Mammalian Genome (2018) 29:190–204
https://doi.org/10.1007/s00335-018-9738-7

Leveraging human genetic and adverse outcome pathway (AOP) data to inform susceptibility in human health risk assessment

Holly M. Mortensen[1] · John Chamberlin[2] · Bonnie Joubert[3] · Michelle Angrish[4] · Nisha Sipes[5] · Janice S. Lee[4] · Susan Y. Euling[6,7]

Received: 4 October 2017 / Accepted: 31 January 2018 / Published online: 23 February 2018

Abstract

Estimation of susceptibility differences in human health risk assessment (HHRA) has been challenged by a lack of available susceptibility and variability data after exposure to a specific environmental chemical or pharmaceutical. With the increasingly large number of available data sources that contain polymorphism and other genetic data, human genetic variability that informs susceptibility can be better incorporated into HHRA. A recent policy, the 2016 The Frank R. Lautenberg Chemical Safety for the twenty-first Century Act, requires the US Environmental Protection Agency to evaluate new and existing toxic chemicals with explicit consideration of susceptible populations of all types (life stage, exposure, genetic, etc.). We propose using the adverse outcome pathway (AOP) construct to organize, identify, and characterize human genetic susceptibility in HHRA. We explore how publicly available human genetic datasets can be used to gain mechanistic understanding of molecular events and characterize human susceptibility for an adverse outcome. We present a computational method that implements publicly available human genetic data to prioritize AOPs with potential for human genetic variability. We describe the application of this approach across multiple described AOPs for health outcomes of interest, and by focusing on a single molecular initiating event. This contributes to a long-term goal to improve estimates of human susceptibility for use in HHRA for single and multiple chemicals.

Introduction

Genetic susceptibility and variability in response to environmental chemicals across human populations is an active area of investigation for human health risk assessment (HHRA) (NRC 2010; Krewski et al. 2014). The National Research Council (NRC) defined susceptibility as "the capacity to be affected" (NRC 2009) and stated that "variability in human susceptibility has not received sufficient or consistent attention in many EPA health risk assessments..." in their *Science and Decisions: Advancing Risk Assessment* report (NRC 2009). Four additional NRC reports include the need for incorporating human interindividual variability data in risk assessment (NRC 2009, 2011a, 2014a, 2014b, 2016). Two of these reports (NRC 2009, NRC) encourage the EPA to integrate information from twenty-first century methods. The NRC also makes recommendation on how emerging data streams (e.g., big data and toxicogenomic) can be integrated and used to improve risk-related evaluations (NRC 2017). Additionally, two EPA policies require accounting for *life stage* susceptibility for environmental chemicals, including the

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00335-018-9738-7) contains supplementary material, which is available to authorized users.

✉ Holly M. Mortensen
mortensen.holly@epa.gov

[1] NHEERL, US Environmental Protection Agency, Research Triangle Park, NC 27709, USA

[2] Oak Ridge Associated Universities, Research Triangle Park, NC 27709, USA

[3] National Institute of Environmental Health Sciences, Research Triangle Park, NC 27709, USA

[4] NCEA, US Environmental Protection Agency, Research Triangle Park, NC, USA

[5] NTP, National Institute of Environmental Health Sciences, Research Triangle Park, NC 27709, USA

[6] NCEA, US Environmental Protection Agency, Washington, DC 20460, USA

[7] Present Address: Office of Children's Health Protection, US Environmental Protection, Washington, DC 20460, USA

Food Quality Protection Act (FQPA 1996) and the Safe Drinking Water Act amendments (SDWA 1996). Most recently, the Frank R. Lautenberg Chemical Safety for the twenty-first Century Act (Public Law 114-182 2016), also referred to as the amended Toxic Substances Control Act (TSCA), modernizes and accelerates the pace of US EPA's toxic chemical evaluation process and, includes a requirement for consideration of all types of susceptibility. The law defines a 'potentially exposed or susceptible subpopulation' as "…a group of individuals within the general population identified by the Administrator who, due to either greater susceptibility or greater exposure, may be at greater risk than the general population of adverse health effects from exposure to a chemical substance or mixture, such as infants, children, pregnant women, workers, or the elderly'' (Public Law 114-182 2016). The law also requires prioritization of potentially high-risk toxic chemicals, in part based on "unreasonable risk to a potentially exposed or susceptible subpopulation" (Public Law 114-182 2016).

Current EPA approaches to incorporating genetic susceptibility information into HHRA include an assessment template that prompts the description of various susceptibility factors, including genetic factors. For example, genetic factors are considered in the "At Risk" Chapter of the Integrative Science Assessments (ISA) that support national ambient air quality standards (NAAQS) (Sacks et al. 2010) and a weight of evidence approach is used to determine potential modifying factors in the ISAs (Vinikoor-Imler et al. 2014). IRIS toxicological reviews, which develop noncancer oral reference doses (RfDs) or inhalation reference concentrations (RfCs), and unit risk estimates for cancer potency by each exposure route, include a susceptibility section that synthesizes the available information qualitatively. In rare cases such as the US EPA IRIS cancer assessment for benzene, human polymorphism data leading to differences in response to benzene, specifically single nucleotide polymorphism (SNP) data in benzene metabolic enzymes (CYP2E1, NADPH-dependent quinone oxidoreductase, MPO, GSH transferase, etc.), were included (US EPA 2002). However, comprehensive genetic data addressing specific chemical effects are often lacking and further, methods for quantitatively utilizing genetic information are limited. For this reason, one of the challenges for HHRA is quantifying human variability and defining the full range of response values that define a susceptible population. In EPA noncancer assessments, the approach for deriving a reference response value is to identify the point of departure and then, apply a set of default adjustment or uncertainty factors to account for missing data. For the case of intraspecies (interhuman) variability, the intraspecies [human uncertainty factor (UFH)] default value of 10X [comprised of 3X for toxicokinetics (TK) and 3X for toxicodynamic (TD) differences] is applied in the absence of data on human variability.

The approach presented herein aims to improve upon these existing measures through the identification and use of publically available genetic data, a first step in the prediction of susceptibility. The use of publicly available human genetic data eliminates the need for considering uncertainty when extrapolating from other organisms to humans. One approach for integrating mechanistic and polymorphism data to characterize human genetic susceptibility to chemical exposure has been described by Mortensen and Euling (2013). Here, we build upon this approach by using the chemical agnostic adverse outcome pathway (AOP) (Ankley et al. 2010; Villeneuve et al. 2014a, b) as a framework for organizing and integrating data sources for genetic susceptibility, and to identify mechanistically relevant molecular targets. By investigating the genetic components of environmental toxicant susceptibility etiology, we can better understand the mechanistic components implicated in adverse outcomes, and characterize the individual and population level heterogeneity and heritability of exposure risk. We present a computational approach that integrates mechanistic data associated with an AOP with data capturing human genetic variability and function. This AOP-anchored genetic susceptibility approach uses both functional genomic and human polymorphism data to characterize individual and population level genetic variation that may impact responses to environmental chemicals. Here, we first review the current state of the science in the incorporation of genetic data to define susceptibility to environmental chemicals (also refer to Chui and Rusyn, this issue). Secondly, we outline a general computational approach that implements existing adverse outcome information with available functional genomic data to define human genetic susceptibility. Lastly, we discuss future directions and the preparation for case study to illustrate aspects of the approach. The approach presented here is a first step towards defining genetic susceptibility using publicly available data, and contributes to the long-term goal of a more comprehensive model of susceptibility and risk for HHRA in the future.

Review of existing approaches for the study of human diversity in chemical response

To identify examples using genetics to define susceptibility to environmental chemicals, we used EPA's HERO database (http://www.hero.epa.gov) to simultaneously search PubMed, Toxnet, and Web of Science (described in S1). Subsearches were designed in collaboration with EPA science librarians to identify the literature on (1) approaches for incorporating genetic susceptibility information into HHRA, including a query about whether the AOP construct had been used previously; and (2) human genetic susceptibility data

sources (review articles, primary research, and databases of information). Subsequent sorting of the literature was performed using defined inclusion and exclusion criteria. In addition to the HERO literature searches, "snowball searches" (based on backward searches and discussions with experts) and Google searches were used to identify publicly available human genetic sequence annotation resources and phenotypic inference tools. Key conclusions from the literature subsearch results are that (1) there is substantial interest in utilizing polymorphism data to inform HHRA, but limited data exists describing the associations between chemical exposures and health outcomes; and (2) no prior approach had utilized the AOP construct for this purpose. In general, current efforts to characterize susceptibility for chemical risk assessment typically fall into three subject areas: (1) disease/biomarker characterization, (2) in vitro human diversity panel approaches, and (3) in silico toxicokinetic approaches.

Disease-biomarker studies

Davis and Burgoon (2015) studied a SNP (rs13266634 in SLC20A8 zinc transporter) previously identified to be associated with type II diabetes mellitus (T2DM) (Cauchi et al. 2010) and calculated the attributable risk based on reported SNP frequency in three cohorts (Mexican, Asian, and European). The authors then applied this information to the California demographic data (2007–2011 American Communities Survey: Caucasian, Asian, and Mexican (of any race) populations) to identify community-specific "hotspots" at increased risk of developing T2DM. Demographic data were obtained from the US Census Bureau and counties with > 25% of their population residing in census tracts in the highest quintile for population attributable risk (PAR) were identified as "hotspots." Although this approach is limited to the data available in different populations, this genetic risk measure could be used to adjust for T2DM health outcomes in a population-specific manner.

A second disease focused study implemented the Environmentally sensitive genes (ESG) from the NIEHS' Environmental Genome Project (http://egp.gs.washington.edu/); (Wilson 2004) to identify variants in genes statistically associated with myeloproliferative neoplasms (MPN) (Gross-Davis et al. 2015). Genes that contained nsSNPs, with a > 5% minor allele frequency (MAF), were considered. One hundred and fourteen Phase I and Phase II metabolism genes were initially investigated. The Genome Variation Server (GVS) (http://gvs.gs.washington.edu) was used to facilitate nsSNP identification and selection for genotyping, where 82 genes with coding region SNPs were selected. Twenty-one genes were identified to be relevant to a mutagenic pathway, and variants in 13 ESGs were found to be statistically significantly associated with MPN risk (Gross-Davis et al. 2015). While this study provides qualitative information on gene-environment interactions, measurements of exposure were not included in this study.

