

THE 1000 GENOMES TOXICITY SCREENING PROJECT: UTILIZING THE
POWER of HUMAN GENOME VARIATION FOR POPULATION-SCALE IN
VITRO TESTING

Nour Abdo

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in
partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department
of Environmental Sciences and Engineering, Gillings School of Global Public Health

Chapel Hill
2014

Approved by:

Ivan Rusyn

Fred Wright

Rebecca Fry

Kari North

Avram Gold

© 2014
Nour Abdo
ALL RIGHTS RESERVED

ABSTRACT

Nour Abdo: The 1000 Genomes Toxicity Screening Project: Utilizing the power of human genome variation for population-scale *in vitro* testing
(Under the direction of Ivan Rusyn, M.D., Ph.D.)

Incorporation of novel toxicity screening approaches is a crucial tool for tackling the complex contemporary challenges in evaluating the human health hazards of exposure to chemicals. Current *in vitro* testing paradigms still have major gaps that need addressing, such as population-based *in vitro* approaches to qHTS screening. This study evaluated the *hypothesis* that comparative population genomics with efficient *in vitro* experimental design can be used for the evaluation of the potential hazard, mode of action, and the extent of population variability in response to chemicals. **In Aim 1**, we evaluated and assessed the validity of *in vitro* genetically-anchored population human model system in assessing chemical toxicity and identifying candidate genetic susceptibility. We screened 81 human lymphoblast cell lines with 240 chemicals at 12 different concentrations and assessed the toxic response using different endpoints (cell death and caspase production). We evaluated the toxic responses to a panel of chemicals observed in lymphoblast cell lines, and compared them to other toxic responses seen with different cell lines that originate from different sources. **In Aim 2**, we expanded our model to include more than one population. The goals were to (1) quantitatively assess population-based toxicological hazard to environmental contaminants, (2) determine the extent of human inter-individual variability in chemical toxicity, identify susceptible sub-populations or races, (3) understand the genetic determinants of the inter-individual variability, (4) generate testable hypotheses about toxicity pathways by leveraging genetic and genomic data from 1000 Genomes

and HapMap Projects, and (5) use the data obtained from this research to build predictive in silico models. *In Aim 3*, we addressed some of the remaining challenges in our model, such as the ability to screen chemical mixtures. We explored the potential and efficiency of our model in assessing new challenges such as the evaluation of environmental chemical mixtures in a population in vitro screening, and the extrapolation of the in vitro hazard to an oral equivalent dose. In summary, this research not only will use novel tools to investigate population genetically anchored variability, but it will also offer exceptional methodology for incorporating scientifically-based estimates of uncertainty in risk assessment.

DEDICATION

To my family, Ivanka, Rami, Danny and Rani

I couldn't have done all this work without your continuous support and guidance. I appreciate all the prayers, sacrifices, and advice you have given me along the way. I also want to thank my sister Ruja for all her positive encouragement.

ACKNOWLEDGEMENTS

I am extremely indebted to my advisor, Dr. Rusyn for his continuous guidance and support throughout my study. I am also grateful for the advice and tremendous help from Dr. Wright without whom this work would not have been possible. I also deeply appreciate my wonderful committee (Dr. Rusyn, Dr. Wright, Dr. Fry, Dr. North, and Dr. Gold) who supported and guided me in my research. Great teachers like my committee are not easy to find and I appreciate all their time, patience, and ability to make time for my questions and helping me to grow professionally.

I wish to thank many people who I have collaborated with and they contributed tremendously to my projects. It is of great pleasure to recognize:

- Dr. Ray Tice from National Institute of Health : National Institute of Environmental Health Sciences/National Toxicology Program
- Christopher Austin, Menghang Xia, Ruili Huang, Srilatha Sakamuru from the National Center for Advancing Translational Sciences.
- Dr. Alison Motsinger-Reif, Dr. Damian Shea, Dr. Chad Brown, Dr. John Jack, Dr. Yihui Zhou from North Carolina State University
- Dr. Munir Pirmohamed, and Philippe Marlot from University of Liverpool, United Kingdom.
- Dr. Eric Lock and Paul Gallins from Biostatistics Department at University of North Carolina at Chapel Hill.
- Dr. Barbara Wetmore from The Hamner Institute
- Dr. Mathew Martin and Dr. John Wambaugh from the US Environmental Protection Agency.

- Current and former members of Rusyn lab including Oksana Kosyk, Grace Chapel, Andrew Shapiro, Jessica Wignall, Dr. Hong Sik Yoo, Dr. Mary Kushman, Svitlana Shymonyak, Blair Bradford, Dr. Shinji Furuya, Abhishek Venkatratnam, Shannon O'shea, Samantha Tulenko, Alexandra Hamilton, Sangeetha Kumar, Shannon O'shea.

TABLE OF CONTENTS

LIST OF TABLES	xi
LIST OF FIGURES	xii
LIST OF ABBREVIATIONS	xiv
CHAPTERS	
I. GENERAL INTRODUCTION	1
A. THE OVERARCHING NECESSITY TO REGULATE THOUSANDS OF CHEMICALS.....	1
1. The shortcomings of traditional toxicological methods to inform risk assessment.....	3
2. What alternatives to animal models exist to meet the high demand of regulating thousands of chemicals? Toxicity Testing in the 21st Century.....	4
3. Current <i>in vitro</i> paradigms do not measure population variability and rely on rigid unproven assumptions.....	6
B. GAUGING POPULATION VARIABILITY.....	7
C. LYMPHOBLASTOID CELL LINES IN PHARMACOGENOMIC DISCOVERY AND CLINICAL TRANSLATION.....	10
1. Establishment of lymphoblastoid cell lines (LCLs).....	10
ii. The 1000 Genomes Project.....	10
iii. Transformation of B-Lymphocytes to LCLs	12

2. Utility of LCLs in pharmacogenomics discovery.....	13
D. UTILIZING LYMPHOBLASTOID CELL LINES IN TOXICOGENOMIC POPULATION-BASED IN VITRO TOXICITY SCREENING.....	15
1. Advantages and Limitations of Utilizing LCLs in Toxicogenomic Discovery.....	15
2. Toxicity Phenotype	17
3. How <i>in vitro</i> data could inform systematic risk assessment?	20
E. SPECIFIC AIMS	22
1. AIM 1: To evaluate and assess the validity of an in vitro, genetically–anchored human population model system in assessing chemical toxicity and identifying candidate genetic susceptibility.....	23
2. AIM 2: To quantitatively evaluate inter-individual variability for diverse environmental chemicals using high throughput large scale in-vitro model across genetically -defined genetically-diverse populations.....	24
3. AIM3: To investigate remaining challenges of in vitro genetically–anchored population human model, such as the potential to screen complex mixtures.....	24
F. REFERENCES	25

II.	CHAPTER 2: QUANTITATIVE HIGH-THROUGHOUT SCREENING FOR CHEMICAL TOXICITY IN A POPULATION-BASED IN VITRO MODEL.....	35
	A. ABSTRACT	36
	B. INTRODUCTION	37
	C. MATERIALS AND METHODS	39
	D. RESULTS	45
	E. DISCUSSION	52
	F. REFERENCES	72
III.	CHAPTER 3: GENETIC MAPPING OF IN VITRO SUSCEPTIBILITY TO CYTOTOXIC COMPOUNDS- THE 1000 GENOMES HIGH THROUGHPUT SCREENING STUDY.....	76
	A. ABSTRACT	78
	B. INTRODUCTION	79
	C. MATERIALS AND METHODS	82
	D. RESULTS	88
	E. DISCUSSION	97
	F. REFERENCES	147
IV.	CHAPTER 4: IN VITRO SCREENING FOR INTER-INDIVIDUAL AND POPULATION VARIABILITY IN TOXICITY OF PESTICIDE MIXTURES	152
	A. ABSTRACT	153
	B. INTRODUCTION	154
	C. MATERIALS AND METHODS	155
	D. RESULTS	157

