Integrating mechanistic and polymorphism data to characterize human genetic susceptibility for environmental chemical risk assessment in the 21st century

Holly M. Mortensen
U.S. EPA, mortensen.holly@epa.gov

Susan Y. Euling
U.S. EPA

Contents lists available at ScienceDirect

Toxicology and Applied Pharmacology

journal homepage: www.elsevier.com/locate/ytaap

Integrating mechanistic and polymorphism data to characterize human genetic susceptibility for environmental chemical risk assessment in the 21st century ☆

Holly M. Mortensen [a,*], Susan Y. Euling [b]

[a] Office of Research and Development, US Environmental Protection Agency, National Center for Computational Toxicology, US EPA, 109 TW Alexander Dr., Mailcode B205-01, Research Triangle Park, NC 27711, USA
[b] Office of Research and Development, US Environmental Protection Agency, National Center for Environmental Assessment, US EPA, 1200 Pennsylvania Ave., NW, Mail Code 8623P, Washington, DC 20460, USA

ARTICLE INFO

Article history:
Received 5 October 2010
Revised 28 December 2010
Accepted 24 January 2011
Available online xxxx

Keywords:
Bioinformatics
Human genetic variation
Susceptibility
Toxicity pathway

ABSTRACT

Response to environmental chemicals can vary widely among individuals and between population groups. In human health risk assessment, data on susceptibility can be utilized by deriving risk levels based on a study of a susceptible population and/or an uncertainty factor may be applied to account for the lack of information about susceptibility. Defining genetic susceptibility in response to environmental chemicals across human populations is an area of interest in the NAS' new paradigm of toxicity pathway-based risk assessment. Data from high-throughput/high content (HT/HC), including -omics (e.g., genomics, transcriptomics, proteomics, metabolomics) technologies, have been integral to the identification and characterization of drug target and disease loci, and have been successfully utilized to inform the mechanism of action for numerous environmental chemicals. Large-scale population genotyping studies may help to characterize levels of variability across human populations at identified target loci implicated in response to environmental chemicals. By combining mechanistic data for a given environmental chemical with next generation sequencing data that provides human population variation information, one can begin to characterize differential susceptibility due to genetic variability to environmental chemicals within and across genetically heterogeneous human populations. The integration of such data sources will be informative to human health risk assessment.

Introduction

Differential response to an environmental chemical exposure is due to interactions between heredity and environment. Here we define *susceptibility* as the increased likelihood of an adverse effect in relation to a human subpopulation factor such as life stage, demographic feature, or genetic characteristic (EPA, 2005). Multiple intrinsic and extrinsic susceptibility factors, including age, sex, ethnicity, nutrition status, and lifestyle, contribute to human variability in biological response to environmental agents. Because of these multiple interacting factors, the identification of causal factors responsible for the variable response for even a single environmental agent is often unclear given the available human data, and underlines the need for a better mechanistic understanding of the risk factors that result in variable response to environmental chemicals in humans.

When available, information on human genetic variation can be used in human health risk assessment qualitatively and/or quantitatively to characterize variability and susceptibility differences in response to a chemical exposure. For example, available information could be used to estimate a risk level based on a study of a susceptible population, and/or contribute to the information and rationale for selecting the uncertainty factor (UF), thereby accounting for uncertainty in susceptibility among human populations. One example of accounting for intraspecies uncertainty, as well as other uncertainties in risk assessment, is the Environmental Protection Agency's (EPA) Integrated Risk Information System (IRIS). IRIS accounts for uncertainty in the risk assessment process by using various uncertainty (or

Abbreviations: ACToR, Aggregated Computational Toxicology Resources; dbSNP, Single Nucleotide Polymorphism Database; CNV, Copy Number Variant; EGP, NIEHS-Environmental Genome Project; EWAS, Environment-Wide Association Study; GWAS, Genome-Wide Association Study; HapMap, International Haplotype Map (HapMap) Project; HT/HC, High-throughput/High content; HTS, High-throughput screening; IRIS, Integrated Risk Information System; MOA, Mode of action; NAS, National Academy of Sciences; NCBI, National Center for Biotechnology; NIEHS, National Institute of Environmental Health Sciences; NR, Nuclear Receptor; -omics, neologism relating multiple fields in biology that end in -omics (e.g., genomics, transciptomics, proteomics, metabolomics); PDR, Polymorphism Discovery Resource; RfC, Reference Concentration; RfD, Reference Dose; SNP, Single Nucleotide Polymorphism; TD, Toxicodynamics; TD2, Diabetes Type 2; TK, Toxicokinetics; ToxCast, EPA's chemical prioritization research program; TSS, Transcription Start Site; UF, Uncertainty/Variability Factor; US EPA, US Environmental Protection Agency.

☆ Disclaimer: This manuscript has been reviewed by the U.S. Environmental Protection Agency and approved for publication. The views expressed in this manuscript are those of the authors and do not necessarily reflect the views or policies of the U.S. Environmental Protection Agency.

* Corresponding author at: National Center for Computational Toxicology, Research Triangle Park, NC 27711, USA. Fax: +1 919 541 1194.
E-mail address: mortensen.holly@epa.gov (H.M. Mortensen).

0041-008X/$ – see front matter © 2011 Published by Elsevier Inc.
doi:10.1016/j.taap.2011.01.015

variability) factors (UFs) (for further information, see http://www.epa.gov/iris/), including the intraspecies UF that is applied when there is a lack of data about population susceptibility or uncharacterized variation in chemical response among humans. Note that the intraspecies UF is not specifically accounting for *genetic* differences, but instead is accounting for all types of intraspecies differences that could contribute to an increased susceptibility. Limitations to defining susceptibility within the current risk assessment approach can include data gaps on human susceptibility (e.g., genetic susceptibility, lack of understanding of the relationships among different susceptibility factors) for many chemicals and inconsistency in applying the intraspecies uncertainty factor.

Two recent reports from the National Academy of Science (NAS), *Science and Decisions* (National Research Council, 2009) and *Toxicity Testing in the 21st Century* (National Research Council, 2007), are part of a concerted effort to review and overhaul the current risk assessment process, and thereby improve US EPA human health risk assessments. These two NAS reports underscore two needs for improving toxicity testing and risk assessment of environmental chemicals: 1) increasing mechanistic information; and 2) characterizing human susceptibility. *Toxicity Testing in the 21st Century* (National Research Council, 2007) presents a future vision of a toxicity pathway-based toxicity testing that measures effects on the mechanism of action as the underlying basis. The goal is to develop an understanding of human biological mechanism for environmental chemicals by using appropriate human *in vitro* toxicity pathway-based assays, in combination with mechanistic computational systems biology models. The methods to be utilized in the new paradigm (National Research Council, 2007) have several advantages over current assays used in risk assessment including increased sensitivity and measurement of precursor effects, increased number of chemicals with available toxicity data (based on increased analysis efficiency, increased cost-effectiveness and decreased animal use), and measurement of effects in humans thereby eliminating cross-species extrapolation issues. Disadvantages include the limited availability of *in vitro* and *in vivo* extrapolation methods in humans (i.e., establishing valid assays with known causal relationships between chemical exposure, *in vitro* response, and phenotypic outcome in humans), and the lack of a concrete path forward to achieve the 21st century pathway-based approach. One important component of the NAS' strategic plan is “the use of toxicity pathways information in risk assessment,” although bridging the gap between current *in vivo* toxicity-based to pathway-based approaches remains an unscripted challenge. *Science and Decisions* (National Research Council, 2009) provides a number of recommendations for the overall risk assessment process. Regarding susceptibility, the report states that variability in human susceptibility “has not received sufficient or consistent attention…” and recommends that uncertainty and variability be characterized and communicated in the quantitative steps of EPA risk assessments.

To address the need to characterize human genetic susceptibility for informing human health risk assessment, we suggest a general approach for integrating two existing data sources: 1) mechanistic data implicating chemical–gene targets and 2) population polymorphism data characterizing genes implicated in chemical target pathways. In order to outline the approach of combining mechanistic data with human polymorphism data to define human susceptibility, in this article we review the latest genomic and informatic approaches for chemical target identification and characterization of mechanism of action, and discuss how genotype data from dense sequencing projects can be combined with mechanistic data to inform risk assessment about human susceptibility to environmental chemicals. Lastly, we discuss future needs and directions in light of the NAS' recommendations for improving toxicity testing and using information about variability and susceptibility in human health risk assessment of environmental chemicals.

Data source I: Chemical mechanistic information

One obstacle in moving toward a human pathway-based toxicity testing paradigm is the obvious lack of mechanism of action (see Fig. 1) data for humans. We currently have limited knowledge of the effects of environmental chemicals on published molecular pathways and in related human disease etiology, including knowledge of key events in human toxicity (National Research Council, 2010). In addition, there is little known about the relationship among events that result in toxicity in humans, causal steps in the continuum between exposure and pathway perturbation, and how differences at the individual or population level manifest in disease. The identification of potential chemical targets is the first step in elucidating key molecular interactions that trigger alterations in toxicity (Krewski et al., 2010), and is a major research priority. For a shift to toxicity pathway-based risk assessment to be realized, there is a need to characterize human toxicity pathways for multiple, integrated, and diverse biological processes in humans. The identification of the molecular targets in humans that, when perturbed, define the key events in a pathway or process that is implicated in a probable adverse outcome would be a critical step forward in understanding chemical toxicity related processes (Mortensen et al., 2010) and susceptibility in human disease.

