), or equivalently, the cumulative distribution of EC_{10} values.

Table 1

Common toxicity values used in risk assessment for dose-response assessment

Toxicity Values ^a	Definition	Population Variability Addressed?
<i>NOAEL</i> : No Observed Adverse Effect Level	The highest exposure level at which there are no biologically significant increases in the frequency or severity of adverse effect between the exposed population and its appropriate control.	No.
<i>BMD(L)</i> : Benchmark Dose (Lower confidence limit)	A dose of a substance that when ingested produces a predetermined change ("benchmark response") in the response rate of an adverse effect relative to the background response rate of this effect.	No.
<i>RfD</i> : Reference Dose	An estimate of the dose of a substance (with uncertainty spanning perhaps an order of magnitude) to which a human population can be exposed (including sensitive subgroups) that is likely to be without an appreciable risk of deleterious effects during a lifetime.	Uncertainty factor (10-fold) for human variability.
<i>MRL</i> : Minimal Risk Level	An estimate of the daily human exposure to a hazardous substance that is likely to be without appreciable risk of adverse non-cancer health effects over a specified duration of exposure.	Uncertainty factor (10-fold) for human variability.
<i>OSF</i> : Oral Slope Factor	An upper-bound estimate of risk per increment of oral dose that can be used to estimate risk probabilities for different exposure levels.	Increased risk at early life stages for mutagenic compounds.
<i>OED(L)</i> : Oral Equivalent Dose (Lower confidence limit)	Daily oral dose necessary to produce steady-state <i>in vivo</i> blood concentrations equivalent to the AC ₅₀ (concentration at 50% of maximum activity) or LEC (lowest effective concentration) values in the <i>in vitro</i> assays.	Modeled toxicokinetic variability.
<i>HD_M¹</i> : "Target Human Dose" for a specific Magnitude and Incidence	The human dose at which a fraction I of the population shows an effect of magnitude (or severity) M or greater (for the critical effect considered).	Probabilistic factor separately addressing variability and uncertainty.

Sources: (ATSDR 2013; U.S. EPA 1989, 2002., 2005., 2012; Wetmore 2015; WHO/IPCS 2014)

^a A numerical expression of the dose-response relationship that, when combined with exposure, gives information to characterize risk.