E. DISCUSSION	165
F. REFERENCES	198
V. GENERAL DISCUSSION	202
A. SUMMARY AND CONCLUSIONS	201
1. QUANTITATIVE HIGH-THROUGHOUT SCREENING FOR CHEMICAL TOXICITY IN A POPULATION-BASED IN VITRO MODEL	201
2. GENETIC MAPPING OF IN VITRO SUSCEPTIBILITY TO CYTOTOXIC COMPOUNDS-THE 1000 GENOMES HIGH THROUGHPUT SCREENING STUDY	202
3. IN VITRO SCREENING FOR INTER- INDIVIDUAL AND POPULATION VARIABILITY IN TOXICITY OF PESTICIDE MIXTURES	205
B. SIGNIFICANCE, IMPACT, AND INNOVATION OF THIS STUDY	208
C. LIMITATIONS	212
D. FUTURE DIRECTIONS	215
E. REFERENCES	217

LIST OF TABLES

Table 1.1	Largest pharmacogenomics studies that utilized LCLs for GWAS analysis	14
Table 2.1	List of 240 Screened Chemicals	66
Table 3.1	Average EC10 Cytotoxicity Values and Fold-Ranges	111
Table 3.2	Chemicals showing significant EC10 variation across populations or by sex	117
Table 3.3	MAGWAS Multivariate Association Results	121
Table 3.4	Significant EC10 –SNP associations among set of 1.3m SNPs	124
Table 3.5	Significant EC10 –SNP associations among larger set of 12m SNPs.....	134
Table 3.6	SNP set pathway	145
Table 3.7	LASSO prediction accuracy of expression vs. EC ₁₀ for chemicals with proportion of variance explained R ² >0.01. Prediction accuracy of expression vs. EC ₁₀	146
Table 4.1	The Screened Population’s Distribution	190
Table 4.2	Individual Pesticides in Chlorinated Pesticide Mixture.	191
Table 4.3	Individual Pesticides Current-use Pesticide Mixture	192
Table 4.4	Summary Statistics of EC10s for each mixture	194
Table 4.5	Margin of Exposure for Current-Use Pesticide Mixture	195
Table 4.6	Margin of Exposure for Chlorinated Pesticide Mixture	196
Table 4.7	Top Results from Pathway Analysis of Pesticide Mixtures	197

LIST OF FIGURES

Figure 1.1	Schematic comparison of current and envisioned risk assessment paradigm „.....	6
Figure 1.2	The 1000 Genomes Project Populations	11
Figure 2.1	Experimental reproducibility	56
Figure 2.2	Distribution of cytotoxicity across chemicals	57
Figure 2.3	Cytotoxicity Distributions across Cell Lines	58
Figure 2.4	Inter-individual Variability Range for the 240 Chemicals	59
Figure 2.5	Toxicity-genotype Associations	61
Figure 2.6	Toxicity-RNA expression Associations.....	62
Figure 2.7	Comparison of Cytotoxicity across Different Studies	64
Figure 2.8	Comparison of Cytotoxicity Range across Different Studies.....	65
Figure 3.1	Illustrative Fits for Cytotoxicity Estimation	100
Figure 3.2	Distributions of the LCLS among the 9 Populations	101
Figure 3.3	Flow Chart of Data Processing to Obtain Cytotoxicity Response Values and EC ₁₀	102
Figure 3.4	Comparison of the Current Study to Other Comparable LCL cell line/screening studies and Reproducibility	103
Figure 3.5	Magnitude of Inter-Individual Variability Across Chemicals and Populations	105
Figure 3.6	P-values vs. partial R ²	107
Figure 3.7	Genotype-Cytotoxicity Associations	108
Figure 3.8	MAGWAS vs. PLINK for an example chemical/region	109
Figure 3.9	Heritability of EC ₁₀	110
Figure 3.10	Heatmaps of Gene- and Pathway-Chemical Associations	111

Figure 4.1	Inter-individuals and Population Variability and Reproducibility of the Mixtures.	178
Figure 4.2	Boxplots of Population Differences for Each Mixture	180
Figure 4.3	Illustrative Flow Chart of Predicted Exposure Limits Calculations	181
Figure 4.4	Distribution of EC10s and Oral Equivalent Dose Across 146 cell lines for each mixture.....	182
Figure 4.5	Illustrative Flow chart of Oral Equivalent Dose Calculations	184
Figure 4.6	Oral Equivalent Doses with the Four Different Scenarios.	185
Figure 4.7	Inter-individuals and Population Variability and Reproducibility of the Mixtures.....	186
Figure 4.8	MAGWAS Results for Current-Use Pesticide Mixture....	188

LIST OF ABBREVIATIONS

ATP	adenosine triphosphate
CEPH	Centre d'Etude du Polymorphisme Humain
FBS	fetal bovine serum
FDR	False Discovery Rate
GWAS	Genome Wide Association Study
qHTS	Quantitative high-throughput screening
MAGWAS	Multivariate ANCOVA Genome-Wide Association Software
SNP	single nucleotide polymorphisms
PK	Pharmacokinetics
FWER	Family-wise error rate

CHAPTER 1: GENERAL INTRODUCTION

A. THE OVERARCHING NECESSITY TO REGULATE THOUSANDS OF CHEMICALS.

Several federal agencies in the United States have been bestowed the responsibility to regulate a diverse variety of environmental compounds. For example, the Toxic Substances Control Act (TSCA), which was passed in 1976, granted authority to the U.S. Environmental Protection Agency (EPA) to oversee the safety of chemical products in commerce to both the public and the environment, and to ensure continuous reviewing and regulation of chemicals in commerce (EPA 2012). Consequently, those federal agencies are responsible for implementing regulations that sets the maximum permissible thresholds for environmental compounds in drinking water, establish acceptable limits of exposure in occupational settings, and to determine tolerances for pesticides residues in food, among other tasks (National Research Council 2006).

Previously, the US EPA has relied primarily on information obtained from *in vivo* animal models. Traditional toxicity testing is carried out in laboratory animals to assess the hazards and risks associated with exposure to environmental agents. Animal models have afforded valuable information on the possible harmful effects of exposure to a chemical and the associated dose at which effects may be observed. With limited confounding, controlled experimental design, and whole intact body system, *in vivo* animal models were deemed an invaluable resource for understanding toxicity risk.

With traditional animal models, each chemical requires multiple tests, can use up to 5000 animals (or even 12,000 for some pesticides), costs millions of dollars, and take up to 5 years of testing or more. These disadvantages are in addition to ethical concerns that have been raised about animal welfare (Abbot 2005). There are more than 10,000 chemicals in commercial use in the United States, with hundreds being introduced each year (Anastas et al. 2010). However, only a small fraction of these chemical have been adequately evaluated for their potential risk to human health (Anastas et al. 2010; Judson et al. 2008). Accordingly, there has been a growing concern from governmental agencies about the need to adequately and accurately assess thousands of chemicals in a rapid and efficient manner. New paradigms are inevitably needed for the fast and accurate evaluation of the potential human health hazard of environmental chemicals (Collins et al. 2008).

The emerging massive demand for to identify data sources for regulating chemicals is not limited to the United States. There is a global awareness of the challenges associated with toxicology testing. The European Union's Registration, Evaluation, Authorisation, and Restriction of Chemicals (REACH) legislation is one of the international agencies that have set out to better understand and manage risks to human health and the environment that arise from the manufacture and use of chemicals (Anastas et al. 2010).

Incorporation of novel toxicity screening approaches is a crucial tool for tackling the complex contemporary challenges in evaluating the human health hazards of exposure to chemicals. A shift in toxicity testing from *in vivo* to *in vitro* methods may efficiently prioritize compounds, reveal new mechanisms, and enable predictive modeling. The emergence of new toxicity methods and strategies might address several risk assessment needs, as well as bring about new challenges. The following sections discuss current toxicity approaches and major

remaining challenges to meet risk assessment demands for chemicals with variable modes of action.