Several systems-based approaches have recently been developed in order to better understand chemical–gene effects, characterize mechanism of toxicity in various toxicological contexts, and identify the genetic loci implicated in toxicity (Hamadeh et al., 2002a,b; Butcher et al., 2004; Goh et al., 2007; Edwards and Preston, 2008; Davis et al., 2009; Linghu

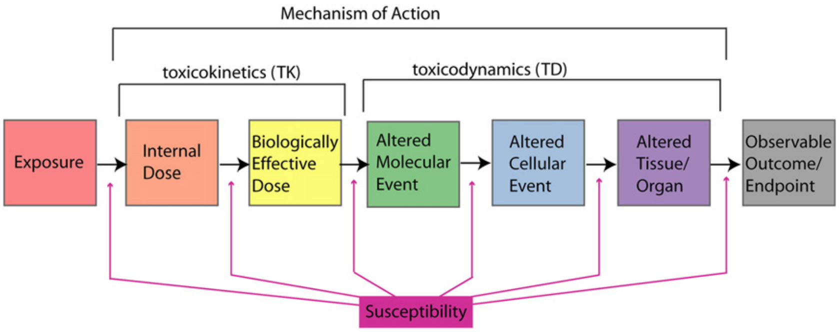

Fig. 1. Illustration of mechanism of action. Black arrows indicate the sequence of events in mechanism of action. Purple arrows indicate potential steps in the continuum leading to observable disease outcome in humans where susceptibility could influence the outcome. Mechanism of action is defined here as all of the steps, including toxicodynamic (TD) and toxicokinetic (TK), between chemical exposure and the final outcome at the level of the whole organism. Variation within humans could affect the resulting outcome at any TD and/or TK step(s), indicated by the different boxes (adapted from Eubanks, 1994; Guyton et al., 2008; Jayapal et al., 2010).

et al., 2009; Wiegers et al., 2009). Fig. 2a–c illustrates how consolidation of information on *chemical–gene association* can be extended to a pathway-level understanding by identifying disrupted target loci through empirical studies. The integration of available information on chemical–gene-pathway–outcome relationships provides information about the putative or established mechanism of action, where target loci are identified (Fig. 2a,b). Differing information may be available for non-human animal species thus providing for or limiting the extrapolation of pathway effects to humans. Further, the chemical may affect multiple, distinct or overlapping biological pathways within a single species. The association of chemical, gene, and pathway does not directly implicate a perturbation of a gene in a pathway, but highlights where empirical evidence supports an association. Similar to candidate disease gene studies, this list of chemical target loci can be expanded based on inclusion of related loci or reduced based on loci thought to be in chromosomal linkage with a candidate loci (Fig. 2b,c). Two principal approaches, direct experimental methods and informatic integration, can contribute to the elucidation of chemical molecular mechanism of action as described in Fig. 2a–b. These methods and some considerations involved in their applications are described below.

Direct experimental approaches

Both *in vivo* and *in vitro* direct experimental sources contribute to understanding of chemical mechanism of toxicity and identification of genetic loci implicated in toxicity. *In vivo* model systems are the backbone of current toxicology testing of environmental chemicals, and data from *in vivo* animal experiments is often used in the current

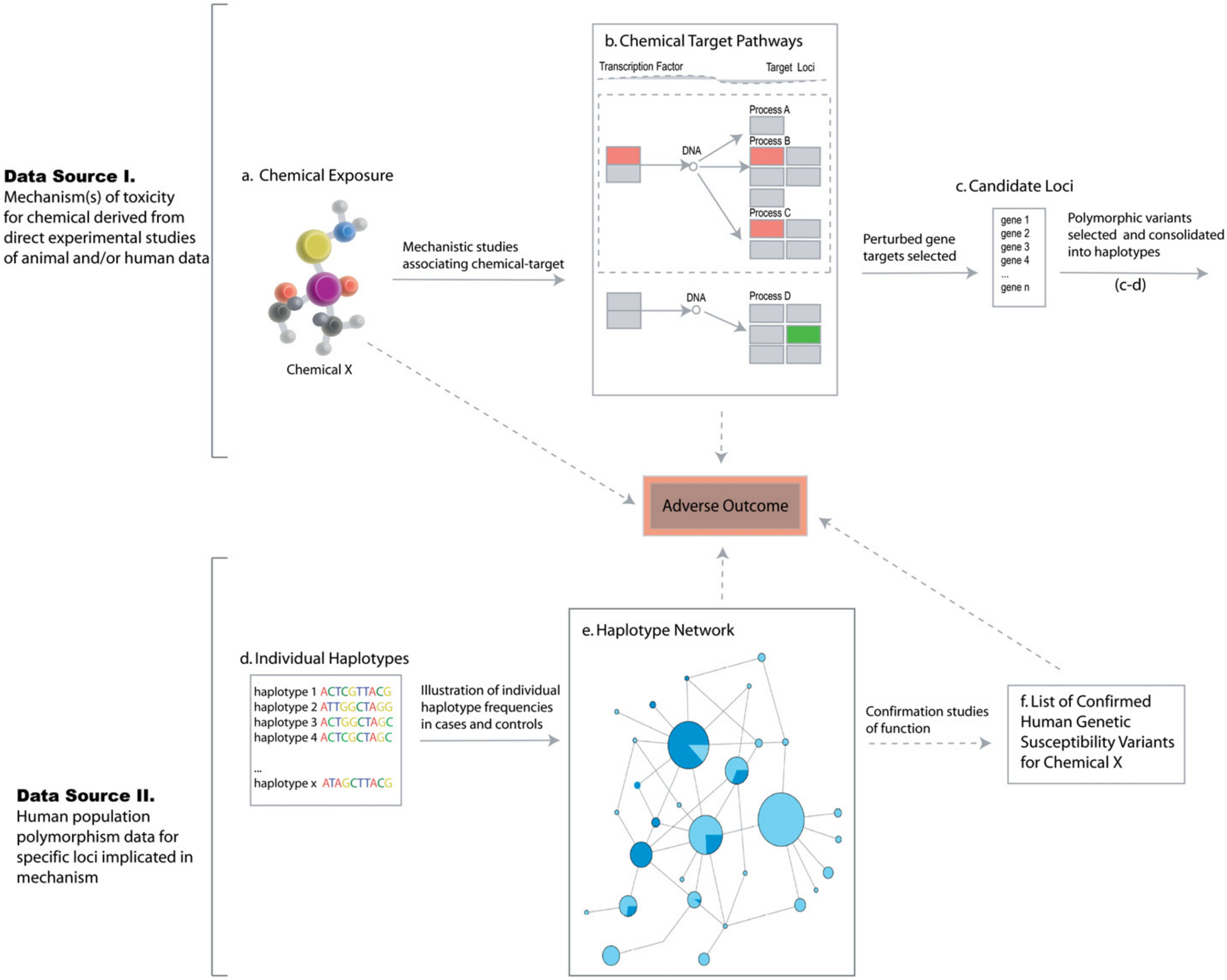
Data Source I.
Mechanism(s) of toxicity for chemical derived from direct experimental studies of animal and/or human data
a. Chemical Exposure
Chemical X
Mechanistic studies associating chemical-target
b. Chemical Target Pathways
Transcription Factor
Target Loci
DNA
Process A
Process B
Process C
DNA
Process D
Perturbed gene targets selected
c. Candidate Loci
gene 1
gene 2
gene 3
gene 4
...
gene n
Polymorphic variants selected and consolidated into haplotypes
(c-d)
Adverse Outcome
Data Source II.
Human population polymorphism data for specific loci implicated in mechanism
d. Individual Haplotypes
haplotype 1 ACTCGTTACG
haplotype 2 ATTGGCTAGG
haplotype 3 ACTGGCTAGC
haplotype 4 ACTCGCTAGC
...
haplotype x ATAGCTTACG
Illustration of individual haplotype frequencies in cases and controls
e. Haplotype Network
Confirmation studies of function
f. List of Confirmed Human Genetic Susceptibility Variants for Chemical X

Fig. 2. Overview of the proposed process for the genetic characterization of variable response to environmental chemicals. Horizontal arrows indicate manipulation of data at various stages of analysis. a, b. Data Source I includes direct experimental data, associating chemical X with molecular target(s) (e.g., gene or protein), obtained from animal toxicology, human epidemiology, or *in vitro* methods. Data sources for mechanistic information for chemical X can be obtained from publically available resources (Table 1). Mechanistic information is illustrated using a pathway-based framework. Red and green boxes indicate genetic targets affected by chemical X, where only RED target modification is associated with chemical X and the adverse outcome. c. Candidate or target loci can be characterized for variation at the nucleotide level. d. Data Source II is human genetic data, at the nucleotide level. SNP haplotypes can be constructed from SNPs of interest. Use of tagged SNPs (i.e., representative SNPs from a region of the genome with high linkage disequilibrium) can be used to limit redundant information. There are currently insufficient methods to infer function with confidence and functional effect of particular SNPs may not be known at this stage. e. Cases and controls for a particular adverse outcome can be illustrated in terms of observed haplotype frequencies using haplotype networks, defining susceptible haplotypes. Cases can be further classified based on population association to inform the risk assessor of susceptible group (haplotype network adapted from Mortensen, 2008; Mortensen et al., 2011). f. Following susceptibility haplotype identification, confirmatory studies of function may be needed. The effect of molecular variation on functional phenotype can be assessed using clinical, epidemiological or computational inference-based methods like those listed in Table 2.