In vivo and in vitro diversity panel approaches

Much of the work in genetics and medicine surrounding human diversity panels or more specifically, human diversity cell lines, has been for the purpose of gaining understanding into human variation, genetic diversity, and human evolution, though these human genetic diversity studies were not incorporated in environmental toxicology until relatively recently.

In a recent toxicological application of the 1000 Genomes resource, Abdo (2015a, b) evaluated lymphoblastoid cell lines (LCLs) from the 1000 Genomes Project against a 179 chemical subset from the NTP's 1408 chemical library (1000 Genomes Project Consortium 2010) to identify genetic variants associated with cytotoxic responses. Using GWAS data, Abdo (2015a, b) found genetic variants in transmembrane and solute carrier genes that were associated with cytotoxicity. Depending on the level of chemical concentration in vitro (factor of 3), these variants captured the response in the top 1% of "sensitive" individuals. However, there were some chemicals outside of this range (factor of 10 rather than 3) that may be good targets for further evaluation. Works like these can provide the evidence needed to prioritize chemicals and genetic loci for further evaluation. Chui et al. (2016) proposed a tiered workflow whereby chemicals eliciting the most divergent population responses are computationally identified using in silico models (Eduati et al. 2015), and then further characterized to determine population-wide responses. An integrative in vitro and in silico tiered approach may have the most realistic application to chemical prioritization and translation to hazard identification by including computational approaches to assess physiochemical properties and structure–activity relationships.

Where the in vitro models are insufficient (i.e., miss key biological spaces, lack comparative metabolism data, and lack toxicokinetics) in vivo models can gain traction. There are several new rodent population models available that were recently reviewed by Harrill (2017). The Collaborative Cross (CC) (Churchill et al. 2004 , Churchill et al. 2012; Threadgill and Churchill 2012), a breeding scheme designed to generate a large panel of genetically diverse recombinant inbred strains from 8 genetically distinct founder strains, was recently applied for the first time in toxicological studies. Cichocki (2017) and Venkatratnam (2017) used CC sample populations to investigate how genetic variation influenced tetrachloroethylene and trichloroethylene, respectively, metabolism to metabolites including trichloroacetic acid (TCA). TCA is a known PPARα agonist thought to mediate toxicities elicited by metabolism of the chlorinated olefins such as TCE (Corton 2008), a metabolite also influenced by

genetic variation (Bradford et al. 2011). Even though rodent studies would still require an interspecies uncertainty factor of 10, we can begin to use these in vivo data to build evidence-based biological plausibility for pathways linking exposure to outcome that incorporate critical toxicodynamic, toxicokinetic, and dose–response information. For example, notable findings from Cichocki (2017) and Venkatratnam (2017) found that TCA levels did not covary across the CC stains, but the PPARa responsive genes *Acox1* and *Cyp4a10* mRNA levels did, suggesting interactions between genetic variants that influenced the PPARa network response to TCA. This critical MOA information would be missed if not for in vivo data. Furthermore, these studies can help identify critical genetic variants that might put human populations at risk of an adverse health outcome from chlorinated olefin exposure. Clearly, there is a need to organize in vitro, in silico, and in vivo findings into an integrative AOP framework, further supported by human exposure, biomarker, and epidemiologic evidence (if available), so that causal relationships can be defensibly determined.

Modeling human toxicokinetic variability through in silico approaches

Polymorphism data are routinely incorporated into computational models to quantitate population variability to drug responses (Rostami-Hodjegan and Tucker 2007; Rostami-Hodjegan 2012; Sager et al. 2015). Many of these models use in vitro-in vivo extrapolation (IVIVE) approaches to integrate in vitro chemical-specific data (e.g., intrinsic hepatic clearance, fraction of the chemical unbound in plasma) into toxicokinetic (TK) models [e.g., physiologically based pharmacokinetic (PBPK)], where TK variability is known to exist among different populations and may impact tissue dosimetry (Clewell et al. 2002). Specifically, variability in age (Johnson et al. 2006), race (Inoue et al. 2006), gender (Polak et al. 2012), and genetic polymorphisms (including drug metabolizing enzymes) (Dickinson et al. 2007; Gertz et al. 2014) have been simulated in these models using Monte Carlo methods to estimate diversity in pharmacokinetic behavior. The power of these IVIVE approaches allows us to make predictions of specific compound responses using in vitro human data without the need for allometric scaling of animal data. These modeling approaches have been used in estimating population TK variability for environmental chemicals (Clewell and Andersen 1996). The combining of genetic polymorphism distributions with their effects on TK distributions illustrates the potential to provide quantitative estimates of TK variability for sensitive populations (Johanson et al. 1999, Timchalk et al. 2002).

Computational approaches are utilizing human variability parameters and in vitro high-throughput screening (HTS) data together in population-specific dosimetry models for estimating toxicity potential of non-pharmaceutical compounds (Rotroff et al. 2010; Wetmore et al. 2012, 2014; Wambaugh et al. 2015). The idea was to convert the chemical-assay HTS data, gathered from the Federal Tox21 partnership (Kavlock et al. 2009) and the U.S. EPA's ToxCast program (Dix et al. 2007) to equivalent doses utilizing information on in vitro measured TK parameters (fraction of the chemical unbound in plasma and the intrinsic metabolic clearance), physicochemical properties and population variability using the Simcyp simulator (Jamei et al. 2009a, b). Simcyp has a population variability function, which utilizes information on age, weight, height, sex, genetics, race, and disease. Specifically, the frequencies of CYP and UGT enzyme polymorphism are captured in this program (Ginsberg et al. 2009). Specifying a population of 100 healthy adults (20–50 years old) of both sexes, a subset of chemicals were identified as having equivalent HTS doses lower than or equal to estimated U.S. population exposures (Rotroff et al. 2010; Wetmore et al. 2012; Wambaugh et al. 2015), indicating potential biological activity for these environmental chemicals. Furthermore, individual CYP and UGTs expression profiles were measured for 9 chemicals, where life-stage and ethnicity models revealed up to a 13.1-fold difference in steady-state blood concentrations for the most sensitive population over the median (Wetmore et al. 2014). Subsequent work has broadened the use of these approaches by providing a publicly available tool [High-Throughput Toxicokinetic (HTTK) R-package] to perform these analyses and incorporate population variability (Wambaugh et al. 2015; Pearce et al. 2017; Ring et al. 2017). Whether the population parameters used in these computational models are sufficient for the purposes of risk assessment and how computational approaches that incorporate TK and TD can be integrated with more comprehensive genetic susceptibility information across diverse outcomes has yet to be systematically investigated.

Incorporation of genomic information into an AOP framework

The AOP framework provides an integrative tool for adapting existing data, including HTS assay information like that generated by the Federal ToxCast and Tox21 programs (Dix et al. 2007; Collins et al. 2008). AOPs are generally described by key events (KE) within different levels of organization (molecular, cellular, tissue, organism, population) that when perturbed lead to an adverse outcome (AO). The MIE is the first KE in a pathway, thought to trigger the adverse outcome, but can also act as a KE in alternate AOPs. The entire path (linear or networked) is anchored by an upstream molecular initiating event (MIE), the first point of chemical–molecular interaction, and leading one or more

key events (KE) to an adverse outcome (Ankley et al. 2010; Villeneuve et al. 2014a, b).

The pace at which genetic information can be incorporated into risk assessment decisions depends to a certain extent on expert-driven AOP development. The AOP development process has been described (Villeneuve et al. 2014a, b). Formal guidance on AOP development can be found in the "Guidance on Developing and Assessing AOPs" document (IOMC 2017). Additionally, confidence frameworks have been developed for the implementation of AOPs for regulatory purposes (Patlewicz et al. 2015). These efforts have been underway for several years and publicly available web-based applications exist, such as the AOP KnowledgeBase (AOP-KB) (http://aopkb.org/index.html), which includes modules like the AOPWiki (https://aopwiki.org/), a central repository for all AOPs developed as part of the Organization for Economic Co-operation and Development (OECD) AOP development effort by the Extended Advisory Group on Molecular Screening and Toxicogenomics. EPA efforts like the AOP-DB, a database tool that integrates publicly available AOP resources with gene, pathway, and disease information (Pittman et al. 2017), make it possible to link AOP information from the AOPWiki to gene target and in vitro assay information from ToxCast, for example (S2). These data will soon be available to the public through the EPA Chemistry Dashboard (https://comptox.epa.gov/dashboard), and possibly as a user-friendly frontend in the near future dependent on EPA research funding available. Resources like the AOP-KB, AOP-Wiki, and AOP-DB are integral in the computational prediction of AOPs discussed by Oki (2016).

In order to apply genetic information into risk assessment paradigms, it is critical to integrate genetic data so that the relationship between chemical-gene target and population polymorphism data is clearly linked to an adverse outcome (Fig. 1). Given the AOP concept is relatively novel, it is not surprising that there is a limited number of well-characterized and documented AOPs (https://aopwiki.org/). For this reason, it is important for any susceptibility approach to integrate disease, pathway, and chemical association information when available. Table 1 lists the data sources necessary to integrate AOP information with functional genomic targets, as well as the data sources used to characterize selected targets for tissue level expression and individual and population level variation. Figure 2 describes the proposed computational workflow for the characterization of human individual and population level susceptibility to an adverse outcome. The derivation of a relevant gene list associated with a biological process and leading to an adverse outcome (Step 1), underlines that a gene set could be selected for a single or multiple biological processes simultaneously. The selected gene set is then validated (Step 2) to establish that is in fact relevant for the biological process in question. The individual genes in question are then characterized in terms of their regulatory regions (Step 3), so that the functionally relevant SNPs can be selected (Step 4), and finally characterized for individual and population level variation (Steps 5 and 6).