1. *The shortcomings of traditional toxicological methods to inform risk assessment.*

Using animal models to evaluate toxicological response depends primarily on observing an adverse health outcome with high doses of a chemical. Uncertainties associated with animal data are usually handled by the use of a 10-fold interspecies uncertainty factor when extrapolating from laboratory animals to humans (EPA, 2004). However, the 10-fold interspecies uncertainty factor does not account for the additional uncertainty in extrapolating from high doses in animal experiments to environmental exposure levels that are orders of magnitude lower (Andersen and Krewski 2009). Although studies using animal models have improved dramatically in recent years in term of dosing, existing challenges and limitations remain. These include the high cost and labor required to handle animals, the extensive amount of time needed to conduct, assess and fully evaluate the health outcomes seen in animals, the challenges in translating human relevance in toxicological outcomes that are species dependent, the limited capability of assessing multiple compounds simultaneously, the inability to assess mixtures or several compounds at once, and most importantly the limited ability to provide timely relevant information to support informed regulation of environmental compounds. Due to the above reasons, growing concern and frustration has been expressed by health protection and regulatory agencies, leading to the request for improved alternative models that are faster and more efficient in tackling the

hundreds of thousands of commercial chemicals that require proper assessment and evaluation (Andersen and Krewski 2009).

2. *What alternatives to animal models exist to meet the high demand of regulating thousands of chemicals? Toxicity Testing in the 21st Century.*

The growing concern from regulatory agencies, such as the Registration, Evaluation, Authorization and Restriction of Chemicals (REACH) in Europe and the Toxic Substances Control Act reform in the US, to meet the high demand of regulation of thousands of chemicals that are being released to the environment has resulted in a shift in toxicity assessment from *in vivo* to *in vitro* and *in silico* methods (Plunkett et al. 2010). The proposed shift should be carried out by the development of rapid screening methods based on the mechanistic understanding of biological processes (Plunkett et al. 2010; Judson et al. 2010). It is necessary to merge high-information content biology and modeling with mechanistic research to build a predictive framework of an intact biological system (Chiu et al. 2011).

Several recent governmental initiatives have been established in the US to meet the requirements and needs of exploration of new methodologies in toxicology to be incorporated into current and future risk assessments. For example, the Tox21 program (Collins et al. 2008), a partnership between four governmental agencies (EPA, NTP: National Toxicology Program, NCGC: NIH Chemical Genomics Center, and FDA: Food and Drug Administration), is currently screening thousands of chemicals for their potential to disturb biological pathways that may result in human disease, with a broad spectrum of *in vitro* assays utilizing quantitative high throughput screening (qHTS) format (Xia et al. 2008a). Such data from toxicologically relevant

in vitro endpoints can be utilized as toxicity-based triggers to assist in decision-making (Reif et al. 2010), act as predictive surrogates for *in vivo* toxicity (Zhu et al. 2008; Martin et al. 2010), or to generate testable hypotheses on the mechanisms of toxicity (Xia et al. 2008a). Another governmental initiative is The NexGen program, a collaborative effort between EPA, NTP, ATSDR, NCGC and CalEPA (US EPA, 2011), which is developing new approaches and methods to better utilize novel molecular toxicology data to understand the risks posed by environmental exposures. The NexGen program provides the opportunity to address challenges and opportunities of converting data generated through Tox21 into knowledge and ultimately into the scientific basis for NexGen risk assessments (US EPA 2011).

Nevertheless, a major gap that is not being currently addressed in either Tox21 or NexGen, but is the focus of the research detailed herein, is the population-based in vitro approach to qHTS screening. The availability of genetically diverse defined renewable sources of human cells, such as transformed lymphoblasts, will allow for the investigation of the hazard and degree of inter-individual biological variability in the human population, as well as to understand and comprehensively characterize the role of human genome sequence variation in observed inherited variation in toxicity phenotypes.

3. *Current in vitro paradigms do not measure population variability and rely on rigid unproven assumptions.*

Risk assessments need to be regularly reviewed and modified to meet the demands of new discoveries and evolving technologies. With the recent shift from *in vivo* to *in vitro* in toxicity testing, the Committee on Toxicity Testing and Assessment of Environmental

Agents of the National Research Council (National Research Council 2007) has recommended certain visionary guidelines for assessing *in vitro* testing as depicted in Figure 1 (fully

reproduced from (Crump et al. 2010). The suggested paradigm proposes a similar modified approach to deal with *in vitro* data to the previous current *in vivo* paradigm. The future model includes different components to make it a

successful and valuable tool. Two of those important components and suggestions are: (i) to test cell lines from many human donors that are representative of diverse populations and (ii)

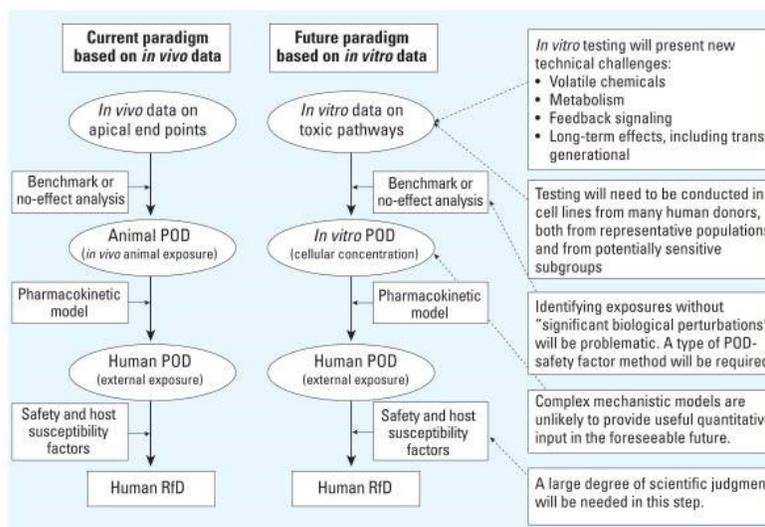


Figure 1.1: Schematic comparison of current and envisioned risk assessment paradigm adapted from EHS (Crump et al. 2010). Fully reproduced from PMC under

“The following PMC journals are U.S. Government publications:

- Addiction Science & Clinical Practice (vol. 1 through vol. 6)
- Emerging Infectious Diseases
- Environmental Health Perspectives
- Preventing Chronic Disease

All material published in these journals is in the public domain and may be used and reproduced without special permission. However, anyone using the material is requested to properly cite and acknowledge the source. ”

<http://www.ncbi.nlm.nih.gov/pmc/about/copyright/>

taken June 20th 2014

to incorporate certain safety and host susceptibility factors that require a large degree of scientific judgment to derive a meaningful human reference dose (Crump et al. 2010). While the Tox21 program is currently screening thousands of environmental chemicals for their potential to affect biological pathways that may result in human disease (Xia et al. 2008a), the current screening paradigms do not contain enough human cell lines to assess the degree of inter-individual biological variability in the human population. Such variability is of particular importance in assessing potential human health hazard (National Research Council 2007). Instead, the current default assumption for inter-individual variability, which is not based on any biological assessment, is a 3-fold difference in toxicodynamic studies and a 3-fold difference in toxicokinetics studies (International Programme for Chemical Safety 2001).

B. GAUGING POPULATION VARIABILITY

Exposure to environmental chemicals may result in a variety of health outcomes that are different in type or magnitude among individuals and/or populations. Those differences in response are due to underlying human variability and should be addressed for a valuable human health risk assessment of chemicals (Guyton et al. 2009; Hattis 1997; National Research Council 2009). The variety of health outcomes from exposure to chemicals can be a result of intrinsic factors (genetic variability), extrinsic factors (life style, environment, etc), or the interaction of both intrinsic and external factors (Zeise et al. 2013). Human variability is currently assessed by applying “uncertainty” or “adjustment” factors (U.S. EPA, 2011). A factor of 1, 3, or 10 has been used to account for inter-individual variability in human population. In some cases, the factor is

further divided to separately account for variation in pharmacokinetics and pharmacodynamics (US EPA 2011) (Sonich-Mullin et al. 2001).