Please cite this article as: Mortensen, H.M., Euling, S.Y., Integrating mechanistic and polymorphism data to characterize human genetic susceptibility for environmental chemical risk assessment in the 21st century, Toxicol. Appl. Pharmacol. (2011), doi:10.1016/j.taap.2011.01.015

human health risk assessment paradigm. Animal models have long been used in the discovery of novel candidate genes in disease studies (LePage and Conlon, 2006). Many genes implicated in human disease susceptibility have been characterized using a comparative approach utilizing animal models of human disease (Moore, 1999; Young, 2001; Phillips et al., 2002; Ewart-Toland and Balmain, 2004; Simmons, 2008). Conservation of function among organisms is pivotal to the use of traditional toxicological testing assays (National Research Council, 2000), as well as the basis for most research into the genetics of human complex traits and disease (Boguski, 2002). The biological mechanisms known to be associated with disease were identified in many cases because they were highly conserved across species. Conversely, the disruption of function of conserved genetic elements can potentially lead to radically different phenotypic outcomes in different species (McGary et al., 2010). An example relevant to both pharmacology and toxicology is the phylogenetically diverse nuclear receptor (NR) peroxisome proliferator-activated receptor alpha (PPARα). Nuclear receptors control transcription through binding to promoter regions of DNA. PPARα, like other NRs, exhibits many of the species differences in NR structure and function observed between rodents and humans. Species differences in response to PPARα agonists have been attributed to differences in sequence similarity in the DNA response elements, differences in hepatic expression level (i.e., high expression in rodents in comparison to humans in response to many substrates), hepatic microRNA (miRNA) expression, and the presence of PPAR splice variants in humans (Miller and Willson, 2001; Klaunig et al., 2003; Gonzalez and Shah, 2008; Peters et al., 2005; Peters et al., 2008). Molecular variation at this level may affect differences in the downstream pathway, which manifest in disease in one species but not necessarily another. Thus, one limitation to the use of animal assays for identifying human toxicity pathways/mechanisms of action is uncertainty about the human mode or mechanism of toxicity, because the evolutionary conservation of a biological process cannot necessarily be assumed.

In vitro methods that are used in the characterization of chemical mechanism, including -omics methods, focus on understanding complex systems (e.g., genes and proteins). These approaches complement traditional *in vivo* approaches to target identification and elucidation of mechanism in humans. Through the integration of -omics technologies, an analysis of changes at the molecular level for multiple interacting systems (or at the level of the genome) becomes feasible and may provide a means for predicting toxicity in less time than performing classical toxicological studies (Butcher et al., 2004). Analysis methods that contribute to defining chemical mechanism of action include, but are not limited to, *microarrays* that measure gene or protein expression in different cells or tissues under various conditions, *high throughput screening (HTS) assays* that include cell-based systems, and *ex vivo* tissue methods that do not require a culture medium and allow experimentation in or on a tissue outside of an organism in a simulated environment. Study design choices (e.g., dose, dose–response, duration of exposure, route of exposure, developmental timing) are important to consider when utilizing studies, *in vitro* and *in vivo*, that will contribute to a picture of a chemical's mechanism of action for an outcome of interest. For

Table 1
Open source molecular toxicology resources and databases.

Resource	Description	URL
ACToR—Aggregated Computational Toxicology Resources	Compiles publically available data collections on environmental chemicals, including EPA in vivo guideline study results (ToxRefDB) and chemical activity data (ToxCast).	http://www.actor.epa.gov
Array Express	Repository for functional genomic experiments, focused on expression studies, data mining, comparison and analysis tools.	http://www.ebi.ac.uk/microarray-as/ae/
CEBS—Chemical Effects in Biological Systems	Data integration tool for integrating data relating to environmental health, pharmacology, and toxicology.	http://www.niehs.nih.gov/research/resources/databases/cebs/index.cfm
Connectivity Map	Collection of genome-wide transcriptional expression data from cultured human cells treated with bioactive small molecules.	http://www.broadinstitute.org/cmap/
CTD—Comparative Toxicogenomic Database	Elucidates molecular mechanisms by which environmental chemicals affect human disease.	http://www.ctd.mdibl.org/
GAD—Genetic Association Database	Archive of human genetic association studies of complex diseases and disorders.	http://www.geneticassociationdb.nih.gov/
GEO—Gene Expression Omnibus	Repository for a wide range of high-throughput experimental data, with query and analysis functionality.	http://www.ncbi.nlm.nih.gov/geo/
GO—Gene Ontology	Describes gene products in terms of their associated biological processes, cellular components and molecular functions in a species-independent manner.	http://www.geneontology.org/
GSEA—Gene Set Enrichment Analysis	Determines whether a set of genes shows concordant differences between two biological states or phenotypes.	http://www.broadinstitute.org/gsea/
KEGG—Kyoto Encyclopedia of Genes and Genomes	Integrated database resource consisting of 16 main databases, broadly categorized into systems information, genomic information, and chemical information.	http://www.genome.jp/kegg/
Pathway Commons	Collection of publicly available pathways from multiple organisms.	http://www.pathwaycommons.org/pc/home.do
pCEC—Profiles of Chemical Effects on Cells	Stores and handles gene expression profiling information and the categorization of toxicity data, and separates chemicals into a variety groups by the type of influence.	http://www.project.nies.go.jp/eCA/cgi-bin/index.cgi
PharmaGKB—The Pharmacogenomics Knowledge Base	Database of primary genotype and phenotype data, annotated gene variants and gene–drug–disease relationships obtained from literature review. Also includes summaries of important pharmacogenetic genes and drug pathways.	http://www.pharmgkb.org/
Reactome	A curated knowledgebase of biological pathways.	http://www.reactome.org/
STRING—Search Tool for the Retrieval of Interacting Genes/Proteins	Known and predicted protein–protein interaction database.	http://www.string.embl.de
SuperToxic	Predicts the toxicity of compounds, informs potential targets in biochemical pathways, and shows potential binding partners.	http://www.bioinformatics.charite.de/supertoxic/
ToxExpress	Database of integrated toxicity-based gene expression and classic toxicology endpoints.	http://www.genelogic.com/knowledge-suites/toxexpress-program
ToxRTool—Toxicological Data Reliability Assessment Tool	Evaluates quality and reliability of *in vivo* and *in vitro* toxicological data.	http://www.ecvam.jrc.ec.europa.eu
TOXNET—Toxicology Data Network	Database of toxicology, hazardous chemicals, environmental health, and toxic releases.	http://www.toxnet.nlm.nih.gov/

Please cite this article as: Mortensen, H.M., Euling, S.Y., Integrating mechanistic and polymorphism data to characterize human genetic susceptibility for environmental chemical risk assessment in the 21st century, Toxicol. Appl. Pharmacol. (2011), doi:10.1016/j.taap.2011.01.015

example, mechanistic information, for a chemical and a particular developmental outcome, from toxicity studies that exposed during a critical window of development may be more meaningful than adult exposure studies.

Informatic integration

Multiple efforts have been initiated to consolidate diverse datasets of chemical toxicity, gene expression, and pathway level biological processes from *in vitro* animal and human data, and *in vivo* studies when available (Table 1). Database and knowledgebase approaches, integrating multiple data types, make it possible to look for associations and use statistical methods to look for correlation among the different data sources. Database sources differ in a number of aspects, such as hand curated or automated (for discussion, see Judson, 2010). It is clear that available data must be integrated to come to an overall conclusion on the toxicity of a chemical, as well as understand the mechanisms of chemical toxicity (Mattingly, 2009). One example of a project that integrates toxicity data sources is the US EPA web-based resource, the Aggregated Computational Toxicology Resource (ACToR), which houses publicly available data on chemical identity, structure, and physical properties (Judson et al., 2008, 2009). ACToR also houses data from the ToxCast high throughput screening program, with *in vitro* assay data profiles for what will soon be 1000 environmental chemicals (http://www.epa.gov/comptox/toxcast/, (Dix et al., 2007), as well as *in vivo* toxicology data from ToxRefDB (Knudsen et al., 2009; Martin et al., 2009a,b). Data consolidated within ACToR have been used to create bioactivity profiles based on *in vitro* chemical response data and machine learning predictive signatures to infer phenotypic outcomes (Judson et al., 2010). Methods, such as these, can be used to associate chemicals with gene targets that may be important in particular *in vivo* phenotypic outcomes.

Data within ACToR have also been integrated with human gene, pathway, disease, and drug target information from multiple data sources, though not yet publicly available, for use in association mapping and toxicity inference (Mortensen et al., 2009, 2010). Association between chemical exposure, genes, and disease can be used in the absence of mechanistic information in humans. The systems biology approach advocated here ultimately utilizes the association between drug or environmental chemical exposure and disease, the human "phenotypic anchor," in the absence of direct evidence of mechanism of toxicity. A number of recent informatic approaches implementing molecular networks have been developed that tackle the problem of identifying putative toxicity pathways by using gene enrichment, chemical and disease association, and drug target endpoints (Goh et al., 2007; Gohlke et al., 2009; Linghu et al., 2009; Mortensen et al., 2009, 2010; Patel et al., 2010). Though these types of association methods are supported by empirical, often curated, knowledge, the mechanisms of action are typically not well known, making them most useful for prioritization efforts for further testing and hypothesis generation.