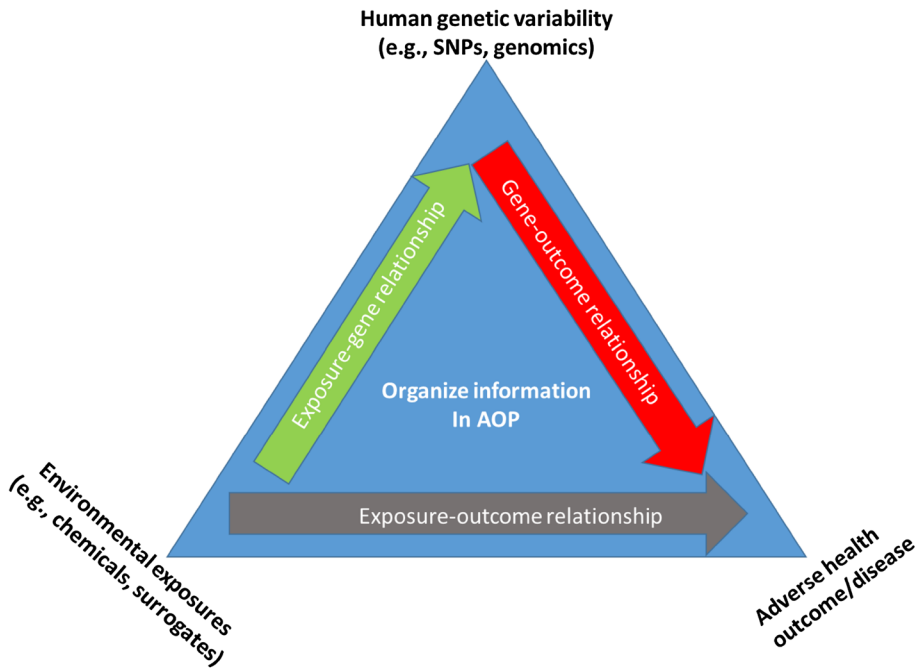

Fig. 1 Conceptual model of associations needed between genetic human genetic variability, environmental chemical exposure, and adverse outcomes to inform genetic susceptibility in HHRA. Note that multiple data sources linking environmental exposures to human genetic variability information, and in turn linking genetic susceptibility to environmental exposures to adverse outcomes (illustrated by arrows), are needed to be informative to HHRA

Table 1 Data source, level of biological organization, and data type included

Data source	Level of biological organization							Data type	Data source location
	Molecular	Biological pathway	Cellular	Tissue	Organ	Individual	Population		
dbSNP	Y	N	N	N	N	Y	Y	R, P, C, I	https://www.ncbi.nlm.nih.gov/projects/SNP/
dbVar	Y	N	N	N	N	Y	Y	R, C, A	https://www.ncbi.nlm.nih.gov/dbvar
dbGaP	Y	–	–	–	–	Y	Y	A	https://www.ncbi.nlm.nih.gov/gap
ClinVar	Y	–	–	–	–	Y	N	A	https://www.ncbi.nlm.nih.gov/clinvar/
CTD	Y	Y	Y	–	–	–	–	C	http://ctdbase.org/
REACTOME	Y	Y	–	–	–	N	N	C	http://www.reactome.org/
1000 Genomes	Y	–	–	–	–	Y	Y	R, P	http://www.internationalgenome.org/data/
UniProt	Y	–	–	–	–	–	–	C	uniprot.org
Ensembl	Y	Y	Y	Y	Y	Y	Y	A, C, P	http://ensembl.org/
ENCODE	Y	–	Y	–	–	–	N	R	ENCODE https://www.encodeproject.org/
GTEx	Y	–	–	Y	Y	Y	Y	R, P	GTEx https://gtexportal.org/home/
AOP-Wiki	Y	Y	–	–	–	N	N	C	AOP Wiki https://aopwiki.org/
NHGRI-EBI GWAS Catalog	Y	–	–	–	–	Y	Y	A	https://www.ebi.ac.uk/gwas/
GIANT	Y	Y	Y	Y	Y	–	–	A, P	http://giant.princeton.edu/download/
s1500 Gene set	Y	Y	Y	–	–	–	–	R, P	https://ntp.niehs.nih.gov/results/tox21/s1500-gene-set-consensus-strategy-index.html
NIEHS SNPs Program	Y	–	–	–	–	–	–	R, P	http://egp.gs.washington.edu/
DisGeNET	Y	–	–	–	Y	Y	–	A	http://www.disgenet.org
ToxCast	Y	–	–	–	N	N	N	R, P	https://www.epa.gov/chemical-research/toxicity-forecaster-toxcasttm-data
OMIM	–	–	–	–	Y	Y	–	C	http://www.ncbi.nlm.nih.gov/omim

Table format adapted from Oki et al. (2016)

Y definitely covers this level of organization, *N* definitely does not covers this level of organization, *I* inferred data, – ambiguous, *A* data aggregator, *C* curated data, *R* raw data, *P* processed data

Steps 1 and 2: mechanism-relevant gene lists for environmental toxicology

To characterize population genetic variation for a single observed, complex phenotype, it is critical to identify the mechanistic targets (e.g., genes, proteins, and epigenetic factors) controlling the phenotype as completely as possible. Because the knowledge of any given complex phenotype, and the completeness of genetic elements known to control it, depends wholly on the state of the science at a given time, a mechanistically relevant gene list is temporally dependent and can be derived in a variety of ways, such as from an AOP with defined molecular MIE and KEs, a disease or phenotype of interest, a list of assay targets with environmental association, or a computationally derived list based on pathways or ontologies, for example. Figure 3 illustrates the overlap in number of molecular targets between four data sources (refer to Table 1 for links to sources), expert-derived AOPs obtained from the joint EPA-OECD AOP-Wiki (http://www.AOPwiki.org);

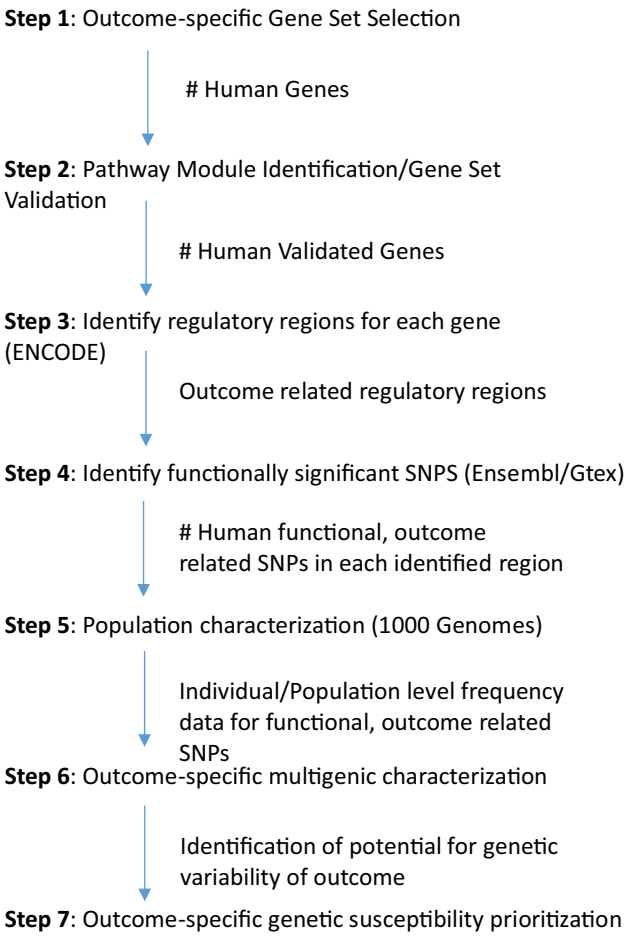

Fig. 2 Proposed computational workflow illustrating how a given outcome-specific, gene set of interest can be selected (Step 1), validated (Step 2), regulatory regions identified (Step 3), SNPs of interested selected (Step 4), characterized for human individual and population level, multigenic susceptibility to an adverse outcome (Step 6), and finally, prioritized by outcome for potential for human genetic susceptibility (Step 7)

refer to (Pittman et al. 2017), two notable environmentally responsive gene (ERG) sets derived as part of NIH funded initiatives, the NIEHS Environmental Genome Project (EGP) (see Mortensen and Euling 2013 for discussion) and the Tox21 Phase III "s1500" Gene Set (Federal Register 2015) based on the 1000 landmark genes or L1000 (Lamb et al. 2006), and disease-associated genes obtained from DisGeNET, a human disease gene repository (Piñero et al. 2016). Selection of environmentally responsive and disease phenotype targets at Step 1 (Fig. 2) that are present in existing AOPs is one way to begin the gene list selection for a global analysis.

Molecular (e.g., gene–gene, protein–protein) interactions have been described in terms of their modularity (Ideker et al. 2002; Bar-Joseph et al. 2003; Snel and Huynen 2004; Qi and Ge 2006; Mitra et al. 2013). The modularity of functional gene sets becomes important for the current approach in that we are selecting unlinked, molecular targets whose coordinated function can result in an adverse outcome or phenotype of interest. Because an adverse outcome, like a disease, is rarely the consequence of a single gene defect, but results from the perturbations of one or more functionally related gene modules (Zaghloul and Katsanis 2010; Barabasi et al. 2011), and because the gene targets thought to be associated with an adverse or disease outcome may be incomplete, we propose a workflow that includes an independent validation method (Step 2, Fig. 2) of the outcome gene sets. With the inclusion of an independent validation step, such as pathway-based association (PBA) or other statistical measure, we increase the probability of capturing all functionally relevant targets, as well as any population-specific variants, for downstream characterization. There has been a substantial amount of work to computationally determine gene lists using PBA and related methods for predictive modeling of disease risk (Torkamani et al. 2008; Dudley and Butte 2009; Fridley and Biernacka 2011; Yang et al. 2011; Fernandez et al. 2013; Mooney and Wilmot 2015), as well as chemical toxicity studies (Fujibuchi et al. 2009; Smalley et al. 2010; Judson et al. 2012), and many methods now incorporate network, multilocus methods that utilize GWAS information and prior biological knowledge (Azencott et al. 2013; Croteau-Chonka et al. 2015; Ayati and Koyuturk 2016; Hormozdiari et al. 2016a, b; Reyes-Gibby et al. 2017). Approaches used in GWAS, such as kernal machine testing for variable selection, could also be used in this context in a unique way. This strategy is just one of many to prioritize groups of SNPs related to biological processes (He et al. 2016). Tissue and cell-specific approaches (Greene et al. 2015) (Table 1), that incorporate Bayesian and network methodologies, are also useful in illuminating functional gene groupings in specific tissue types and related diseases.

A possible iteration to Step 2, in some ways to serve as replication, could be to implement pathway analysis approaches used in GWAS to verify the biology of groups of SNPs selected in an AOP. Many large-scale GWAS meta-analyses validate the SNP-specific, statistical analyses with pathway analyses to understand the implicated biology (Gharib et al. 2015). For a specific AOP, researchers following Fig. 2 outline could implement pathway analyses as a secondary check of the AOP-associated molecular variants for Step 2.