Characterizing genetic variability can enhance our understanding of population variability in response to chemicals. The finding of genetic loci associated with susceptibility can potentially inform us of important cellular proteins that affect health outcome and can uncover novel toxicity pathways. Endeavors to map human variability have been focused on discovering genetic variations (Schadt and Bjorkegren 2012), or other -omics variations including epigenomics, transcriptomics, proteomics, and metabolomics (Chen et al. 2008; Emilsson et al. 2008; Illig et al. 2010; Manolio 2010; Cornelis et al. 2010; Schadt 2009). Genetic polymorphisms may act together or separately to alter susceptibility to adverse effects of exposure to environmental agent. Thus, understanding genetic susceptibility has the potential for predicting toxicity pathways (National Research Council 2009). For instance, adverse health outcome to occupational exposure to welding fumes has been associated with genetic polymorphisms in DNA repair and detoxifying genes. Consequently, sensitive individuals have evidence of higher chromosomal and DNA damage (Iarmarcovai et al. 2005).

Modern approaches to genetic epidemiology include valuable approaches to integrate toxicity exposure in population-based setting to potentially associated genetic loci. Several studies have been able to utilize genetic epidemiology to determine the relationship between specific genes in the population and adverse health outcomes from environmental chemical exposure. For example, human epidemiological studies have provided information on DNA damage in arsenic-exposed populations (Andrew et al. 2006). Further *in vitro* studies were able to elucidate the specific DNA-repair pathways affected by arsenic (Andrew et al. 2006). While genetic epidemiology holds a great value in population risk assessment, several factors might hinder its

ability to overcome new challenges in toxicology. First, it is difficult to estimate and accurately measure environmental exposure amounts. Furthermore, it is challenging to link the exposure with the outcome. The level of exposure (by amount in air, or water) does not necessarily reflect the actual dose that enters the body. While certain biomarkers could reflect internal ingested/inhaled exposure, they are seldom measured. Biomarkers might be invasive, expensive to measure, or/and not characterized or known for certain chemicals. Genetic epidemiology depends on actual real human exposures. This means we can only evaluate chemicals that individuals are exposed to and cannot make assessments on new chemicals. Second, genetic epidemiology can be enormously expensive and requires large sample sizes (Burton et al. 2009). Except for a few diseases, health outcomes are complex and the contribution of individual genetic loci is modest at best (Manolio et al. 2009). In order to have enough power to detect potentially associated loci, thousands of individuals need to be exposed and genotyped for a meaningful study. With humans, confounding factors are a major issue, where individuals are exposed to other potential chemicals and have predisposed internal and external factors that could confound the outcome or increase variability of outcomes. While large sample sizes and careful epidemiological designs can potentially decrease confounding, it is difficult to achieve pure design principles with the reality of real human exposures. Third and most importantly, the overarching challenge of evaluating thousands of chemicals and mixtures in a short amount of time in epidemiological studies is not tenable.

Other potential ways to investigate population variability is either through *in vivo* or *in vitro* toxicity testing paradigms. Several *in vivo* studies with genetically defined mouse models were designed to discover the genetic determinants of susceptibility and population variability (Rusyn et al. 2010). However, the extrapolation of population variability from animal models to human

population, in addition to the human relevance of the detected susceptibility loci in mice has yet to be fully determined (Rusyn et al. 2010). Moreover, animal model studies are labor intensive, not amenable to high-throughput screening, and can only effectively evaluate one chemical at a time.

In vitro population based testing paradigms have been deemed useful in identifying adverse health outcomes that were not detected in preclinical and clinical testing for pharmaceutical products (IOM 2007), and in tailoring chemotherapy treatment based on patient's genetics (Phillips and Mallal 2010) or tumor type (La Thangue and Kerr 2011). With the expense of developing new drugs, pharmacogenomics studies have been aimed to maximize effective therapy response and to minimize adverse reactions by prescribing treatment based on a patient's genetic profile (Wheeler and Dolan 2012).

C. LYMPHOBLASTOID CELL LINES IN PHARMACOGENOMIC DISCOVERY AND CLINICAL TRANSLATION.

1. Establishment of lymphoblastoid cell lines (LCLs).

I. The 1000 Genomes Project

The 1000 Genomes Project is an international consortium with multiple centers, platforms, and funders to construct a foundational data set for human genetics (Kuehn 2008; Clarke 2012). The main purpose of the project is set to discover virtually all common human variations by investigating many genomes at the base pair level (Kuehn 2008; Clarke 2012). It aims to

discover population level human genetic variations of all types (95% of variation > 1% frequency), define haplotype structure in the human genome, and develop sequence analysis methods, tools, and other reagents that can be transferred to other sequencing projects (Kuehn 2008; Clarke 2012).

The consortium recruited healthy adult volunteers from different continents, representing wide variation in genetic ancestries. From blood samples, the consortium isolated B-Lymphocytes to be later transformed to lymphoblastoid cell lines (LCLs) and genotyped via standard array technologies, and eventually these individuals will be sequenced to high coverage (Clarke 2012). The consortium also included cell lines developed as part of the HapMap project, including early CEPH populations (Clarke 2012). Figure 2 illustrate the up-to-date populations that were included in the 1000 Genomes Project.

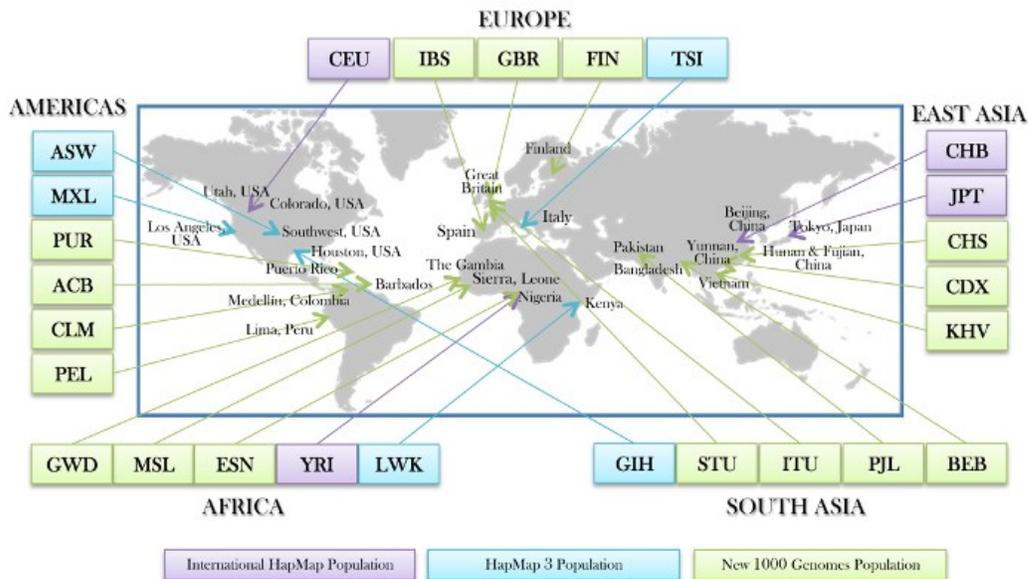


Figure 1.2: The 1000 Genomes Project Populations. Obtained from <http://www.1000genomes.org/cell-lines-and-dna-coriell>

II. Transformation of B-Lymphocytes to LCLs

Immortalization is the process of transforming normal primary B-Lymphocytes to LCLs and giving them the ability to indefinitely proliferate. Consequently, having an unlimited life span with no other additional changes (Miller 1982). Normal cells are mortal, in part because their telomeres shorten with each cell division (Higaki et al. 2004). Telomeres are repetitive nucleotides at the end of chromosomes that act as a buffer layer to protect chromosomes from fusing together and genes from being lost during cell division. With cell division and replication, the DNA duplicates and the telomere sequences become shorter (Qian et al. 2014). Replenishing telomeres is dependent upon the enzyme telomerase reverse transcriptase (Qian et al. 2014). LCLs are immortalized by developing strong telomerase activity and other cellular changes upon Epstein-Barr virus (EBV) integration (Miller et al. 1982). *In vitro* infection with EBV of B-Lymphocytes aids in their transformation to LCLs (Miller et al. 1982). EBV, which is a human herpes virus, is a common virus in humans and has been associated with mononucleosis (Weiss and O'Malley 2013), autoimmune diseases such as multiple sclerosis, dermatomyositis, systemic lupus erythematosus, rheumatoid arthritis (Toussiroit and Roudier 2008; Dreyfus et al. 2011; Pender et al. 2012; Ascherio and Munger 2010), various forms of cancer such as Hodgkin's lymphoma, Burkitt's lymphoma, nasopharyngeal carcinoma (Niedobitek et al. 2001; Epstein 2001), and conditions associated with human immunodeficiency virus (HIV) such as hairy leukoplakia and central nervous system lymphomas (Kawa 2000; Maeda et al. 2009).