For genetic susceptibility to environmental chemicals to be characterized there is a broad need for a more detailed description, beyond association and weight of evidence approaches, to understanding the mechanisms of toxicity. Efforts have begun to consolidate what is currently known about various chemical mechanisms of action for both humans and other species into a useable database platform that can be easily integrated with chemical data found within ACToR (Mortensen and Judson, *unpublished results*). The novel approach of Sone et al. (2010) integrates data from the TOXLINE and IRIS databases to manually define putative mechanisms of action categories. These authors plan to expand their mechanisms of action classification into a larger HEALS (Health Effects Alert System) database (Sone, personal communication). Future efforts involving the integration of diverse data types poses several challenges to the field of bioinformatics; however, the consolidation of mechanistic information will allow for a greater understanding of system and pathway level changes related to toxicity and subsequently for the identification and characterization of molecular factors that contribute to human susceptibility.

Data source II: Sequencing projects to characterize human variation

Until recently genetic studies in humans were primarily focused on one or a few candidate genes to investigate disease–gene phenotype associations. This type of study surveys gene sequence variation at candidate loci across multiple individuals or populations. Genetic information related to candidate loci is consolidated into haplotypes, or combinations of polymorphisms at the nucleotide level (e.g., Single Nucleotide Polymorphisms (SNPs), Copy Number Variants (CNVs)) across an individual chromosome (refer to Fig. 2d). The frequencies of individual variants are then compared between the individuals representing cases and controls for a disease phenotype of interest (Fig. 2e), or alternatively from populations or groups that exhibit differing disease phenotypes. From this, the individual haplotypes associated with a disease can be identified within and between groups, identifying potentially susceptible or at risk populations. Further testing can confirm the precise functional genetic variant(s) responsible for an association with a disease phenotype (Fig. 2f). The effect of particular molecular variants on functional phenotype can be assessed using clinical (Tishkoff et al., 2007) and molecular epidemiological-based methods (Spitz and Bondy, 2010), and computational inference tools (see Table 2 for open source functional inference tools).

Ideally, for a disease of interest, there is *a priori* information about the genetic basis of the disease, based on conspicuous determinants typically identified in genetic epidemiological studies. Classic successes that have identified the genetic determinants of disease phenotypes include Huntington's disease, Alzheimer's disease, and some forms of breast cancer (Risch and Merikangas, 1996). In the past 10 years, the explosion of genomic information, bioinformatics tools, and open access to large datasets has revolutionized research on complex traits, including common human diseases. With the sequencing of the human genome and next-generation sequencing technologies, the genetic variation responsible for common disease phenotypes can be explored more readily. Large-scale studies of human genetic variation, such as the International HapMap Consortium, have focused on the identification of the underlying genetic variation in and across human populations, the structure of that variation across the genome, and its link to phenotype. The implementation of dense sequencing data thus far has been largely in characterizing disease risk and pharmaceutical response. Other applications of the HapMap data have provided information about human study design, population structure and statistical issues (Clark et al., 2005; Clayton et al., 2005; Goldstein, 2009), the characterization of natural selection (Nielsen et al., 2005; Voight et al., 2006; Sabeti et al., 2007), and the characterization of regions of linkage and recombination (Myers et al., 2005; Weir et al., 2005). However, inquiry into common diseases has paved the way for similar approaches to understanding differential human response to environmental chemicals. With identified chemical targets, it is possible to characterize variation in chemical response, with the aim of informing human genetic susceptibility to environmental chemicals. Following the HapMap efforts, several other deep sequencing projects have been initiated and have specific objectives with the potential to increase our understanding of human susceptibility to environmental chemicals (Table 2).

Identification of the 'Environmental Genome'

Shortly after the first completed resequencing of the human genome, a collaborative effort between the National Institute of

Table 2
Open source sequence annotation resources and phenotype inference tools.

Resource	Description	URL
1000 Genomes—A Deep Catalog of Human Genetic Variation	Extensive public catalog of human genetic variation, including SNPs and structural variants, and their haplotype contexts.	http://www.1000genomes.org/page.php?page=data
ANNOVAR	Functional annotation of HT sequence data with reported SNPs from dbSNP, MAF>1% from the 1000 Genome Project, or subset of non-synonymous SNPs with SIFT score >0.05.	http://www.openbioinformatics.org/annovar/
DAS—Distributed Annotation System	Gathers genome annotation information from multiple web sites, collates the information, and displays it to the user in a single view.	https://www.sanger.ac.uk/Software/analysis/das/
dbSNP—Single Nucleotide Database	Public archive of genetic variation, within and across different species, containing characterized and uncharacterized molecular variations.	http://www.ncbi.nlm.nih.gov/projects/SNP/snp_summary.cgi
EGP—Environmental Genome Project	Resequencing of 600 environmentally relevant genes thought to play a role in susceptibility to environmental exposures in a panel of 95 individuals representing the ethnic diversity found in the United States.	http://www.niehs.nih.gov/research/supported/programs/egp/
ENCODE—Encyclopedia of DNA Elements	Identifies functional elements in the human genome sequence.	http://www.genome.gov
GeneSNPs Database	Publicly available graphical display of the newly identified SNPs discovered through resequencing environmentally responsive genes (EGP).	http://www.genome.utah.edu/genesnps/
GREAT—Genomic Regions Enrichment of Annotations Tool	Assigns biological meaning to a set of non-coding genomic regions by analyzing the annotations of the nearby genes.	http://www.great.stanford.edu.
NHGRI GWAS Catalog	Online catalog of GWAS publications that assay at least 100,000 SNPs; SNP-trait associations are indicated.	http://www.genome.gov/gwastudies
PolyPhen— Polymorphism Phenotyping	Predicts impact of an amino acid substitution on the structure and function of a human protein using physical and comparative (species homology) considerations.	http://www.genetics.bwh.harvard.edu/pph/
Seattle SNPs	Focused on identifying, genotyping, and modeling the associations between SNPs in candidate genes and pathways that underlie inflammatory responses in humans.	http://www.pga.gs.washington.edu/finished_genes.html
SIFT—Sorting Intolerant From Tolerant	Predicts whether an amino acid substitution affects protein function based on sequence homology and the physical properties of amino acids.	http://www.sift.jcvi.org/
SNPnexus	Database designed to simplify and assist in the selection of functionally relevant SNPs for large-scale genotyping studies of multifactorial disorders.	http://www.snp-nexus.org/index.html

Environmental Health Sciences (NIEHS) and the University of Washington, the Environmental Genome Project (EGP) (Wilson and Olden, 2004), was formed with the goal of implementing sequencing technology to identify loci that contribute greater than average to human susceptibility to environmental agents. An initial list of environmentally responsive genes was identified, and gene targets were selected for testing including loci implicated in DNA repair, cell cycle control, drug metabolism, apoptosis, and cell differentiation and signal transduction mechanisms. Systematic resequencing of these genes was conducted in 95 human individuals from the polymorphism discovery resource (PDR) panel (Collins et al., 1998) that represents the ethnic diversity present in the US. During phases 1 and 2 of this study, the EGP identified 92,486 SNP variants within 647 environmentally responsive candidate genes. Based on the hypothesis that individuals with functionally significant polymorphisms within environmentally responsive gene regions may be particularly susceptible to genotoxic environmental agents, SNPs were selected for population screening across both coding and non-coding regions of these loci (Livingston et al., 2004). These data are publicly available and housed in the GeneSNPs database (Table 2). In 2006, HapMap and NCBI dbSNP updates were added to the GeneSNPs database, allowing extension of the data to include validated genotypes and calculate population specific haplotypes. Most studies that have analyzed data from this important GeneSNPs database have focused on human population genetics (e.g., changes in allele frequencies associated with evolutionary processes such as natural selection, genetic drift, mutation and gene flow) rather than individual or population level susceptibility across the genome (Gutenkunst et al., 2009; Ionita-Laza et al., 2009; Lohmueller et al., 2010). The EGP has also been used to validate other extensive human genome sequencing projects (Nickerson, personal communication) such as the 1000 Genomes Project (Table 2). Because these data include environmentally responsive genes, many of which are the targets of environmental agents or implicated in biological processes involving chemical response, these data could be useful in establishing baseline levels of variation across human population groups for the 'environmental response genome'.

Predicting functional consequences

Over 64 million distinct genetic variants are currently available in the public domain, for multiple organisms including *Homo sapiens* (dbSNP, build 131, see Table 1). However, only a small fraction of these variants have been characterized in terms of their function, occurrence in relevant populations (allele frequency and population variation), or association with a phenotype or disease. The HapMap Project has contributed to much of our current understanding of underlying genetic variation in diverse humans populations and has elucidated many of the variants associated with common disease phenotypes (e.g., diabetes, obesity, breast cancer (Manolio and Collins, 2009)). SNP variants, and more recently structural variation (Eichler et al., 2007), have been investigated to understand the pattern and nature of differences within the human genome, and their functional effect.