Steps 3 and 4: evaluating regulatory variation

Once we have obtained the validated gene list associated with AOs, it is possible to identify and characterize the regulatory elements that may contribute to variation in function (Steps 3 and 4, Fig. 2). Non-coding sequence polymorphisms, located in 5′ promoter regions, 3′ untranslated

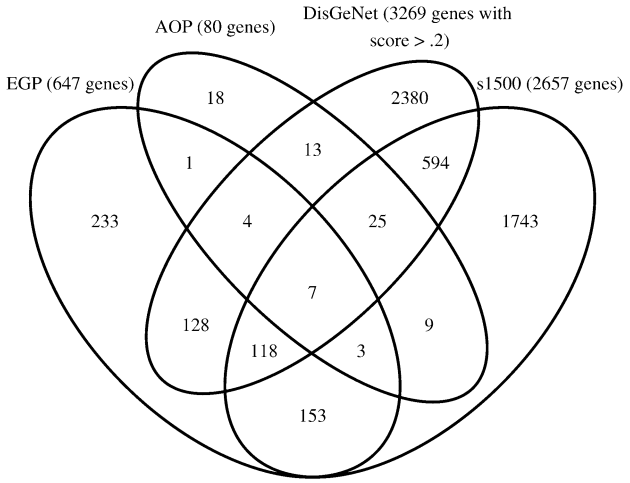

Fig. 3 Venn diagram illustrating the overlap in number of molecular targets between four data sources, expert-derived AOPs from the AOP-Wiki (http://www.AOPwiki.org), two notable environmentally responsive gene sets NIEHS s1500 (https://ntp.niehs.nih.gov/results/tox21/s1500-gene-set-consensus-strategy-index.html) and the NIEHS EGP (http://egp.gs.washington.edu/), and disease-associated genes present in the DisGeNET database (http://www.disgenet.org) (Refer to Table 1)

regions (UTRs), enhancers, or elsewhere in the genome may impact the regulation of gene expression through a myriad of mechanisms, including but not limited to protein stability, isoform variation, abundance, or tissue specificity. A variety of computational tools have been developed for the evaluation of polymorphisms in protein-coding sequences, which estimate the functional significance of exonic polymorphisms in terms of protein structure (Kumar et al. 2009; Adzhubei et al. 2010). Non-coding or intronic regions of the genome are far less tractable. We know from GWAS that many trait-associated variants map to non-coding regions (Deplancke et al. 2016), which supports the idea that a high a proportion of observed phenotypic variability in humans may be due to variation in gene regulation (Lonsdale et al. 2013). Regulatory variants are also thought to occur at distances of 100 s of kilobases from associated coding regions (GuhaThakurta 2006), making the computational screening for regulatory variants using publicly available tools challenging. Further, trans-regulatory processes are also important, due to 3-day conformational changes in chromatin structure, long distance enhancer-promoter contacts, non-strand-specific gene expression, etc. Large-scale efforts like the ENCODE (Encyclopedia of DNA Elements) Project Consortium (2007) aim to identify and characterize all functional elements in the human genome (protein-coding genes, regulatory elements, etc.). To incorporate important regulatory variants into the current workflow, we focus on the integration of three data sources (Step 4, Fig. 2; Table 1; S3): (1) the ENCODE Project which identifies functional elements in human genome sequence (ENCODE Project Consortium 2004, ENCODE Project Consortium); (2) Ensembl (Yates et al. 2015; Zerbino et al. 2015), which provides regulatory element data for humans in an annotated reference genome, the Ensembl Regulatory Build; and (3) Genotype-Tissue Expression (GTex) Portal (Lonsdale et al. 2013), an NIH Common Fund Program, which provides data on tissue specific and shared regulatory human gene Expression Quantitative Trait Loci (eQTL) variants.

Together, these data sources when integrated with individual and population level variation sources (Table 1) make it possible to establish the basis of a global or systems level approach to the evaluation of genetic variability and chemical susceptibility. By combining eQTL data with candidate gene sets such as those involved in an AOP, we can evaluate functionally specific target variation in expression on a pathway, tissue, and outcome-relevant level within and between populations. One current limitation of the most recent GTex version 6 (V6p) is that samples disproportionately represent white American (84.3%) and African American (13.7%) populations, with a total number of donor individuals ~ 500. Nonetheless, the integration of coding, non-coding and regulatory variation with tissue-specific, functional effects is essential to future efforts to characterize the impact of human genetic variability on chemical susceptibility.

Steps 5, 6, and 7: incorporating publicly available human variation sources for outcome-specific prioritization

With a defined region for each gene, typically ~ 5 KB both upstream and downstream of each exonic region, and a selection of functionally characterized variants corresponding to each AOP, we can begin to characterize those variants in terms of their allele frequencies (Step 5, Fig. 2) and build multigenic haplotypes across human population groups that are outcome specific (Step 6, Fig. 2). Here, we use data obtained from the 1000 Genomes Project, accessible through the National Center for Biotechnology (NCBI) Database of single-nucleotide polymorphisms (dbSNP) (Table 1). dbSNP also contains data on small-scale variants (insertions, deletions, microsatellites, and non-polymorphic variants). NCBI, a division of the National Library of Medicine (NLM) at the National Institutes of Health (NIH), is now the primary repository for molecular genetic data. These data for humans come from two large initiatives: the International HapMap Project (The International HapMap Consortium 2005; International HapMap Consortium 2007) and the 1000 Genomes Project (1000 Genomes Project Consortium 2012, 1000; Genomes Project Consortium 2015b). With the information gained in Step 6, we are able to prioritize on an outcome-specific basis the level of genetic variability observed at the individual and population levels in Step 7.

Future directions

Setting the stage for a case study

With the goal of developing an AOP-based workflow for the evaluation of human population genetic susceptibility to environmental toxicants, we investigated a variety of data sources. Table 1 is organized to illustrate the presence of information at each level of complexity in the AOP framework. Table 1 lists each relevant data source, including those that describe mechanistic information pertaining to adverse outcomes, and those sources used to obtain human individual and population level variability information. Figure 2 illustrates the data integration workflow proposed in the current approach. Figure 3 illustrates the overlap of gene targets for four environmentally relevant datasets, described above. Though we envision the proposed workflow to be performed globally, across many AOs, we focus for the present review on the goal of selection of a candidate AOP, with a MIE suitable for future case study. In preparation for future case study, we outlined the workflow for a single molecular gene target, or MIE. To identify a suitable target, we queried the AOPwiki for KEs known to be variable across human populations, and found three molecular targets with documented variation in response in humans: Glucose-6-phosphate dehydrogenase (G6PD) (Tishkoff 2001), human thyroid peroxidase (TPO) (Fu et al. 2016; Graf et al. 2017), and cytochrome 2E1(CYP2E1) (Lee et al. 2008) (Fig. 4). We built association networks to illustrate the relationships between the selected AOP-MIEs and associated chemicals, by querying the AOP-DB for AOP-gene targets and chemical-gene associations (Davis et al. 2016; Pittman et al. 2017). Figure 4 illustrates that the majority of TSCA high-priority chemicals (US EPA 2014) are associated with the

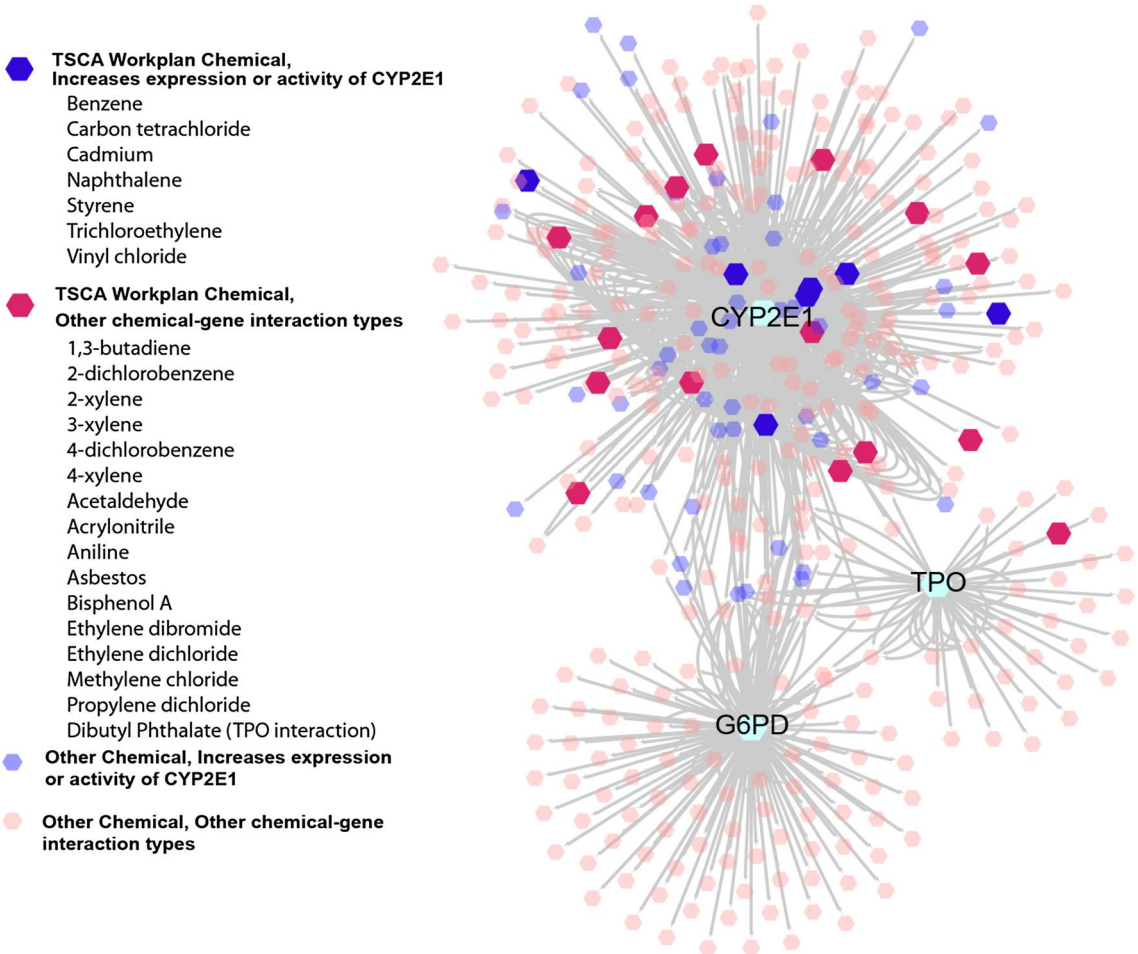

Fig. 4 Chemical-gene association network illustrating chemicals associated with CYP2E1 in humans, TPO, and G6PD, according to the CTD database. Large nodes indicate TSCA Workplan Chemicals, and small nodes indicate non-TSCA Workplan chemicals. Blue nodes indicate increased expression or activation of CYP2E1, in concordance with the AOP, "chronic activation of CYP2E1 leading to liver cancer." Red and pink nodes indicate chemical-gene association with CYP2E1, according to the literature-based chemical-gene information in CTD. (Color figure online)