It was discovered in 1968 that EBV infects resting primary human B cells, activates them, and establishes a latent infection in them (Henle and Henle 1980). After *in vitro* infection with EBV, the B-lymphocytes continue proliferating and give rise to stable LCLs with genomic virion DNA that is approximately 170 Kbp (Kalla and Hammerschmidt 2012; Amon et al. 2004). The

virion DNA exists as independent complete multicopy circular extrachromosomal plasmid-like DNA, and is not integrated into the cell's chromosomes (Miller 1982). Moreover, the virion DNA replicates as the cellular DNA in the nucleus replicates in latency, infected proliferating B cells (Kalla and Hammerschmidt 2012) and mostly does not express its genetic information except for some viral products (Miller 1982).

2. Utility of LCLs in pharmacogenomics discovery

The utilization of lymphoblastoid cell lines (LCLs) as a human model has emerged as a promising tool in the study of the genetics of drug response. Many studies have used LCLs for pharmacogenomics discoveries by observing response to chemotherapeutics (Wheeler and Dolan 2012), radiation (Smirnov et al. 2009; Niu et al. 2010), statins (Simon et al. 2006; Mangravite et al. 2008; Wilke et al. 2008; Medina et al. 2008), selective serotonin reuptake inhibitors (Morag et al. 2011), immunosuppressants (Wheeler and Dolan 2012), pain relievers and β -blockers (Wheeler and Dolan 2012).

Pharmacogenomics studies using LCLs have not only proven to yield suggestive associations between genetic variants and drugs, but these associations have been supported in real clinical data. For example, LCL-based GWAS led to the discovery of an association between cytarabine arabinoside cytotoxicity and *FKBP5* gene expression (Li et al. 2008). In a follow up clinical targeted gene study, both event-free and overall survival for acute myeloid leukemia patients treated with cytarabine arabinoside were associated with SNPs located in the *FKBP5* gene (Mitra et al. 2011). Successful clinical translation of LCL pharmacogenomics discoveries is not limited to the previous example. Several other studies have found value in performing GWAS and/or eQTL when compared to a cohort of cancer patients (Wheeler and Dolan 2012).

The following table (Table 1.1) illustrates the biggest GWA pharmacogenomics studies, showing the number of cell lines and the drugs that were utilized in each study.

Table 1.1. Largest pharmacogenomics studies that utilized LCLs for GWAS analysis.

Author (Year)	Journal	# of Cell Lines	# of Compounds
(Wheeler et al. 2013)	Pharmacogenomics J.	608	2
(Brown et al., 2014)	Pharmacogenomics J.	520	29
(Innocenti et al. 2009)	Cancer Chemother Pharmacol.	372	4
(Gamazon et al. 2010)	PNAS	343	5
(Stark et al. 2010)	Pharmacogenomics J	270	11
(Aksoy et al. 2009)	Pharmacogenet Genomics	240	3
(O'Donnell et al. 2010)	Pharmacogenet Genomics.	206	2
(Li et al. 2009)	Cancer Res	197	2
(Li et al. 2010)	Drug Metab Dispos.	194	3
(Huang et al. 2007a)	Molec. Therap.	176	4
(Huang et al. 2007b)	PNAS	176	2
(Fridley et al. 2011)	Pharmacogenet Genomics.	175	2
(Li et al. 2009)	PLoS One	174	2
(Peters et al. 2011)	Pharmacogenomics.	124	29
(Huang et al. 2011)	RNA biol.	107	3
(Wheeler et al. 2011)	PLoS One.	83	4
(Kulkarni et al. 2012)	BMC Med Genomics.	55	2
(Brown et al. 2012)*	Pharmacogenet Genomics	516	1

This table is sorted by the number of LCLs in each study. Studies with more the 50 LCLs and more than one drug were included. *This study represents the biggest study that evaluated one drug and included the largest number of LCLs.

D. UTILIZING LYMPHOBLASTOID CELL LINES IN TOXICOGENOMIC POPULATION-BASED IN VITRO TOXICITY SCREENING.

Pharmacogenomics studies were among the first to recognize the utility of LCLs as a model in testing pharmaceutical drugs. They mainly employed LCLs to discover genetic loci that are associated with either drug toxicity and/or survival prognosis. With the realization of the full potential of LCLs in addressing major gaps in current risk assessment, the present work aimed to extend their utility to (i) quantitatively assess and address population based toxicological effects or hazard of environmental contaminants, (ii) determine the extent of human inter-individual variability in chemical toxicity, (iii) identify susceptible sub-populations or races, (iv) understand the genetic determinants of the inter-individual variability, (v) generate testable hypotheses about toxicity pathways by leveraging genetic and genomic data from 1000 Genomes and HapMap Projects and (vi) use the data obtained from this research to build predictive in silico models.

1. Advantages and Limitations of Utilizing LCLs in Toxicogenomic Discovery

Our novel central hypothesis is that human lymphoblast cells derived from various human populations can be efficiently utilized to better understand the hazard and magnitude of human inter-individual variability in response to different environmental chemicals. LCLs are derived from various human populations representing diverse genetic ancestries and are publicly available for researchers from the 1000 Genomes Project (1000 Genomes Project Consortium et al. 2012). Because they are immortalized, LCLs represent a renewable source for repeated cultures. Thus, they provide a relatively cost-effective testing system with controlled

experimental manipulation. Unlike human epidemiological studies or *in vivo* animal experiments, the same cell lines could be treated with different concentrations of multiple compounds at the same time. Therefore, LCLs are amenable for high throughput screening and lack the confounders that are often present in *in vivo* studies. Genome-wide genotype (The SNP Consortium; International HapMap Project; The 1000 Genomes Project; Human Variation Panel dbGaP Access) and gene expression data (Gene Expression Omnibus; Gene Expression Omnibus b; CEU RNA-Seq Data), including next-generation sequencing (DNA and RNA-Seq) data, is publicly available for hundreds of established LCLs.

Like any toxicological model system, *in vitro* assessment using LCLs presents a number of technical challenges for extrapolation to humans, including lack of metabolism, and the inability to establish cell-cell interactions or evaluate chronic toxicity. The liver is the major site of xenobiotic metabolism and plays a central role in preventing accumulation of a wide range of compounds by converting them into a form suitable for elimination (Lerapetritou et al. 2009). Human lymphoblast cell lines are not hepatocytes and their main role is not meant to metabolize chemicals. Although lymphocytes do not have the metabolic capacity of the liver, or even that of freshly isolated hepatocytes, they do express a number of nuclear receptors, as well as most genes of phase I and II metabolism, and transporters (Siest et al. 2008). A comparison of the population-wide (250+ individuals of various races, ages and gender) variability in mRNA levels for several dozen liver-specific thyroid hormone-related genes between human liver (Schadt et al. 2008) and lymphoblast cell lines (Stranger et al. 2007) showed that most of the nuclear receptors and metabolism genes are expressed in lymphoblasts, albeit at a 10- to 100-times lower quantity than hepatocytes (unpublished data).