The primary goal for genetic association studies is to identify the functional variants associated with a phenotype of interest. Prioritization using gene and protein network properties has also been informative in the identification and selection of candidate disease genes (Goh et al., 2007; Chen et al., 2009; Linghu et al., 2009; Nitsch et al., 2010), and similar network methods have been described to identify and rank order potential chemical target loci to investigate based on biological relationship (Mortensen et al., 2010). When the mechanism of action is not well-characterized in humans, network methods are useful in identifying loci related to a phenotype or adverse outcome by using a chemical–gene association approach, as described previously (see Fig. 2a–c). With a list of identified chemical targets and a comprehensive catalogue of common variants in the human population, the researcher is then faced with the question of how to most effectively identify potentially causative susceptibility variants. One general approach for identifying informative SNPs is to compare correlation among SNPs of similar frequency (referred to as linkage disequilibrium). Additionally, in looking for functional effects of genetic variants, exons, regulatory regions, and conserved noncoding regions (CNS) among species are obvious regions to scan for functional polymorphisms.

Studies of the heritability of gene expression indicate an abundance of *cis*- regulatory variation in the human genome and lack of *trans*- effects, suggesting that regulatory variation may be the primary effect contributing to phenotypic variation in humans (Kudaravalli et al., 2009) as well as a target of positive natural selection (Kudaravalli et al., 2009). Whether variants of functional significance map to exon regions and near transcription start sites (TSS) only, or exist within intronic regions has been an area of contention. Stranger et al. (2007) found that the vast majority of detected *cis*- regulatory effects map very close to the TSS, and additionally note that these regions are enriched for regions that are highly conserved (between species), suggesting that most of the large effect variants will be in genic and immediate intergenic regions.

When a region of interest is identified, there is concern that SNP characterization in publically available data may not be dense enough to discover all functional variants. A recent study by Ionita-Laza et al. (2009) addressed this issue by asking how many variants are left undiscovered following dense sequencing of the human genome. These authors compared three genetic discovery projects (i.e., NIEHS SNPs, Seattle SNPs, and the ENCODE data (Table 2)), and found that only a relatively small sample number of individuals (~150) is sufficient to identify 80% of the variants with a frequency of at least 0.1% (the frequency which defines a SNP), and that to discover all variants requires a larger population sample size (~3000 individuals). This is consistent with the finding that the population sample size necessary to accurately survey a susceptibility allele (i.e., a gene variant associated with a particular phenotype of interest) will depend on the assumptions of how the allele and its effect vary across subpopulations (Pritchard and Donnelly, 2001). Interestingly and relevant to human health risk assessment, the analysis of Ionita-Laza et al. (2009) indicates that environmental response genes show a much greater diversity compared with the average genome. This was observed to be especially true for African populations (Ionita-Laza et al., 2009), which are known to be more genetically and phenotypically diverse on average, compared to non-African groups (European or Asian), as a result of having a longer evolutionary history, as well as having experienced extensive variation in climate, diet and exposure to infectious disease (Campbell and Tishkoff, 2008; Tishkoff et al., 2009).

Use of human mechanistic and polymorphism data in characterization of genetic susceptibility for environmental chemicals

The value of the consolidation and integration of the two data types outlined here is that together these data can provide information about potential genetic susceptibility across human populations, contributing to what is known of differential human response to environmental chemicals. We identified two chemicals for which research has made use of both the available mechanistic and human susceptibility information.

Benzene is one example of a chemical with an abundance of both human mechanistic and genetic susceptibility data. Further, ongoing toxicogenomic studies to identify additional mechanisms and susceptibility factors are underway as part of a systems biology approach to studying benzene (Zhang et al., 2010). Information on genetic susceptibility to benzene exposure provides a complex picture of multiple modes of action and human susceptibility genes identified in both TK and TD steps of the mechanism of action, with some polymorphisms conferring protective phenotypes and others increasing susceptibility in response to benzene exposure. Differences in susceptibility to benzene hematotoxicity were established in several studies of Chinese workers (Rothman et al., 1997; Wan et al., 2002; Lan et al., 2009). Rothman et al. (1997) identified individuals with both rapid CYP2E1 activity (conferring high rates of metabolism to benzene oxide) and two copies of the mutant NQ01 alleles (conferring low detoxification activity). These CYP2E1-NQ01 haplotypes were observed to have a statistically significant increased risk of benzene poisoning compared to controls (i.e., individuals with a slow CYP2E1 activity and two wild-type NQ01 alleles). Later studies indicate a combined effect of three loci in benzene poisoning, NQO1, CYP2E1, and GSTT1, as well as modifying lifestyle factors such as smoking and alcohol consumption (Wan et al., 2002). TD differences in susceptibility to benzene have also been characterized. Polymorphisms in genes involved in DNA repair and genomic maintenance have been identified as associated with risk of benzene-induced hematotoxicity (Lan et al., 2009). At least one of the genes, NQ01, has effects on multiple pathways leading to both TK and TD differences, presumably by impacting other pathways such as stabilizing p53 and microtubule maintenance (Ross and Zhou, 2010). Recent systems biology approaches to the study of benzene that focus on genomic screens for candidate genes related to benzene exposure should further clarify the interactions among benzene toxicity, susceptibility genes, mRNA and DNA methylation (Zhang et al., 2010). Additionally, a case study has been recently initiated to explore how to further utilize the more recent benzene genomic and polymorphism data in risk assessment as part of the NexGen EPA project (Sonawane, personal communication; Ginsberg et al., 2009).

The study of Fry et al. (2008) builds on previous studies of variation in transcription profiles in human lymphoblastoid cells derived from healthy individuals (Cheung et al., 2005; Stranger et al., 2005; Dixon et al., 2007). These authors ask what influence the observed variation has on the response to environmental and chemotherapeutic agents, specifically the DNA alkylating agent MNNG. Fry et al. (2008) examined differentially expressed genes in human lymphoblastoid cells, previously identified through microarray studies (Sabeti et al., 2007). Using the PDR test population of 24 cell lines of diverse ancestry (Collins et al., 1998), Fry et al. (2008) identified 48 of the differentially expressed genes that were predictive (with 94% accuracy) of differences in cellular sensitivity to MNNG. Response to MNNG, and other environmental alkylating agents, has been characterized across many species, and pathways that influence the associated DNA repair mechanism are thought to be well-conserved. They found that transcripts with high MNNG sensitivity are enriched for a common regulatory factor, Oct-1, known to respond to DNA alkylation damage (Zhao et al., 2000), as well as proteins associated with tumorigenesis. The study findings contribute to our understanding of the interindividual differences in the mechanism of DNA damage response. However, one limitation of the study of Fry et al. (2008) is that variation at the level of the individual cell line compared to variation within or between population groups was not addressed, probably due to their use of a subset of the larger PDR panel (24 of 450) that does not comprise a global population set. The study of Fry et al. (2008) represents the only application of a human population diversity panel to the characterization of chemical response and their approach can be applied to other human outcomes. Future studies that include characterization of variants associated with chemical sensitivity to MNNG in human populations would aid in understanding how variation between individuals corresponds to population level variation. Although these two excellent research examples share many of the goals and data sources put forth in the current proposal, they both have specific data availability limitations.

Conclusions

Currently, there is no single resource that integrates chemical mechanistic data with human polymorphism data to allow for identification of gene targets in humans. Here we propose that human genetic susceptibility information useful to human health risk assessment could be gained by integrating these two existing data sources. While the integration of these types of data for environmental chemicals is one area that begs for exploration, the challenge of how to integrate systems biology concepts with interindividual and population variation data has been noted (Zanger, 2010). One recommended need is a web-based database including gene–chemical targets with identified human susceptibility polymorphism data from a global population panel for estimating population risk and characterizing genetic susceptibility on a whole genome scale. However, there are a number of outstanding issues, reviewed below, to realizing this proposed approach and some potential next steps in utilizing genomic data to inform human genetic susceptibility.

First, in order to utilize human genetic susceptibility information of this nature, it is important to understand the mechanism(s) of action of a toxicant, at the level of the molecular factors involved in each biological process (Data Source I; Fig. 2a,b), as well as the potential interactions among processes. One obvious data gap is in complete mechanism of action information, including multiple modes of action and interacting pathways under different conditions, operating in the human. Further, use of the approach presented could be significantly improved with the organization and consolidation of mechanistic information related to human toxic response and association with disease outcome. However, the approach presented does not require a complete understanding of chemical mechanism of action, where all targets and affected pathways have been identified. Rather, one can utilize the available mechanistic data as a starting point in what we and other authors (Andersen et al., 2010; Daston and Naciff, 2010; Watanabe et al., 2011) have described as an iterative process. For example, even information of a proposed or hypothesized mode of action can suggest a biological process, where genes of interest can then be identified and expanded upon by including all available pathway information. Human SNP data, when available, can improve our understanding of the mechanism of action by characterizing gene regions of interest (e.g., gene and protein structure, linkage disequilibrium, and non-synonymous amino acid changes), further elucidating variants that can then be tested for their relationship to functional changes. Finally, the observation of individual or population variation at SNPs found to be informative to mechanism could contribute information about variation in response. Other molecular mechanisms and processes related to genotype–phenotype (e.g., epigenetic changes such as DNA methylation, chromatin remodeling) are just beginning to be interrogated in humans. Understanding the complexity of human biology at a mechanistic level is certainly the research challenge of this century, for which the success of the 21st century paradigm for toxicity testing and risk assessment depends heavily.