AOP-MIE CYP2E1, but not the other two candidates, and indicates the specific TSCA high-priority chemicals associated with this AOP-MIE in humans. Figure 5 illustrates TSCA high-priority chemicals associated with the AOP-MIE CYP2E1 for human and mouse, and indicates that 1-bromopropane (1-BP), a TSCA chemical recently identified to cause adverse neurological, reproductive, kidney, and liver effects in humans (Federal Register 2016), association with CYP2E1 is based entirely on evidence from the mouse (Garner et al. 2007; Liu et al. 2009). In addition, to illustrate and confirm the mechanism of the AOP [Chronic Cyp2E1 Activation Leading to Liver Cancer (AOP220: Webster, Lambert, Yauk, http://aopwiki.org; OECD Project 1.24)], we explored disease association in the network as shown in Fig. 6, where liver-associated disease is indicated, confirming that the outcome mechanism described for this AOP is accurate.

Future challenges and opportunities: characterizing genetic susceptibility using an AOP framework for HHRA

The computational approach developed here is generalizable to select AOP molecular targets that can be identified using the AOP-DB (Pittman et al. 2017) that is seamlessly connected to the AOPWiki, making it possible to both look across all AOPs or at a specific AOP of interest. The next challenges include validating AOP relevant genes for causality between the genes and the health outcome. There are available methods for validating AOP relevant genes, discussed above, including methods originally developed for other purposes e.g., GWAS to infer genes and predict trait-associated networks (Akula et al. 2011; Lee et al. 2011). It is clear that the incorporation of epigenetic data in this workflow, though beyond the scope of the current paper, is necessary for a more complete characterization of susceptibility for risk assessment (Cote et al. 2017), and could easily be incorporated as an additional step in the

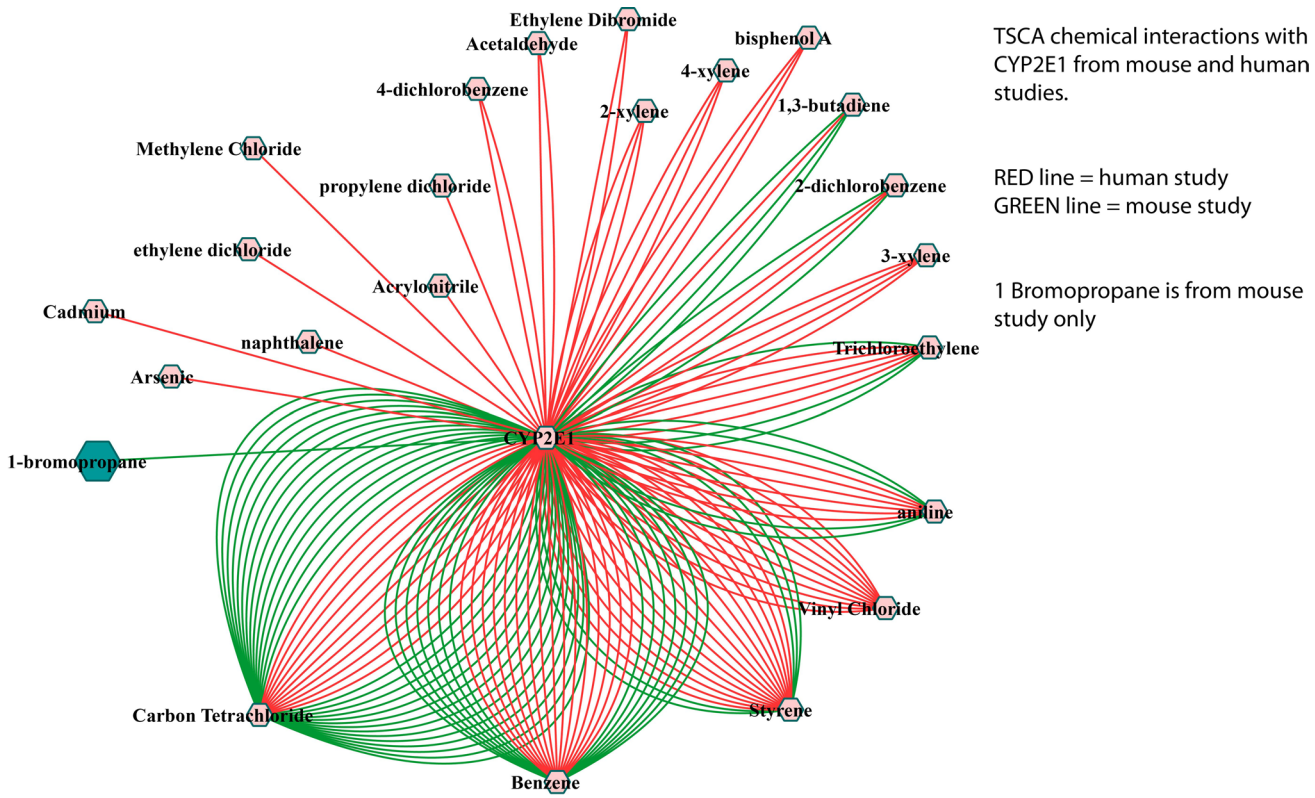

Fig. 5 Chemical-gene association network illustrating chemicals associated with CYP2E1 in humans and mouse, according to the CTD Database, and selecting for only TSCA Workplan Chemicals. Red edges indicate human studies; green edges indicate mouse studies. 1-Bromopropane (1-BP) has been highlighted, whereby information associating 1-BP with CYP2E1 from CTD is derived from mouse studies only. (Color figure online)

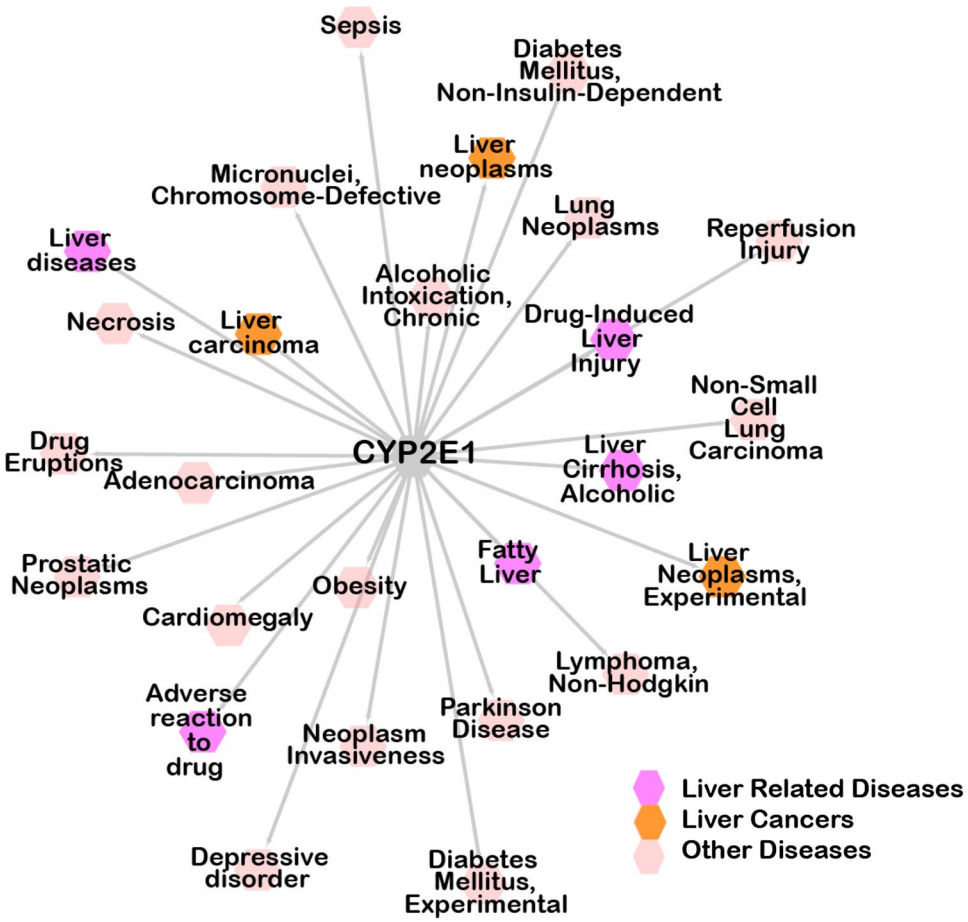

Fig. 6 Disease-gene association network illustrating diseases associated with CYP2E1 in humans, according to the DisGeNet database. Score is indicated by edge length; shorter edges indicate stronger associations. The strongest association is with fatty liver, with score 0.206. Associations weaker than 0.01 are omitted

approach described here. A second challenge is the interpretation and visualization of population susceptibility for each multigenic AOP in a way that is readily accessible for risk assessors. However, the prioritization of outcomes for genetic susceptibility potential readily identifies which outcomes are subject to genetic variability difference in humans. Though the AOP-anchored approach presented here for specifically genetic susceptibility is but one part of a more comprehensive model of human susceptibility to environmental chemicals needed in HHRA, we have described the first step for the inclusion of existing, publicly available information to identify the potential for human susceptibility across adverse outcomes.