Potential confounders that affect the utility of LCLs include baseline growth rates, EBV copy numbers and ATP levels (Choy et al. 2008). While growth rate of LCLs was associated with chemotherapeutic-induced cytotoxicity in one study (Stark et al. 2010), it was not, in aggregate, associated with the cytotoxicity of 100+ chemicals in Lock et al. (2012). Altered apoptosis has been observed as a result of EBV transformation in LCLs with cancer drugs (Liu et al. 2004).

The immortalization process or EBV transformation has been observed to affect gene-expression and promoter-methylation profiles of majority of genes compared to primary B cells (Caliskan et al. 2011). However, the difference in expression levels between the primary and immortalized cells was small in magnitude (<1.5 fold). Moreover, the inter-individual variability in gene expression was the same between primary B cells and LCLs (Caliskan et al. 2011). Furthermore, many expression quantitative trait loci eQTLs observed in LCLs were observed in primary tissues like the liver, lung, and skin (Schadt et al. 2008; Bullaughey et al. 2009; Ding et al. 2010).

2. Toxicity Phenotype

With the recent shift in toxicology from *in vivo* to *in vitro*, hundreds of toxicity assays were developed to assess different cellular or biological endpoints. While those assays vary widely in their phenotypes and qualities, a good assay should be sensitive, robust, and able to quantitatively measure toxicity endpoint in a high throughput manner.

Generally, toxicity phenotypes can be classified to include general hazard, carcinogenicity, genotoxicity, developmental toxicity, reproductive toxicity, and chronic toxicity (Judson et al. 2009). With the Tox21 initiative, thousands of chemicals have been screened for

their potential hazard (Knudsen et al. 2013). More than 650 *in vitro* assays including biochemical assays, human cells and cell lines, and alternative models such as mouse embryonic stem cells and zebrafish embryo development have been developed and utilized for these chemicals, to understand their mechanisms and potential hazards (Knudsen et al. 2013).

The definition of “hazard” in chemical evaluation is broad and includes multiple potential definitions. A “hazardous” chemical evaluation could be derived from an acute and subchronic rodent study, or from material safety data sheets (Judson et al. 2009). To evaluate the *in vitro* potential hazard for a wide range of compounds, two essential things need to be properly selected: an appropriate end-point phenotype that represents general toxicity and is sensitive and broadly applicable to a variety of compounds, and a human relevant concentration range to determine the concentration for which the chemical will elicit toxicity. Determining the concentration at which chemical might elicit an endpoint is not enough, by itself, to quantify hazard. Because the *in vitro* assay endpoint does not incorporate metabolic clearance and plasma protein binding, ranking the chemicals by nominal assay concentrations might result in over- or under- estimation of the steady-state of a chemical (Rotroff et al 2010). However, incorporation of reverse dosimetry and exposure estimates when exploring concentration-response relationships for individual chemicals can aid tremendously in assessing human hazard and has been assessed by EPA and others (Rotroff et al. 2010; Judson et al. 2011) as a proposed approach.

The “CellTiter-Glo™ Luminescent Cell Viability Assay” is a commonly used assay in toxicity screening for cytotoxicity evaluation. It is a homogeneous “add-mix-measure” method to determine the number of viable cells in culture based on quantification of Adenosine Triphosphate (ATP) content (Promega). The quantity of ATP contents is a direct measurement of

the number of viable metabolically active cells (Promega). The assay results in cell lysis and a luciferase reaction. It is well-known that luciferase from fireflies or bacteria can be used to measure ATP (Fan and Wood 2007). A highly stabilized form of luciferase created by optimized strategies (Hall et al., 1998) is utilized in the CellTiterGlo assay, reacting with ATP to produce luciferyl adenylate, which is oxidized to produce a glow-like luminescent signal, among other products. The final signal has been shown to be proportional to the amount of ATP (Fraga et al. 2006). The luminescent signal, produced by the modified luciferase reaction, has a prolonged half-life (>5 hrs), making it resistant to a broad spectrum of detergents and suitable for batch-mode processing of multiple plates (Riss et al., 2006; Fan et al., 2005).

In a normal living cell, the generation of ATP requires harmonized interactions of many enzymes (Fan and Wood 2007). ATP concentrations are sustained upon equilibrium between consumption and demand enzyme pathways (Fan and Wood 2007). Upon cell death, the enzymes that generate ATP cease to function while the enzymes that consume ATP keep working, resulting in ATP consumption. Therefore, the concentration of intracellular ATP drops instantaneously upon cell death with a corresponding loss of luminescence when assayed using luciferase (Fan and Wood 2007).

CellTiter-GloTM Luminescent Cell Viability Assay has many advantages that make it a great choice for quantitative high throughput screening. First, the unique homogeneous format reduces pipetting errors that may be introduced during the multiple steps required by other ATP measurement methods, thus decreasing experimental inconsistencies. Second, the assay is extremely sensitive in the measurement of number of cells below the detection limits of standard colorimetric and fluorometric assays, reducing the number of cells required per assay. Third, the assay is fast (10 minutes) and flexible for different detection techniques, including automated

high-throughput protocols. Fourth, among all cytotoxicity assays, bioluminescent ATP assays have been the most widely used in high-throughput applications (Riss et al. 2006; Melnick et al. 2006). Finally, the assay is very robust with extremely stable luminescent and a long half-life (>5 hrs). The homogeneous assay procedure involves the addition of a single reagent (CellTiter-Glo™ Reagent) directly to cells cultured in serum-supplemented medium. Cell washing, removal of medium or multiple pipetting steps are not required (Promega).

Moreover, the CellTiter-Glo™ Luminescent Cell Viability Assay has been utilized and extensively evaluated for *in vitro* screening of cytotoxicity in high-throughput settings with proper evaluation of time points at the National Institutes of Health Chemical Genomics Center (NCGC) (Xia et al. 2008a). It was used to screen the NTP 1,408-compound library for cytotoxicity in 13 cell types (9 human, 2 rat, 2 mouse) (Xia et al. 2008b). The cell types originated from different tissues and included cell lines, cell strains, and primary cell populations (Tice et al. 2013). While some compounds were cytotoxic to all cell types, others were only toxic in some cell lines but not others. The results indicated that no single cell type would be universally informative for cytotoxicity or other endpoints, such as apoptosis (Xia et al. 2008a; Tice et al. 2013; Huang et al. 2008).

3. How can *in vitro* data inform systematic risk assessment?

Thousands of chemicals are being currently screened with hundreds of *in vitro* assays in the Tox21 and Toxcast projects. The predominant question is how to move forward from an *in vitro* endpoint to an overall comprehensive risk assessment? One way, as illustrated in Figure 1,

is to apply the findings of an *in vitro* study to a more traditional *in vivo* risk assessment. First, the hazard is established from a concentration-response *in vitro* assay. It is recommended that the *in vitro* testing be established from cell lines representing populations with sensitive individuals. Second, PK modeling is applied to derive the dose that results in the hazardous outcome. Third, uncertainty factors are applied to establish a human reference dose (Crump et al. 2010).

Another approach is to find the biological pathway altering dose (BPAD) for high-throughput risk assessment (HTRA). With this framework, a biological pathway is defined that is linked to adverse effects. Then, the *in vitro* concentration that perturbs the specified biological pathway is measured. Furthermore, PK modeling is applied to estimate the *in vivo* dose that could result in the hazardous concentration. From there, uncertainty and population variability estimates are incorporated to identify the protective exposure limit (Judson et al. 2011).

A third approach for toxicity assessment of *in vitro* assays is the tiered step-wise decision tree (Thomas et al. 2013). In this approach, chemicals are ranked based on their relative selectivity and the point of departure (POD) is established from the *in vitro* assays. Reverse toxicokinetic modeling (Rotroff et al. 2010; Wetmore et al. 2013; Wetmore et al. 2012) can then be applied to estimate the corresponding dose for the *in vitro* POD. The calculated external dose can then be computed to human exposure estimates to establish the margin of exposure (MOE). In the second tier, after selecting chemicals from tier I, *in vivo* animal testing is carried out with a focus on refining dose-response, PK evaluations, and other estimates. The third tier is characterized by traditional animal studies that are currently used to assess chemicals.