Secondly, one of the challenges of making effective use of genetic and genomic data in risk assessment, and toxicological applications in general, is in our ability to identify the biologically important genetic variants for each toxic response phenotype (e.g., to associate phenotype with genotype) (Data Source II; Fig. 2f). Usage of publically available data in the inference of human variability at loci and surrounding regions implicated in chemical response is one approach to characterizing existing variation. The consolidation of this information will reduce the number of technical aspects involved in target identification and selection by identifying what is known of chemical susceptibility loci across the genome, identifying and excluding conserved variants that are functionally significant but that are not variable among individuals and populations. Further, understanding of genomic function outside of the typical promoter-gene structure is limited, and therefore, decision rules for selecting functionally relevant SNPs is an area that should improve in the coming years, as is indicated by the recent increase in functional annotation and phenotype inference tools (Table 2). Additionally, the present discussion does not include the new application of NGS to RNAseq, which has the ability to directly assess the effect of genetic variants on phenotypic response to chemical exposure.

A third issue concerns study design and study comparability within and across the two data sources. For example, the identification of human genetic susceptibility risk factors for environmental agents, where multiple, complex combinations of genetic and environmental factors may contribute to any single phenotype, will require large sample sizes to achieve the power necessary for association mapping of variants. Statistical, specifically power, issues in association mapping of complex disease variants have also been problematic, and may underline the importance of environmental contribution to phenotype. However, recent approaches to more directly understand the environmental contribution to common disease, or chemical response, like the Environment-Wide Association Study (EWAS) (Patel et al., 2010) are promising. Similar to GWAS, the benefits of EWAS is in hypothesis generation, where association between environmental factors and disease can propose tests for correlation and suggest genetic targets for further study. The combination of GWAS and other -omics approaches with EWAS will be useful and will allow for simultaneous assessment of genetic variability of key environmental factors with disease on a global scale, facilitating the characterization of disease causation and estimated risk that has not yet been possible with genome-wide scans alone. The EWAS approach is aligned with what proponents of the 'exposome' (Wild, 2005) are advocating as a means to improve understanding of environmental contribution, specifically, the relation of exposure effects to human disease. Technological advances should, in the near future, make linking exposure to changes at the molecular level feasible (Arnaud, 2010), and this information would certainly inform risk at the individual and population level.

The information that would be generated from applying this approach would be useful to current human health risk assessment practices as well. In the current risk assessment paradigm, information on human genetic susceptibility, when available, is incorporated qualitatively and/or quantitatively. For example, the genetic susceptibility information concerning benzene from several genetic studies, available at the time that the assessment was undertaken, was incorporated qualitatively into the 2003 IRIS benzene assessment (US Environmental Protection Agency, 2002) by describing polymorphisms in metabolic enzymes and the frequencies of those polymorphisms in specific populations. With the NAS envisioned, new risk assessment paradigm of toxicity pathway perturbation, information about mechanism at the level of genes, pathways, and interactions among pathways will continue to become more available for chemicals, where data are currently lacking. There are currently only a small number of pharmaceuticals with relatively well-characterized mechanisms or modes of action (Klein et al., 2001). For example, chemicals that act on xenobiotic metabolizing enzymes, such as cytochrome P450 enzymes, have a well-defined mode of action and have been associated with defined human polymorphisms (e.g., CYP2E1; Ginsberg et al., 2009). Thus, this class of chemicals would be excellent candidates for testing this approach. Additionally, differential susceptibility in response to benzene exposure is an example that could be better characterized by including recent human polymorphism data for the gene and pathway perturbations identified after benzene exposure, from a representative global panel. The approach, outlined herein, of combining next generation sequencing data with mechanistic data for environmental chemicals will inform the relationship between human toxicity pathways, genetic susceptibility, and disease progression (including precursor events) and this information could be applied to risk assessment.

Conflict of interest statement

The authors declare that there are no conflicts of interest.

Acknowledgments

The authors would like to thank Bonnie Joubert, Richard Judson, Jennifer Jinot, and Babasaheb Sonawane for their helpful discussions and comments during the preparation of this manuscript.

References

Andersen, M.E., Clewell III, H.J., Bermudez, E., Dodd, D.E., Willson, G.A., Campbell, J.L., Thomas, R.S., 2010. Formaldehyde: integrating dosimetry, cytotoxicity, and genomics to understand dose-dependent transitions for an endogenous compound. Toxicol. Sci. 118, 716–731.

Arnaud, C.H., 2010. Exposing the exposome. Chemical and Engineering News, pp. 42–44.

Boguski, M.S., 2002. Comparative genomics: the mouse that roared. Nature 420, 515–516.

Butcher, E.C., Berg, E.L., Kunkel, E.J., 2004. Systems biology in drug discovery. Nat. Biotechnol. 22, 1253–1259.

Campbell, M.C., Tishkoff, S.A., 2008. African genetic diversity: implications for human demographic history, modern human origins, and complex disease mapping. Annu. Rev. Genomics Hum. Genet. 9, 403–433.

Chen, J., Aronow, B.J., Jegga, A.G., 2009. Disease candidate gene identification and prioritization using protein interaction networks. BMC Bioinform. 10, 73.

Cheung, V.G., Spielman, R.S., Ewens, K.G., Weber, T.M., Morley, M., Burdick, J.T., 2005. Mapping determinants of human gene expression by regional and genome-wide association. Nature 437, 1365–1369.

Clark, A.G., Hubisz, M.J., Bustamante, C.D., Williamson, S.H., Nielsen, R., 2005. Ascertainment bias in studies of human genome-wide polymorphism. Genome Res. 15, 1496–1502.

Clayton, D.G., Walker, N.M., Smyth, D.J., Pask, R., Cooper, J.D., Maier, L.M., Smink, L.J., Lam, A.C., Ovington, N.R., Stevens, H.E., Nutland, S., Howson, J.M., Faham, M., Moorhead, M., Jones, H.B., Falkowski, M., Hardenbol, P., Willis, T.D., Todd, J.A., 2005. Population structure, differential bias and genomic control in a large-scale, case–control association study. Nat. Genet. 37, 1243–1246.

Collins, F.S., Brooks, L.D., Chakravarti, A., 1998. A DNA polymorphism discovery resource for research on human genetic variation. Genome Res. 8, 1229–1231.

Daston, G.P., Naciff, J.M., 2010. Predicting developmental toxicity through toxicogenomics. Birth Defects Res. C Embryo Today 90, 110–117.

Davis, A.P., Murphy, C.G., Saraceni-Richards, C.A., Rosenstein, M.C., Wiegers, T.C., Mattingly, C.J., 2009. Comparative Toxicogenomics Database: a knowledgebase and discovery tool for chemical–gene–disease networks. Nucleic Acids Res. 37, D786–D792.

Dix, D.J., Houck, K.A., Martin, M.T., Richard, A.M., Setzer, R.W., Kavlock, R.J., 2007. The ToxCast program for prioritizing toxicity testing of environmental chemicals. Toxicol. Sci. 95, 5–12.

Dixon, A.L., Liang, L., Moffatt, M.F., Chen, W., Heath, S., Wong, K.C., Taylor, J., Burnett, E., Gut, I., Farrall, M., Lathrop, G.M., Abecasis, G.R., Cookson, W.O., 2007. A genome-wide association study of global gene expression. Nat. Genet. 39, 1202–1207.

Edwards, S.W., Preston, R.J., 2008. Systems biology and mode of action based risk assessment. Toxicol. Sci. 106, 312–318.

Eichler, E.E., Nickerson, D.A., Altshuler, D., Bowcock, A.M., Brooks, L.D., Carter, N.P., Church, D.M., Felsenfeld, A., Guyer, M., Lee, C., Lupski, J.R., Mullikin, J.C., Pritchard, J.K., Sebat, J., Sherry, S.T., Smith, D., Valle, D., Waterston, R.H., 2007. Completing the map of human genetic variation. Nature 447, 161–165.

Eubanks, M., 1994. Biomarkers: the clues to genetic susceptibility. Environ. Health Perspect. 102 (50–53), 56.

Ewart-Toland, A., Balmain, A., 2004. The genetics of cancer susceptibility: from mouse to man. Toxicol. Pathol. 32 (Suppl 1), 26–30.

Fry, R.C., Svensson, J.P., Valiathan, C., Wang, E., Hogan, B.J., Bhattacharya, S., Bugni, J.M., Whittaker, C.A., Samson, L.D., 2008. Genomic predictors of interindividual differences in response to DNA damaging agents. Genes Dev. 22, 2621–2626.

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

Goh, K.I., Cusick, M.E., Valle, D., Childs, B., Vidal, M., Barabasi, A.L., 2007. The human disease network. Proc. Natl Acad. Sci. USA 104, 8685–8690.

Gohlke, J.M., Thomas, R., Zhang, Y., Rosenstein, M.C., Davis, A.P., Murphy, C., Becker, K.G., Mattingly, C.J., Portier, C.J., 2009. Genetic and environmental pathways to complex diseases. BMC Syst. Biol. 3, 46.

Goldstein, D.B., 2009. Common genetic variation and human traits. N. Engl. J. Med. 360, 1696–1698.

Gonzalez, F.J., Shah, Y.M., 2008. PPARalpha: mechanism of species differences and hepatocarcinogenesis of peroxisome proliferators. Toxicology 246, 2–8.

Gutenkunst, R.N., Hernandez, R.D., Williamson, S.H., Bustamante, C.D., 2009. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. PLoS Genet. 5, e1000695.

Guyton, K.Z., Barone Jr., S., Brown, R.C., Euling, S.Y., Jinot, J., Makris, S., 2008. Mode of action frameworks: a critical analysis. J. Toxicol. Environ. Health 11, 16–31.