Acknowledgements We would like to acknowledge funding for this project from the EPA's Office of Research and Development (ORD) Human Health Risk Assessment (HHRA) National Research Program. We would like to thank Debora Segal, Michael Wright, Annie Jarabek, Brian Chorley, Hisham El-Masri for many helpful discussions on the approach, and thoughtful review of the manuscript. We also thank two anonymous reviewers. We thank Ryan Jones and William Griffin at HERO at EPA for assistance with literature searches.

Compliance with ethical standards

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

References

Abdo N, Wetmore BA, Chappell GA, Shea D, Wright FA, Rusyn I (2015a) In vitro screening for population variability in toxicity of pesticide-containing mixtures. Environ Int 85:147–155

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 (2015b) Population-based in vitro hazard and concentration-response assessment of chemicals: the 1000 genomes high-throughput screening study. Environ Health Perspect 123(5):458–466

Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR (2010) A method and server for predicting damaging missense mutations. Nat Methods 7(4):248–249

Akula N, Baranova A, Seto D, Solka J, Nalls MA, Singleton A, Ferrucci L, Tanaka T, Bandinelli S, Cho YS, Kim YJ, Lee J-Y, Han

B-G, McMahon FJ (2011) A network-based approach to prioritize results from genome-wide association studies. PLoS ONE 6(9):e24220

Ankley GT, Bennett RS, Erickson RJ, Hoff DJ, Hornung MW, Johnson RD, Mount DR, Nichols JW, Russom CL, Schmieder PK, Serrrano JA, Tietge JE, Villeneuve DL (2010) Adverse outcome pathways: a conceptual framework to support ecotoxicology research and risk assessment. Environ Toxicol Chem 29(3):730–741

Ayati M, Koyuturk M (2016) PoCos: population covering locus sets for risk assessment in complex diseases. PLoS Comput Biol 12(11):e1005195

Azencott CA, Grimm D, Sugiyama M, Kawahara Y, Borgwardt KM (2013) Efficient network-guided multi-locus association mapping with graph cuts. Bioinformatics 29(13):i171–i179

Barabasi AL, Gulbahce N, Loscalzo J (2011) Network medicine: a network-based approach to human disease. Nat Rev Genet 12(1):56–68

Bar-Joseph Z, Gerber GK, Lee TI, Rinaldi NJ, Yoo JY, Robert F, Gordon DB, Fraenkel E, Jaakkola TS, Young RA, Gifford DK (2003) Computational discovery of gene modules and regulatory networks. Nat Biotechnol 21(11):1337–1342

Bradford BU, Lock EF, Kosyk O, Kim S, Uehara T, Harbourt D, DeSimone M, Threadgill DW, Tryndyak V, Pogribny IP, Bleyle L, Koop DR, Rusyn I (2011) Interstrain differences in the liver effects of trichloroethylene in a multistrain panel of inbred mice. Toxicol Sci 120(1):206–217

Cauchi S, Del Guerra S, Choquet H, D'Aleo V, Groves CJ, Lupi R, McCarthy MI, Froguel P, Marchetti P (2010) Meta-analysis and functional effects of the SLC30A8 rs13266634 polymorphism on isolated human pancreatic islets. Mol Genet Metab 100(1):77–82

Chiu WA, Wright FA, Rusyn I (2016). A tiered, Bayesian approach to estimating of population variability for regulatory decision-making. ALTEX 34:377

Churchill GA, Airey DC, Allayee H, Angel JM, Attie AD, Beatty J, Beavis WD, Belknap JK, Bennett B, Berrettini W, Bleich A, Bogue M, Broman KW, Buck KJ, Buckler E, Burmeister M, Chesler EJ, Cheverud JM, Clapcote S, Cook MN, Cox RD, Crabbe JC, Crusio WE, Darvasi A, Deschepper CF, Doerge RW, Farber CR, Forejt J, Gaile D, Garlow SJ, Geiger H, Gershenfeld H, Gordon T, Gu J, Gu W, de Haan G, Hayes NL, Heller C, Himmelbauer H, Hitzemann R, Hunter K, Hsu HC, Iraqi FA, Ivandic B, Jacob HJ, Jansen RC, Jepsen KJ, Johnson DK, Johnson TE, Kempermann G, Kendziorski C, Kotb M, Kooy RF, Llamas B, Lammert F, Lassalle JM, Lowenstein PR, Lu L, Lusis A, Manly KF, Marcucio R, Matthews D, Medrano JF, Miller DR, Mittleman G, Mock BA, Mogil JS, Montagutelli X, Morahan G, Morris DG, Mott R, Nadeau JH, Nagase H, Nowakowski RS, O'Hara BF, Osadchuk AV, Page GP, Paigen B, Paigen K, Palmer AA, Pan HJ, Peltonen-Palotie L, Peirce J, Pomp D, Pravenec M, Prows DR, Qi Z, Reeves RH, Roder J, Rosen GD, Schadt EE, Schalkwyk LC, Seltzer Z, Shimomura K, Shou S, Sillanpaa MJ, Siracusa LD, Snoeck HW, Spearow JL, Svenson K, Tarantino LM, Threadgill D, Toth LA, Valdar W, de Villena FP, Warden C, Whatley S, Williams RW, Wiltshire T, Yi N, Zhang D, Zhang M, Zou F (2004). The Collaborative Cross, a community resource for the genetic analysis of complex traits. Nat Genet 36:1133–1137

Churchill GA, Gatti DM, Munger SC, Svenson KL (2012) The diversity outbred mouse population. Mamm Genome 23(9–10):713–718

Cichocki JA, Furuya S, Venkatratnam A, McDonald TJ, Knap AH, Wade T, Sweet S, Chiu WA, Threadgill DW, Rusyn I (2017) Characterization of variability in toxicokinetics and toxicodynamics of tetrachloroethylene using the collaborative cross mouse population. Environ Health Perspect 125(5):057006

Clewell HJ, Andersen ME (1996) Use of physiologically based pharmacokinetic modeling to investigate individual versus population risk. Toxicology 111(1–3):315–329

Clewell HJ, Teeguarden J, McDonald T, Sarangapani R, Lawrence G, Covington T, Gentry R, Shipp A (2002) Review and evaluation of the potential impact of age- and gender-specific pharmacokinetic differences on tissue dosimetry. Crit Rev Toxicol 32(5):329–389

Collins FS, Gray GM, Bucher JR (2008). TOXICOLOGY: transforming environmental health protection. Science 319(5865):906–907

Corton JC (2008) Evaluation of the role of peroxisome proliferator-activated receptor alpha (PPARalpha) in mouse liver tumor induction by trichloroethylene and metabolites. Crit Rev Toxicol 38:857–875

Cote I, McCullough SD, Hines RN, Vandenberg JJ (2017) Application of epigenetic data in human health risk assessment. Curr Opin Toxicol 6:71–78

Croteau-Chonka DC, Rogers AJ, Raj T, McGeachie MJ, Qiu W, Ziniti JP, Stubbs BJ, Liang L, Martinez FD, Strunk RC, Lemanske RF, Liu AH, Stranger BE, Carey VJ, Raby BA (2015) Expression quantitative trait loci information improves predictive modeling of disease relevance of non-coding genetic variation. PLoS ONE 10(10):e0140758

Davis JA, Burgoon LD (2015) Can data science inform environmental justice and community risk screening for type 2 diabetes? PLoS ONE 10(4):e0121855

Davis AP, Grondin CJ, Johnson RJ, Sciaky D, King BL, McMorran R, Wiegers J, Wiegers TC, Mattingly CJ (2016) The comparative toxicogenomics database: update 2017. Nucleic Acids Res 45(D1):D972–D978

Deplancke B, Alpern D, Gardeux V (2016) The genetics of transcription factor DNA binding variation. Cell 166(3):538–554

Dickinson GL, Lennard MS, Tucker GT, Rostami-Hodjegan A (2007) The use of mechanistic DM-PK-PD modelling to assess the power of pharmacogenetic studies ? CYP2C9 and warfarin as an example. Br J Clin Pharmacol 64(1):14–26

Dix DJ, Houck KA, Martin MT, Richard AM, Setzer RW, Kavlock RJ (2007) The ToxCast program for prioritizing toxicity testing of environmental chemicals. Toxicol Sci 95(1):5–12

Dudley JT, Butte AJ (2009). Identification of discriminating biomarkers for human disease using integrative network biology. Pac Symp Biocomput 2009:27–38

Eduati F, Mangravite LM, Wang T, Tang H, Bare JC, Huang R, Norman T, Kellen M, Menden MP, Yang J, Zhan X, Zhong R, Xiao G, Xia M, Abdo N, Kosyk O, N.-N.-U. D. T. Collaboration, Friend S, Dearry A, Simeonov A, Tice RR, Rusyn I, Wright FA, Stolovitzky G, Xie Y, Saez-Rodriguez J (2015) Prediction of human population responses to toxic compounds by a collaborative competition. Nat Biotechnol 33(9):933–940

ENCODE Project Consortium (2004). The ENCODE (ENCyclopedia of DNA elements) project. Science 306(5696):636–640

ENCODE Project Consortium (2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature 447(7146):799–816

Federal Register (2015). List of environmentally responsive human genes selected for use in screening large numbers of substances using toxicogenomic approaches. D. O. H. A. H. SERVICES. 80:20237–20238

Federal Register (2016) 1-Bromopropane (1-BP); availability of TSCA work plan chemical risk assessment for public review and comment. U EPA. 81 FR 12098:12098–12099

Fernandez RM, Bleda M, Luzon-Toro B, Garcia-Alonso L, Arnold S, Sribudiani Y, Besmond C, Lantieri F, Doan B, Ceccherini I, Lyonnet S, Hofstra RM, Chakravarti A, Antinolo G, Dopazo J, Borrego S (2013) Pathways systematically associated to Hirschsprung's disease. Orphanet J Rare Dis 8:187

FQPA (1996). Food Quality Protection Act of 1995: hearings before the Subcommittee on Health and Environment of the Committee on Commerce, House of Representatives, One Hundred Fourth Congress, first session, on H.R. 1627, June 7 and 29, 1995. Washington, U.S. G.P.O: for sale by the U.S. G.P.O., Supt. of Docs., Congressional Sales Office

Fridley BL, Biernacka JM (2011) Gene set analysis of SNP data: benefits, challenges, and future directions. Eur J Hum Genet 19(8):837–843

Fu C, Xie B, Zhang S, Wang J, Luo S, Zheng H, Su J, Hu X, Chen R, Fan X, Luo J, Gu X, Chen S (2016) Mutation screening of the TPO gene in a cohort of 192 Chinese patients with congenital hypothyroidism. BMJ Open 6(5):e010719