Regardless of the paradigm implemented to assess potential adverse health outcomes that are associated with chemical exposures from *in vitro* assays, one thing is standard and consistent among all: assessing human variability and identifying sensitive subpopulations is of key

importance. Testing human cell lines that represent various populations and identify sensitive subpopulations can substantially improve any risk assessment paradigm where decision making will be relied on actual estimated rather than mere statistical assumptions that may or may not be satisfied.

E. SPECIFIC AIMS

Quantitative assessment of the degree of inter-individual biological variability in the human population is a major aspect for properly evaluating potential human health hazards. A comprehensive characterization of human genome sequence variation is important for understanding inherited sources of variation in toxicity phenotypes. Genetic polymorphisms can have a profound influence on disease risk after drug or toxicant exposure. However, such characterization and assessment is difficult to quantitatively evaluate using current *in vivo* animal test systems or *in vitro* methods with established cell lines. The availability of genetically- and geographically-representative diverse and genetically-defined renewable sources of human cells, such as lymphoblasts from the International HapMap and 1000 Genomes projects, enables *in vitro* testing at the population level. As the focus of risk assessment processes shifts toward *in vitro* data, the quantitative assessment of inter-individual variability in response to chemicals, and an understanding of the underlying genetic causes are necessary for regulatory decisions to be based on scientific data rather than on default assumptions.

Previous pilot work showed that utilizing those lymphoblasts for *in vitro* toxicity assessment was very successful in assessing inter-individual variability and the molecular underpinnings of such variability. These novel findings form a solid foundation for this

proposal's central hypothesis that genetic variability and chemical toxicity can be assessed using a lymphoblast cell line-based in vitro model that serves as a model for the human population. To accomplish the objective of this application, this hypothesis was tested by pursuing the following specific aims:

Aim 1: To evaluate and assess the validity of an *in vitro*, genetically–anchored human population model system in assessing chemical toxicity and identifying candidate genetic susceptibility.

In this Aim, we hypothesized that genetic variability and chemical toxicity can be assessed using an *in vitro* model of lymphoblast cell lines that represent a human population. To demonstrate the feasibility of an *in vitro* model system to assess inter-individual and population-wide variability of chemical-induced toxicity phenotypes, cells from over 80 Centre d'Etude du Polymorphisme Humain (CEPH) cell lines were exposed to 12 concentrations of 240 environmental chemicals. The induction of caspase-3/7, indicative of apoptosis, was then assessed and intracellular levels of adenosine triphosphate (ATP) as a surrogate for cell number, was measured to evaluate cytotoxicity. We utilized the available dense genetic data to perform a Genome Wide Association Study (GWAS). We also assessed the validity of the *in vitro* genetics–anchored human model system by comparing it to similar human models and non-human models.

Aim 2: To quantitatively evaluate inter-individual variability for diverse environmental chemicals using high throughput large scale in-vitro model across genetically-defined genetically-diverse populations.

In this aim, we developed an *in-vitro* screening model that had sufficient power to quantitatively assess the potential human health hazards and inter-individual variability in chemical toxicity, identify susceptible sub-populations, understand the genetic determinants of the inter-individual variability, generate testable hypotheses about toxicity pathways by leveraging genetic and genomic data from 1000 Genomes and HapMap Projects, and use the data obtained from this research to build predictive *in silico* models. We screened 1104 lymphoblast cell lines from 9 ancestrally and geographically diverse populations representing 5 continents. Those cell lines were chosen based on availability of dense genetic information and were exposed to 180 diverse environmental chemicals at 8 different concentrations. We assessed cytotoxicity based on measuring intracellular levels of ATP.

Aim 3: To investigate remaining challenges of in vitro genetically-anchored population human model, such as the potential to screen complex mixtures.

This specific aim addressed remaining challenges that require further assessment to increase the utility of the information obtained from this model, including limited metabolic capacity of lymphoblasts and the potential to screen complex mixtures. We selected lymphoblast cell lines from 4 ancestrally and geographically diverse populations based on availability of genome sequence and basal RNA-seq information. The cell lines were exposed to 2 pesticide chemical mixtures, at 8 different concentrations. This design enabled us to investigate the utility of our model in assessing chemical mixtures.

F. REFERENCES

- EPA. 2012. Toxic Substances Control Act (TSCA). Available <http://www.epa.gov/oecaagct/lsc.html>
- National Research Council. 2006. Assessing the Human Health Risks of Trichloroethylene: Key Scientific Issues Washington, DC: The National Academies Press.
- Abbot A. 2005. Nobel laureates: close encounters. *Nature* 436(7048): 170-171.
- Anastas P, Teichman K, Hubal EC. 2010. Ensuring the safety of chemicals. *Journal of exposure science & environmental epidemiology* 20(5): 395-396.
- Judson R, Richard A, Dix D, Houck K, Elloumi F, Martin M, et al. 2008. ACToR--Aggregated Computational Toxicology Resource. *Toxicol Appl Pharmacol* 233(1): 7-13.
- Collins FS, Gray GM, Bucher JR. 2008. Toxicology. Transforming environmental health protection. *Science* 319(5865): 906-907.
- EPA. 2004. Intentional Human Dosing Studies for EPA Regulatory Purposes: Scientific and Ethical Issues: The National Academies Press.
- Andersen ME, Krewski D. 2009. Toxicity testing in the 21st century: bringing the vision to life. *Toxicol Sci* 107(2): 324-330.
- Plunkett LM, Kaplan AM, Becker RA. 2010. An enhanced tiered toxicity testing framework with triggers for assessing hazards and risks of commodity chemicals. *Regul Toxicol Pharmacol* 58(3): 382-394.
- Judson RS, Houck KA, Kavlock RJ, Knudsen TB, Martin MT, Mortensen HM, et al. 2010. In Vitro Screening of Environmental Chemicals for Targeted Testing Prioritization: The ToxCast Project. *Environ Health Perspect* 118(4): 485-492.
- Chiu WA, Euling SY, Scott CS, Subramaniam RP. 2011. Approaches to advancing quantitative human health risk assessment of environmental chemicals in the post-genomic era. *Toxicol Appl Pharmacol*: in press.
- Xia M, Huang R, Witt KL, Southall N, Fostel J, Cho MH, et al. 2008a. Compound cytotoxicity profiling using quantitative high-throughput screening. *Environ Health Perspect* 116(3): 284-291.
- Reif DM, Martin MT, Tan SW, Houck KA, Judson RS, Richard AM, et al. 2010. Endocrine profiling and prioritization of environmental chemicals using ToxCast data. *Environ Health Perspect* 118(12): 1714-1720.
- Zhu H, Rusyn I, Richard A, Tropsha A. 2008. Use of cell viability assay data improves the prediction accuracy of conventional quantitative structure-activity relationship models of animal carcinogenicity. *Environ Health Perspect* 116(4): 506-513.

Martin MT, Dix DJ, Judson RS, Kavlock RJ, Reif DM, Richard AM, et al. 2010. Impact of environmental chemicals on key transcription regulators and correlation to toxicity end points within EPA's ToxCast program. *Chem Res Toxicol* 23(3): 578-590.

US EPA. 2011. Advancing the Next Generation (NexGen) of Risk Assessment Available: <http://www.epa.gov/risk/nexgen/>

National Research Council. 2007. Toxicity testing in the 21st century: A vision and a strategy. Washington, DC: National Academies Press.

Crump KS, Chen C, Louis TA. 2010. The future use of in vitro data in risk assessment to set human exposure standards: challenging problems and familiar solutions. *Environ Health Perspect* 118(10): 1350-1354.

International Programme on Chemical S. 1993. Biomarkers and risk assessment concepts and principles. 9241571551. Geneva, Switzerland: World Health Organization.