Hamadeh, H.K., Bushel, P.R., Jayadev, S., DiSorbo, O., Bennett, L., Li, L., Tennant, R., Stoll, R., Barrett, J.C., Paules, R.S., Blanchard, K., Afshari, C.A., 2002a. Prediction of compound signature using high density gene expression profiling. Toxicol. Sci. 67, 232–240.

Hamadeh, H.K., Bushel, P.R., Jayadev, S., Martin, K., DiSorbo, O., Sieber, S., Bennett, L., Tennant, R., Stoll, R., Barrett, J.C., Blanchard, K., Paules, R.S., Afshari, C.A., 2002b. Gene expression analysis reveals chemical-specific profiles. Toxicol. Sci. 67, 219–231.

Ionita-Laza, I., Lange, C., N, M.L., 2009. Estimating the number of unseen variants in the human genome. Proc. Natl Acad. Sci. USA 106, 5008–5013.

Jayapal, M., Bhattacharjee, R.N., Melendez, A.J., Hande, M.P., 2010. Environmental toxicogenomics: a post-genomic approach to analysing biological responses to environmental toxins. Int. J. Biochem. Cell Biol. 42, 230–240.

Judson, R., Richard, A., Dix, D., Houck, K., Elloumi, F., Martin, M., Cathey, T., Transue, T.R., Spencer, R., Wolf, M., 2008. ACToR–Aggregated Computational Toxicology Resource. Toxicol. Appl. Pharmacol. 233 (1), 7–13.

Judson, R., Richard, A., Dix, D.J., Houck, K., Martin, M., Kavlock, R., Dellarco, V., Henry, T., Holderman, T., Sayre, P., et al., 2009. The toxicity data landscape for environmental chemicals. Environ. Health Perspect. 117 (5), 685–695.

Judson, R., 2010. Public databases supporting computational toxicology. J. Toxicol. Environ. Health 13, 218–231.

Judson, R.S., Houck, K.A., Kavlock, R.J., Knudsen, T.B., Martin, M.T., Mortensen, H.M., Reif, D.M., Rotroff, D.M., Shah, I., Richard, A.M., Dix, D.J., 2010. In vitro screening of environmental chemicals for targeted testing prioritization: the ToxCast project. Environ. Health Perspect. 118 (4), 485–492.

Klaunig, J.E., Babich, M.A., Baetcke, K.P., Cook, J.C., Corton, J.C., David, R.M., DeLuca, J.G., Lai, D.Y., McKee, R.H., Peters, J.M., Roberts, R.A., Fenner-Crisp, P.A., 2003. PPARalpha agonist-induced rodent tumors: modes of action and human relevance. Crit. Rev. Toxicol. 33, 655–780.

Klein, T.E., Chang, J.T., Cho, M.K., Easton, K.L., Fergerson, R., Hewett, M., Lin, Z., Liu, Y., Liu, S., Oliver, D.E., Rubin, D.L., Shafa, F., Stuart, J.M., Altman, R.B., 2001. Integrating genotype and phenotype information: an overview of the PharmGKB project. Pharmacogenetics Research Network and Knowledge Base. The Pharmacogenomics Journal 1, 167–170.

Knudsen, T.B., Martin, M.T., Kavlock, R.J., Judson, R.S., Dix, D.J., Singh, A.V., 2009. Profiling the activity of environmental chemicals in prenatal developmental toxicity studies using the U.S. EPA's ToxRefDB. Reproductive Toxicology (Elmsford, N.Y.) 28, 209–219.

Krewski, D., Acosta Jr., D., Andersen, M., Anderson, H., Bailar III, J.C., Boekelheide, K., Brent, R., Charnley, G., Cheung, V.G., Green Jr., S., Kelsey, K.T., Kerkvliet, N.I., Li, A.A., McCray, L., Meyer, O., Patterson, R.D., Pennie, W., Scala, R.A., Solomon, G.M., Stephens, M., Yager, J., Zeise, L., 2010. Toxicity testing in the 21st century: a vision and a strategy. J. Toxicol. Environ. Health 13, 51–138.

Kudaravalli, S., Veyrieras, J.B., Stranger, B.E., Dermitzakis, E.T., Pritchard, J.K., 2009. Gene expression levels are a target of recent natural selection in the human genome. Mol. Biol. Evol. 26, 649–658.

Lan, Q., Zhang, L., Shen, M., Jo, W.J., Vermeulen, R., Li, G., Vulpe, C., Lim, S., Ren, X., Rappaport, S.M., Berndt, S.I., Yeager, M., Yuenger, J., Hayes, R.B., Linet, M., Yin, S., Chanock, S., Smith, M.T., Rothman, N., 2009. Large-scale evaluation of candidate genes identifies associations between DNA repair and genomic maintenance and development of benzene hematotoxicity. Carcinogenesis 30, 50–58.

LePage, D.F., Conlon, R.A., 2006. Animal models for disease: knockout, knock-in, and conditional mutant mice. Meth. Mol. Med. 129, 41–67.

Linghu, B., Snitkin, E.S., Hu, Z., Xia, Y., Delisi, C., 2009. Genome-wide prioritization of disease genes and identification of disease–disease associations from an integrated human functional linkage network. Genome Biol. 10, R91.

Livingston, R.J., von Niederhausern, A., Jegga, A.G., Crawford, D.C., Carlson, C.S., Rieder, M.J., Gowrisankar, S., Aronow, B.J., Weiss, R.B., Nickerson, D.A., 2004. Pattern of sequence variation across 213 environmental response genes. Genome Res. 14, 1821–1831.

Lohmueller, K.E., Bustamante, C.D., Clark, A.G., 2010. The effect of recent admixture on inference of ancient human population history. Genetics 185, 611–622.

Manolio, T.A., Collins, F.S., 2009. The HapMap and genome-wide association studies in diagnosis and therapy. Annu. Rev. Med. 60, 443–456.

Martin, M.T., Judson, R.S., Reif, D.M., Kavlock, R.J., Dix, D.J., 2009a. Profiling chemicals based on chronic toxicity results from the U.S. EPA ToxRef Database. Environ. Health Perspect. 117, 392–399.

Martin, M.T., Mendez, E., Corum, D.G., Judson, R.S., Kavlock, R.J., Rotroff, D.M., Dix, D.J., 2009b. Profiling the reproductive toxicity of chemicals from multigeneration studies in the toxicity reference database. Toxicol. Sci. 110, 181–190.

Mattingly, C.J., 2009. Chemical databases for environmental health and clinical research. Toxicol. Lett. 186 (1), 62–65.

McGary, K.L., Park, T.J., Woods, J.O., Cha, H.J., Wallingford, J.B., Marcotte, E.M., 2010. Systematic discovery of nonobvious human disease models through orthologous phenotypes. Proc. Natl Acad. Sci. USA 107, 6544–6549.

Miller, R.T., Willson, T.M., 2001. Regulation of xenobiotic metabolism by orphan nuclear receptors. Toxicol. Pathol. 29, 3–5.

Moore, K.J., 1999. Utilization of mouse models in the discovery of human disease genes. Drug Discov. Today 4, 123–128.

Mortensen, H.M., 2008. Genetic variation at the N- acetyltransferase (NAT) genes in global human populations. Doctoral Dissertation: Department of Biology. University of Maryland, College Park.

Mortensen, H. M., Dix, D., Houck, K., Kavlock, R., Shah, I., Judson, R., (2009). The ToxCast™ pathway database identifying toxicity signatures and potential modes of action from chemical screening data. Society of Toxicology, The Toxicologist, pp. Poster #1090.

Mortensen, H. M., Dix, D., Houck, K., Kavlock, R., Shah, I., Judson, R., (2010). Identifying functionally linked gene modules within biological pathways assessed by ToxCast in vitro assays. Society of Toxicology, The Toxicologist, pp. Poster #212.

Mortensen, H.M., Froment, A., Lema, G., Bodo, J.M., Ibrahim, M., Nyambo, T.B., Omar, S.A., Tishkoff, S.A., 2011. Characterization of genetic variation and natural selection at the arylamine N-acetyltransferase genes in global human populations. Pharmacogenomics 12 (11), 1545–1558.

Myers, S., Bottolo, L., Freeman, C., McVean, G., Donnelly, P., 2005. A fine-scale map of recombination rates and hotspots across the human genome. Science 310, 321–324.

National Research Council (2000). Scientific Frontiers in Developmental Toxicology and Risk Assessment. (Board on Environmental Studies and Toxicology, Committee on Developmental Toxicology, Ed.). National Academy of Sciences.

National Research Council (2007). Toxicity Testing in the 21st Century: A Vision and a Strategy. In (Committee on Toxicity Testing and Assessment of Environmental Agents, Ed.). National Academy of Sciences.

National Research Council (2009). Science and Decisions: Advancing Risk Assessment. (N. R. C. Committee on Improving Risk Analysis Approaches Used by the U.S. EPA, Ed.). National Academy of Sciences.

National Research Council (2010). Toxicity Pathway-Based Risk Assessment: Preparing for Paradigm Change: A Symposium Summary(Standing Committee on Risk Analysis Issues and Reviews. Ellen Mantus, Rapporteur.), Washington, DC.

Nielsen, R., Williamson, S., Kim, Y., Hubisz, M.J., Clark, A.G., Bustamante, C., 2005. Genomic scans for selective sweeps using SNP data. Genome Res. 15, 1566–1575.