Fujibuchi W, Kim H, Okada Y, Taniguchi T, Sone H (2009). High-performance gene expression module analysis tool and its application to chemical toxicity data. Methods Mol Biol 577: 55–65

Garner CE, Sloan C, Sumner SC, Burgess J, Davis J, Etheridge A, Parham A, Ghanayem BI (2007) CYP2E1-catalyzed oxidation contributes to the sperm toxicity of 1-bromopropane in mice. Biol Reprod 76(3):496–505

Genomes Project Consortium (2010). A map of human genome variation from population-scale sequencing. Nature 467(7319):1061–1073

Genomes Project Consortium (2012). An integrated map of genetic variation from 1092 human genomes. Nature 491:10

Genomes Project Consortium (2015b). An integrated map of structural variation in 2,504 human genomes. Nature 526(7571):75–81

Gertz M, Tsamandouras N, Säll C, Houston JB, Galetin A (2014) Reduced physiologically-based pharmacokinetic model of repaglinide: impact of OATP1B1 and CYP2C8 genotype and source of in vitro data on the prediction of drug-drug interaction risk. Pharm Res 31(9):2367–2382

Gharib SA, Loth DW, Soler Artigas M, Birkland TP, Wilk JB, Wain LV, Brody JA, Obeidat M, Hancock DB, Tang W, Rawal R, Boezen HM, Imboden M, Huffman JE, Lahousse L, Alves AC, Manichaikul A, Hui J, Morrison AC, Ramasamy A, Smith AV, Gudnason V, Surakka I, Vitart V, Evans DM, Strachan DP, Deary IJ, Hofman A, Glaser S, Wilson JF, North KE, Zhao JH, Heckbert SR, Jarvis DL, Probst-Hensch N, Schulz H, Barr RG, Jarvelin MR, O'Connor GT, Kahonen M, Cassano PA, Hysi PG, Dupuis J, Hayward C, Psaty BM, Hall IP, Parks WC, Tobin MD, London SJ, C. Consortium and C. SpiroMeta (2015) Integrative pathway genomics of lung function and airflow obstruction. Hum Mol Genet 24(23):6836–6848

Ginsberg G, Smolenski S, Neafsey P, Hattis D, Walker K, Guyton KZ, Johns DO, Sonawane B (2009) The influence of genetic polymorphisms on population variability in six xenobiotic-metabolizing enzymes. J Toxicol Environ Health B 12(5–6):307–333

Graf ME, Sookthai D, Johnson T, Schubel R, Katzke V, Bugert P, Hoffmeister M, Kaaks R, Kuhn T (2017). Biological reproducibility of circulating P-Selectin, Thrombopoietin, GPIIb/IIIa and Thrombomodulin over one year. Clin Biochem 50:942–946

Greene CS, Krishnan A, Wong AK, Ricciotti E, Zelaya RA, Himmelstein DS, Zhang R, Hartmann BM, Zaslavsky E, Sealfon SC, Chasman DI, FitzGerald GA, Dolinski K, Grosser T, Troyanskaya OG (2015) Understanding multicellular function and disease with human tissue-specific networks. Nat Genet 47(6):569–576

Gross-Davis C, Heavner K, Frank A, Newschaffer C, Klotz J, Santella R, Burstyn I (2015) The role of genotypes that modify the toxicity of chemical mutagens in the risk for myeloproliferative neoplasms. Int J Environ Res Public Health 12(3):2465–2485

GuhaThakurta D (2006) Computational identification of transcriptional regulatory elements in DNA sequence. Nucleic Acids Res 34(12):3585–3598

Harrill AM, McAllister KA (2017). New rodent population models may inform human health risk assessment and identification of genetic susceptibility to environmental exposures. Environ Health Perspect 125(8):086002

He Q, Cai T, Liu Y, Zhao N, Harmon QE, Almli LM, Binder EB, Engel SM, Ressler KJ, Conneely KN, Lin X, Wu MC (2016) Prioritizing individual genetic variants after kernel machine testing using variable selection. Genet Epidemiol 40(8):722–731

Hormozdiari F, Kang EY, Bilow M, Ben-David E, Vulpe C, McLachlan S, Lusis AJ, Han B, Eskin E (2016a) Imputing phenotypes for genome-wide association studies. Am J Hum Genet 99(1):89–103

Hormozdiari F, van de Bunt M, Segre AV, Li X, Joo JW, Bilow M, Sul JH, Sankararaman S, Pasaniuc B, Eskin E (2016b) Colocalization of GWAS and eQTL signals detects target genes. Am J Hum Genet 99(6):1245–1260

Ideker T, Ozier O, Schwikowski B, Siegel AF (2002) Discovering regulatory signalling circuits in molecular interaction networks. Bioinformatics 18(Suppl 1):S233–S240

Inoue S, Howgate EM, Rowland-Yeo K, Shimada T, Yamazaki H, Tucker GT, Rostami-Hodjegan A (2006) Prediction of in vivo drug clearance from in vitro data. II: potential inter-ethnic differences. Xenobiotica 36(6):499–513

International HapMap Consortium (2007). A second generation human haplotype map of over 3.1 million SNPs. Nature 449(7164):851–861

IOMC (2017). Revised guidance document on developing and assessing adverse outcome pathways. E. D. O. F. E. C.-O. A. DEVELOPMENT. Paris, OECD Environment, Health and Safety Publications

Jamei M, Dickinson GL, Rostami-Hodjegan A (2009a) A framework for assessing inter-individual variability in pharmacokinetics using virtual human populations and integrating general knowledge of physical chemistry, biology, anatomy, physiology and genetics: a tale of 'bottom-up' vs 'top-down' recognition of covariates. Drug Metab Pharmacokinet 24(1):53–75

Jamei M, Marciniak S, Feng K, Barnett A, Tucker G, Rostami-Hodjegan A (2009b) The Simcyp®Population-based ADME simulator. Expert Opin Drug Metab Toxicol 5(2):211–223

Johanson G, Jonsson F, Bois F (1999) Development of new technique for risk assessment using physiologically based toxicokinetic models. Am J Ind Med Suppl 1:101–103

Johnson TN, Rostami-Hodjegan A, Tucker GT (2006) Prediction of the clearance of eleven drugs and associated variability in neonates, infants and children. Clin Pharmacokinet 45(9):931–956

Judson RS, Mortensen HM, Shah I, Knudsen TB, Elloumi F (2012) Using pathway modules as targets for assay development in xenobiotic screening. Mol Biosyst 8(2):531–542

Kavlock RJ, Austin CP, Tice RR (2009) Toxicity testing in the 21st century: implications for human health risk assessment. Risk Anal 29(4):485–487

Krewski D, Westphal M, Andersen ME, Paoli GM, Chiu WA, Al-Zoughool M, Croteau MC, Burgoon LD, Cote I (2014) A framework for the next generation of risk science. Environ Health Perspect 122(8):796–805

Kumar KK, Pugalenthi G, Suganthan PN (2009) DNA-Prot: identification of DNA binding proteins from protein sequence information using random forest. J Biomol Struct Dyn 26(6):679–686

Lamb J, Crawford ED, Peck D, Modell JW, Blat IC, Wrobel MJ, Lerner J, Brunet JP, Subramanian A, Ross KN, Reich M, Hieronymus H, Wei G, Armstrong SA, Haggarty SJ, Clemons PA, Wei R, Carr SA, Lander ES, Golub TR (2006) The Connectivity Map: using