Nitsch, D., Goncalves, J., Ojeda, F., deMoor, B., Moreau, Y., 2010. Candidate gene prioritization by network analysis of differential expression using machine learning approaches. BMC Bioinform. 11, 460.

Patel, C.J., Bhattacharya, J., Butte, A.J., 2010. An Environment-Wide Association study (EWAS) on type 2 diabetes mellitus. PLoS ONE 5, e10746.

Peters, J.M., Cheung, C., Gonzalez, F.J., 2005. Peroxisome proliferator-activated receptor-alpha and liver cancer: where do we stand? J. Mol. Med. (Berlin, Germany) 83, 774–785.

Peters, J.M., Hollingshead, H.E., Gonzalez, F.J., 2008. Role of peroxisome-proliferator-activated receptor beta/delta (PPARbeta/delta) in gastrointestinal tract function and disease. Clin. Sci. (Lond) 115, 107–127.

Phillips, T.J., Belknap, J.K., Hitzemann, R.J., Buck, K.J., Cunningham, C.L., Crabbe, J.C., 2002. Harnessing the mouse to unravel the genetics of human disease. Genes Brain Behav. 1, 14–26.

Pritchard, J.K., Donnelly, P., 2001. Case–control studies of association in structured or admixed populations. Theor. Popul. Biol. 60, 227–237.

Risch, N., Merikangas, K., 1996. The future of genetic studies of complex human diseases. Science 273, 1516–1517.

Ross, D., Zhou, H., 2010. Relationships between metabolic and non-metabolic susceptibility factors in benzene toxicity. Chem.-Biol. Interact. 184, 222–228.

Rothman, N., Smith, M.T., Hayes, R.B., Traver, R.D., Hoener, B., Campleman, S., Li, G.L., Dosemeci, M., Linet, M., Zhang, L., Xi, L., Wacholder, S., Lu, W., Meyer, K.B., Titenko-Holland, N., Stewart, J.T., Yin, S., Ross, D., 1997. Benzene poisoning, a risk factor for hematological malignancy, is associated with the NQO1 609C→T mutation and rapid fractional excretion of chlorzoxazone. Cancer Res. 57, 2839–2842.

Sabeti, P.C., Varilly, P., Fry, B., Lohmueller, J., Hostetter, E., Cotsapas, C., Xie, X., Byrne, E.H., McCarroll, S.A., Gaudet, R., Schaffner, S.F., Lander, E.S., Frazer, K.A., Ballinger, D.G., Cox, D.R., Hinds, D.A., Stuve, L.L., Gibbs, R.A., Belmont, J.W., Boudreau, A., Hardenbol, P., Leal, S.M., Pasternak, S., Wheeler, D.A., Willis, T.D., Yu, F., Yang, H., Zeng, C., Gao, Y., Hu, H., Hu, W., Li, C., Lin, W., Liu, S., Pan, H., Tang, X., Wang, J., Wang, W., Yu, J., Zhang, B., Zhang, Q., Zhao, H., Zhao, H., Zhou, J., Gabriel, S.B., Barry, R., Blumenstiel, B., Camargo, A., Defelice, M., Faggart, M., Goyette, M., Gupta, S., Moore, J., Nguyen, H., Onofrio, R.C., Parkin, M., Roy, J., Stahl, E., Winchester, E., Ziaugra, L., Altshuler, D., Shen, Y., Yao, Z., Huang, W., Chu, X., He, Y., Jin, L., Liu, Y., Shen, Y., Sun, W., Wang, H., Wang, Y., Wang, Y., Xiong, X., Xu, L., Waye, M.M., Tsui, S.K., Xue, H., Wong, J.T., Galver, L.M., Fan, J.B., Gunderson, K., Murray, S.S., Oliphant, A.R., Chee, M.S., Montpetit, A., Chagnon, F., Ferretti, V., Leboeuf, M., Olivier, J.F., Phillips, M.S., Roumy, S., Sallee, C., Verner, A., Hudson, T.J., Kwok, P.Y., Cai, D., Koboldt, D.C., Miller, R.D., Pawlikowska, L., et al., 2007. Genome-wide detection and characterization of positive selection in human populations. Nature 449, 913–918.

Simmons, D., 2008. The use of animal models in studying genetic disease: transgenesis and induced mutation. Nature Education 1.

Sone, H., Okura, M., Zaha, H., Fujibuchi, W., Taniguchi, T., Akanuma, H., Nagano, R., Ohsako, S., Yonemoto, J., 2010. Profiles of Chemical Effects on Cells (pCEC): a toxicogenomics database with a toxicoinformatics system for risk evaluation and toxicity prediction of environmental chemicals. J. Toxicol. Sci. 35, 115–123.

Spitz, M.R., Bondy, M.L., 2010. The evolving discipline of molecular epidemiology of cancer. Carcinogenesis 31, 127–134.

Stranger, B.E., Forrest, M.S., Clark, A.G., Minichiello, M.J., Deutsch, S., Lyle, R., Hunt, S., Kahl, B., Antonarakis, S.E., Tavare, S., Deloukas, P., Dermitzakis, E.T., 2005. Genome-wide associations of gene expression variation in humans. PLoS Genet. 1, e78.

Stranger, B.E., Nica, A.C., Forrest, M.S., Dimas, A., Bird, C.P., Beazley, C., Ingle, C.E., Dunning, M., Flicek, P., Koller, D., Montgomery, S., Tavare, S., Deloukas, P., Dermitzakis, E.T., 2007. Population genomics of human gene expression. Nat. Genet. 39, 1217–1224.

Tishkoff, S.A., Reed, F.A., Friedlaender, F.R., Ehret, C., Ranciaro, A., Froment, A., Hirbo, J.B., Awomoyi, A.A., Bodo, J.M., Doumbo, O., Ibrahim, M., Juma, A.T., Kotze, M.J., Lema, G., Moore, J.H., Mortensen, H., Nyambo, T.B., Omar, S.A., Powell, K., Pretorius, G.S., Smith, M.W., Thera, M.A., Wambebe, C., Weber, J.L., Williams, S.M., 2009. The genetic structure and history of Africans and African Americans. Science 324, 1035–1044.

Tishkoff, S.A., Reed, F.A., Ranciaro, A., Voight, B.F., Babbitt, C.C., Silverman, J.S., Powell, K., Mortensen, H.M., Hirbo, J.B., Osman, M., Ibrahim, M., Omar, S.A., Lema, G., Nyambo, T.B., Ghori, J., Bumpstead, S., Pritchard, J.K., Wray, G.A., Deloukas, P., 2007. Convergent adaptation of human lactase persistence in Africa and Europe. Nat. Genet. 39, 31–40.

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

US Environmental Protection Agency (2005). Supplemental Guidance for Assessing Susceptibility from Early-Life Exposure to Carcinogens. (Risk Assessment Forum, Ed.), Washington, DC USA.

Voight, B.F., Kudaravalli, S., Wen, X., Pritchard, J.K., 2006. A map of recent positive selection in the human genome. PLoS Biol. 4, e72.

Wan, J., Shi, J., Hui, L., Wu, D., Jin, X., Zhao, N., Huang, W., Xia, Z., Hu, G., 2002. Association of genetic polymorphisms in CYP2E1, MPO, NQO1, GSTM1, and GSTT1 genes with benzene poisoning. Environ. Health Perspect. 110, 1213–1218.

Watanabe, K.H., Andersen, M.E., Basu, N., Carvan III, M.J., Crofton, K.M., King, K.A., Sunol, C., Tiffany-Castiglioni, E., Schultz, I.R., 2011. Defining and modeling known adverse outcome pathways: domoic acid and neuronal signaling as a case study. Environmental toxicology and chemistry / SETAC 30, 9–21.

Weir, B.S., Cardon, L.R., Anderson, A.D., Nielsen, D.M., Hill, W.G., 2005. Measures of human population structure show heterogeneity among genomic regions. Genome Res. 15, 1468–1476.

Wiegers, T.C., Davis, A.P., Cohen, K.B., Hirschman, L., Mattingly, C.J., 2009. Text mining and manual curation of chemical–gene–disease networks for the comparative toxicogenomics database (CTD). BMC Bioinform. 10, 326.

Wild, C.P., 2005. Complementing the genome with an "exposome": the outstanding challenge of environmental exposure measurement in molecular epidemiology. Cancer Epidemiol Biomarkers Prev 14, 1847–1850.

Wilson, S.H., Olden, K., 2004. The Environmental Genome Project: phase I and beyond. Mol. Interv. 4, 147–156.

Young, L.J., 2001. Oxytocin and vasopressin as candidate genes for psychiatric disorders: lessons from animal models. Am. J. Med. Genet. 105, 53–54.

Zanger, U., 2010. Pharmacogenetics—challenges and opportunities ahead. Frontiers in Pharmacology:Opinion Article 1.

Zhang, L., McHale, C.M., Rothman, N., Li, G., Ji, Z., Vermeulen, R., Hubbard, A.E., Ren, X., Shen, M., Rappaport, S.M., North, M., Skibola, C.F., Yin, S., Vulpe, C., Chanock, S.J., Smith, M.T., Lan, Q., 2010. Systems biology of human benzene exposure. Chem.-Biol. Interact. 184, 86–93.

Zhao, H., Jin, S., Fan, F., Fan, W., Tong, T., Zhan, Q., 2000. Activation of the transcription factor Oct-1 in response to DNA damage. Cancer Res. 60, 6276–6280.