gene-expression signatures to connect small molecules, genes, and disease. Science 313(5795):1929–1935
Lee MY, Mukherjee N, Pakstis AJ, Khaliq S, Mohyuddin A, Mehdi SQ, Speed WC, Kidd JR, Kidd KK (2008) Global patterns of variation in allele and haplotype frequencies and linkage disequilibrium across the CYP2E1 gene. Pharmacogenomics J 8(5):349–356
Lee I, Blom UM, Wang PI, Shim JE, Marcotte EM (2011) Prioritizing candidate disease genes by network-based boosting of genome-wide association data. Genome Res 21(7):1109–1121
Liu F, Ichihara S, Mohideen SS, Sai U, Kitoh J, Ichihara G (2009) Comparative study on susceptibility to 1-bromopropane in three mice strains. Toxicol Sci 112(1):100–110
Lonsdale J, Thomas J, Salvatore M, Phillips R, Lo E, Shad S, Hasz R, Walters G, Garcia F, Young N, Foster B, Moser M, Karasik E, Gillard B, Ramsey K, Sullivan S, Bridge J, Magazine H, Syron J, Fleming J, Siminoff L, Traino H, Mosavel M, Barker L, Jewell S, Rohrer D, Maxim D, Filkins D, Harbach P, Cortadillo E, Berghuis B, Turner L, Hudson E, Feenstra K, Sobin L, Robb J, Branton P, Korzeniewski G, Shive C, Tabor D, Qi L, Groch K, Nampally S, Buia S, Zimmerman A, Smith A, Burges R, Robinson K, Valentino K, Bradbury D, Cosentino M, Diaz-Mayoral N, Kennedy M, Engel T, Williams P, Erickson K, Ardlie K, Winckler W, Getz G, DeLuca D, MacArthur D, Kellis M, Thomson A, Young T, Gelfand E, Donovan M, Meng Y, Grant G, Mash D, Marcus Y, Basile M, Liu J, Zhu J, Tu Z, Cox NJ, Nicolae DL, Gamazon ER, Im HK, Konkashbaev A, Pritchard J, Stevens M, Flutre T, Wen X, Dermitzakis ET, Lappalainen T, Guigo R, Monlong J, Sammeth M, Koller D, Battle A, Mostafavi S, McCarthy M, Rivas M, Maller J, Rusyn I, Nobel A, Wright F, Shabalin A, Feolo M, Sharopova N, Sturcke A, Paschal J, Anderson JM, Wilder EL, Derr LK, Green ED, Struewing JP, Temple G, Volpi S, Boyer JT, Thomson EJ, Guyer MS, Ng C, Abdallah A, Colantuoni D, Insel TR, Koester SE, Little AR, Bender PK, Lehner T, Yao Y, Compton CC, Vaught JB, Sawyer S, Lockhart NC, Demchok J, Moore HF (2013) The genotype-tissue expression (GTEx) project. Nat Genet 45(6):580–585
Mitra K, Carvunis AR, Ramesh SK, Ideker T (2013) Integrative approaches for finding modular structure in biological networks. Nat Rev Genet 14(10):719–732
Mooney MA, Wilmot B (2015) Gene set analysis: a step-by-step guide. Am J Med Genet B 168(7):517–527
Mortensen HM, Euling SY (2013) Integrating mechanistic and polymorphism data to characterize human genetic susceptibility for environmental chemical risk assessment in the 21st century. Toxicol Appl Pharmacol 271(3):395–404
NRC (2009). Science and decisions: advancing risk assessment. The National Academies Press, Washington, DC
NRC (2010). Toxicity-pathway-based risk assessment: preparing for paradigm change: a symposium summary. The National Academies Press, Washington, DC
NRC (2011a). Review of the environmental protection agency's draft IRIS assessment of formaldehyde. The National Academies Press, Washington, DC
NRC (2014a). Critical aspects of EPA's IRIS assessment of inorganic arsenic: interim report. The National Academies Press, Washington (DC)
NRC (2014b). Review of EPA's integrated risk information system (IRIS) process. The National Academies Press, Washington, DC
NRC (2016). Interindividual variability: new ways to study and implications for decision making: workshop in brief. The National Academies Press, Washington, DC
NRC (2017). Using 21st century science to improve risk-related evaluations. The National Academies Press, Washington, DC
Oki NO, Nelms MD, Bell SM, Mortensen HM, Edwards SW (2016) Accelerating adverse outcome pathway development using publicly available data sources. Curr Environ Health Rep 3(1):53–63
Patlewicz G, Simon TW, Rowlands JC, Budinsky RA, Becker RA (2015) Proposing a scientific confidence framework to help support the application of adverse outcome pathways for regulatory purposes. Regul Toxicol Pharmacol 71(3):463–477
Pearce RG, Setzer RW, Strope CL, Sipes NS, Wambaugh JF (2017). httk: R package for high-throughput toxicokinetics. J Stat Softw 79(4):1–26
Piñero J, Bravo À, Queralt-Rosinach N, Gutiérrez-Sacristán A, Deu-Pons J, Centeno E, García-García J, Sanz F, Furlong LI (2016) DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants. Nucleic Acids Res 45(D1):D833–D839
Pittman M, Edwards SW, Ives C, Mortensen HM (2017). AOP-DB: a database resource for the exploration of adverse outcome pathways through integrated association networks. Toxicol Appl Pharmacol. https://doi.org/10.1016/j.taap.2018.02.006
Polak S, Ghobadi C, Mishra H, Ahamadi M, Patel N, Jamei M, Rostami-Hodjegan A (2012) Prediction of concentration–time profile and its inter-individual variability following the dermal drug absorption. J Pharm Sci 101(7):2584–2595
Public Law 114–182 (2016). Frank R. Lautenberg Chemical Safety for the 21st Century Act U. Congres
Qi Y, Ge H (2006) Modularity and dynamics of cellular networks. PLoS Comput Biol 2(12):e174
Reyes-Gibby CC, Melkonian SC, Wang J, Yu RK, Shelburne SA, Lu C, Gunn GB, Chambers MS, Hanna EY, Yeung S-CJ, Shete S (2017) Identifying novel genes and biological processes relevant to the development of cancer therapy-induced mucositis: an informative gene network analysis. PLoS ONE 12(7):e0180396
Ring CL, Pearce RG, Setzer RW, Wetmore BA, Wambaugh JF (2017) Identifying populations sensitive to environmental chemicals by simulating toxicokinetic variability. Environ Int 106:105–118
Rostami-Hodjegan A (2012) Physiologically based pharmacokinetics joined with in vitro–in vivo extrapolation of ADME: a marriage under the arch of systems pharmacology. Clin Pharmacol Ther 92(1):50–61
Rostami-Hodjegan A, Tucker GT (2007) Simulation and prediction of in vivo drug metabolism in human populations from in vitro data. Nat Rev Drug Dis 6(2):140–148
Rotroff DM, Wetmore BA, Dix DJ, Ferguson SS, Clewell HJ, Houck KA, LeCluyse EL, Andersen ME, Judson RS, Smith CM, Sochaski MA, Kavlock RJ, Boellmann F, Martin MT, Reif DM, Wambaugh JF, Thomas RS (2010) Incorporating human dosimetry and exposure into high-throughput in vitro toxicity screening. Toxicol Sci 117(2):348–358
Sacks JD, Stanek LW, Luben TJ, Johns DO, Buckley BJ, Brown JS, Ross M (2010) Particulate matter-induced health effects: who is susceptible? Environ Health Perspect 119:446–454
Sager JE, Yu J, Ragueneau-Majlessi I, Isoherranen N (2015) Physiologically based pharmacokinetic (PBPK) modeling and simulation approaches: a systematic review of published models, applications, and model verification. Drug Metab Dispos 43(11):1823–1837
SDWA (1996). Safe Drinking Water Act Amendments of 1995: hearing before the Committee on Environment and Public Works, United States Senate, One Hundred Fourth Congress, first session, on S. 1316, a bill to reauthorize and amend Title XIV of the Public Health Service Act (commonly known as the Safe Drinking Water Act"), and for other purposes, October 19, 1995. Washington: For sale by the U.S. G.P.O. Supt. of Docs., Congressional Sales Office
Smalley JL, Gant TW, Zhang S-D (2010) Application of connectivity mapping in predictive toxicology based on gene-expression similarity. Toxicology 268(3):143–146

Snel B, Huynen MA (2004) Quantifying modularity in the evolution of biomolecular systems. Genome Res 14(3):391–397

The International HapMap Consortium (2005). A haplotype map of the human genome. Nature 437(7063):1299–1320

Threadgill DW, Churchill GA (2012) Ten years of the collaborative cross. G3 2:153–156

Timchalk C, Kousba A, Poet TS (2002) Monte Carlo analysis of the human chlorpyrifos-oxonase (PON1) polymorphism using a physiologically based pharmacokinetic and pharmacodynamic (PBPK/PD) model. Toxicol Lett 135(1–2):51–59

Tishkoff SA (2001) Haplotype diversity and linkage disequilibrium at human G6PD: recent origin of alleles that confer malarial resistance. Science 293(5529):455–462

Torkamani A, Topol EJ, Schork NJ (2008) Pathway analysis of seven common diseases assessed by genome-wide association. Genomics 92(5):265–272

US EPA (2002). Toxicological Review of Benzene (NonCancer Effects) (In Support of Summary Information of the Integrated Risk Information System (IRIS)). US EPA, Washington DC

US EPA (2014). TSCA Work Plan for Chemical Assessments: 2014 Update. O. o. P. P. a. Toxics

Venkatratnam A, Furuya S, Kosyk O, Gold A, Bodnar W, Konganti K, Threadgill DW, Gillespie KM, Aylor DL, Wright FA, Chiu WA, Rusyn I (2017). 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 158:48–62

Villeneuve DL, Crump D, Garcia-Reyero N, Hecker M, Hutchinson TH, LaLone CA, Landesmann B, Lettieri T, Munn S, Nepelska M, Ottinger MA, Vergauwen L, Whelan M (2014a) Adverse outcome pathway (AOP) development I: strategies and principles. Toxicol Sci 142(2):312–320

Villeneuve DL, Crump D, Garcia-Reyero N, Hecker M, Hutchinson TH, LaLone CA, Landesmann B, Lettieri T, Munn S, Nepelska M, Ottinger MA, Vergauwen L, Whelan M (2014b) Adverse outcome pathway development II: best practices. Toxicol Sci 142(2):321–330

Vinikoor-Imler LC, Owens EO, Nichols JL, Ross M, Brown JS, Sacks JD (2014) Evaluating potential response-modifying factors for associations between ozone and health outcomes: a weight-of-evidence approach. Environ Health Perspect 122(11):1166–1176

Wambaugh JF, Wetmore BA, Pearce R, Strope C, Goldsmith R, Sluka JP, Sedykh A, Tropsha A, Bosgra S, Shah I, Judson R, Thomas RS, Setzer RW (2015) Toxicokinetic triage for environmental chemicals. Toxicol Sci 147(1):55–67

Wetmore BA, Wambaugh JF, Ferguson SS, Sochaski MA, Rotroff DM, Freeman K, Clewell HJ 3rd, Dix DJ, Andersen ME, Houck KA, Allen B, Judson RS, Singh R, Kavlock RJ, Richard AM, Thomas RS (2012). Integration of dosimetry, exposure, and high-throughput screening data in chemical toxicity assessment. Toxicol Sci 125(1):157–174

Wetmore BA, Allen B, Clewell HJ, Parker T, Wambaugh JF, Almond LM, Sochaski MA, Thomas RS (2014) Incorporating population variability and susceptible subpopulations into dosimetry for high-throughput toxicity testing. Toxicol Sci 142(1):210–224

Wilson SH, Olden K (2004) The environmental genome project: phase i and beyond. Mol Interv 4(3):10

Yang C, Zhou X, Wan X, Yang Q, Xue H, Yu W (2011) Identifying disease-associated SNP clusters via contiguous outlier detection. Bioinformatics 27(18):2578–2585

Yates A, Akanni W, Amode MR, Barrell D, Billis K, Carvalho-Silva D, Cummins C, Clapham P, Fitzgerald S, Gil L, Girón CG, Gordon L, Hourlier T, Hunt SE, Janacek SH, Johnson N, Juettemann T, Keenan S, Lavidas I, Martin FJ, Maurel T, McLaren W, Murphy DN, Nag R, Nuhn M, Parker A, Patricio M, Pignatelli M, Rahtz M, Riat HS, Sheppard D, Taylor K, Thormann A, Vullo A, Wilder SP, Zadissa A, Birney E, Harrow J, Muffato M, Perry E, Ruffier M, Spudich G, Trevanion SJ, Cunningham F, Aken BL, Zerbino DR, Flicek P (2015) Ensembl 2016. Nucleic Acids Res 44(D1):D710–D716

Zaghloul NA, Katsanis N (2010) Functional modules, mutational load and human genetic disease. Trends Genet 26(4):168–176

Zerbino DR, Wilder SP, Johnson N, Juettemann T, Flicek PR (2015) The ensembl regulatory build. Genome Biol 16:56