International Programme for Chemical Safety. 2001. Guidance Document for the Use of Data in Development of Chemical-Specific Adjustment Factors (CSAFs) for Interspecies Differences and Human Variability in Dose/Concentration–Response Assessment. WHO/PCS/01.4.

Guyton KZ, Kyle AD, Aubrecht J, Cogliano VJ, Eastmond DA, Jackson M, et al. 2009. Improving prediction of chemical carcinogenicity by considering multiple mechanisms and applying toxicogenomic approaches. *Mutation research* 681(2-3): 230-240.

Hattis D. 1997. Human variability in susceptibility - how big, how often, for what responses to what agents. *Environmental Toxicology & Pharmacology* 4(3-4): 195-208.

National Research Council. 2009. Science and Decisions: Advancing Risk Assessment. Washington, DC: National Academies Press.

Zeise L, Bois FY, Chiu WA, Hattis D, Rusyn I, Guyton KZ. 2013. Addressing human variability in next-generation human health risk assessments of environmental chemicals. *Environ Health Perspect* 121(1): 23-31.

Sonich-Mullin C, Fielder R, Wiltse J, Baetcke K, Dempsey J, Fenner-Crisp P, et al. 2001. IPCS conceptual framework for evaluating a mode of action for chemical carcinogenesis. *Regulatory toxicology and pharmacology* : RTP 34(2): 146-152.

Schadt EE, Bjorkegren JL. 2012. NEW: network-enabled wisdom in biology, medicine, and health care. *Science translational medicine* 4(115): 115rv111.

Chen L, Tong T, Zhao H. 2008. Considering dependence among genes and markers for false discovery control in eQTL mapping. *Bioinformatics* 24(18): 2015-2022.

Emilsson V, Tholeifsson G, Zhang B, Leonardson AS, Zink F, Zhu J, et al. 2008. Genetics of gene expression and its effect on disease. *Nature* 452(7186): 423-428.

- Illig T, Gieger C, Zhai G, Romisch-Margl W, Wang-Sattler R, Prehn C, et al. 2010. A genome-wide perspective of genetic variation in human metabolism. *Nat Genet* 42(2): 137-141.
- Burton PR, Hansell AL, Fortier I, Manolio TA, Khoury MJ, Little J, Elliott P. 2009. Size matters: just how big is BIG?: Quantifying realistic sample size requirements for human genome epidemiology. *Int J Epidemiol.* 38(1):263-73.
- Manolio TA. 2010. Genomewide association studies and assessment of the risk of disease. *N Engl J Med* 363(2): 166-176.
- Cornelis MC, Agrawal A, Cole JW, Hansel NN, Barnes KC, Beaty TH, et al. 2010. The Gene, Environment Association Studies consortium (GENEVA): maximizing the knowledge obtained from GWAS by collaboration across studies of multiple conditions. *Genet Epidemiol* 34(4): 364-372.
- Schadt EE. 2009. Molecular networks as sensors and drivers of common human diseases. *Nature* 461(7261): 218-223.
- Iarmarcovai G, Sari-Minodier I, Chaspoul F, Botta C, De Meo M, Orsiere T, et al. 2005. Risk assessment of welders using analysis of eight metals by ICP-MS in blood and urine and DNA damage evaluation by the comet and micronucleus assays; influence of XRCC1 and XRCC3 polymorphisms. *Mutagenesis* 20(6): 425-432.
- Andrew AS, Burgess JL, Meza MM, Demidenko E, Waugh MG, Hamilton JW, et al. 2006. Arsenic exposure is associated with decreased DNA repair in vitro and in individuals exposed to drinking water arsenic. *Environ Health Perspect* 114(8): 1193-1198.
- Burton PR, Hansell AL, Fortier I, Manolio TA, Khoury MJ, Little J, Elliott P. 2009. Size matters: just how big is BIG?: Quantifying realistic sample size requirements for human genome epidemiology. *Int J Epidemiol.* 38(1):263-73.
- Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, McCarthy MI, Ramos EM et al. 2009. Finding the missing heritability of complex diseases. *Nature* 461, 747-753
- Rusyn I, Gatti DM, Wiltshire T, Kleeberger SR, Threadgill DW. 2010. Toxicogenetics: population-based testing of drug and chemical safety in mouse models. *Pharmacogenomics* 11(8): 1127-1136.
- Lexchin JR. 2005. Implications of pharmaceutical industry funding on clinical research. *The Annals of pharmacotherapy* 39(1): 194-197.
- Phillips EJ, Mallal SA. 2010. Pharmacogenetics of drug hypersensitivity. *Pharmacogenomics* 11(7): 973-987.
- La Thangue NB, Kerr DJ. 2011. Predictive biomarkers: a paradigm shift towards personalized cancer medicine. *Nature reviews Clinical oncology* 8(10): 587-596.

- Wheeler HE, Dolan ME. 2012. Lymphoblastoid cell lines in pharmacogenomic discovery and clinical translation. *Pharmacogenomics* 13(1): 55-70.
- Kuehn BM. 2008. 1000 Genomes Project promises closer look at variation in human genome. *JAMA : the journal of the American Medical Association* 300(23): 2715.
- Clarke. 2012. *The 1000 Genomes Project A History, Results and Tools*.
- Miller G. 1982. immortalization of human lymphocytes by Epstein-Barr virus. *The Yale journal of biology and medicine* 55(3-4): 305-310.
- Higaki T, Watanabe T, Tamatomi I, Tahara H, Sugimoto M, Furuichi Y, et al. 2004. Terminal telomere repeats are actually short in telomerase-negative immortal human cells. *Biol Pharm Bull* 27(12): 1932-1938.
- Qian Y, Yang L, Cao S. 2014. Telomeres and telomerase in T cells of tumor immunity. *Cell Immunol* 289(1-2): 63-69.
- Weiss LM, O'Malley D. 2013. Benign lymphadenopathies. *Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc* 26 Suppl 1: S88-96.
- Toussirot E, Roudier J. 2008. Epstein-Barr virus in autoimmune diseases. *Best practice & research Clinical rheumatology* 22(5): 883-896.
- Dreyfus DH, Liu Y, Ghoda LY, Chang JT. 2011. Analysis of an ankyrin-like region in Epstein Barr Virus encoded (EBV) BZLF-1 (ZEBRA) protein: implications for interactions with NF-kappaB and p53. *Virology journal* 8: 422.
- Pender MP, Csurhes PA, Pfluger CM, Burrows SR. 2012. CD8 T cell deficiency impairs control of Epstein--Barr virus and worsens with age in multiple sclerosis. *Journal of neurology, neurosurgery, and psychiatry* 83(3): 353-354.
- Ascherio A, Munger KL. 2010. Epstein-barr virus infection and multiple sclerosis: a review. *Journal of neuroimmune pharmacology : the official journal of the Society on NeuroImmune Pharmacology* 5(3): 271-277.
- Niedobitek G, Meru N, Delecluse HJ. 2001. Epstein-Barr virus infection and human malignancies. *Int J Exp Pathol* 82(3): 149-170.
- Epstein MA. 2001. Reflections on Epstein-Barr virus: some recently resolved old uncertainties. *The Journal of infection* 43(2): 111-115.
- Kawa K. 2000. Epstein-Barr virus--associated diseases in humans. *International journal of hematology* 71(2): 108-117.
- Maeda E, Akahane M, Kiryu S, Kato N, Yoshikawa T, Hayashi N, et al. 2009. Spectrum of Epstein-Barr virus-related diseases: a pictorial review. *Japanese journal of radiology* 27(1): 4-19.

- Henle W, Henle G. 1980. Epidemiologic aspects of Epstein-Barr virus (EBV)-associated diseases. *Annals of the New York Academy of Sciences* 354: 326-331.
- Kalla M, Hammerschmidt W. 2012. Human B cells on their route to latent infection--early but transient expression of lytic genes of Epstein-Barr virus. *Eur J Cell Biol* 91(1): 65-69.
- Amon W, Binne UK, Bryant H, Jenkins PJ, Karstegl CE, Farrell PJ. 2004. Lytic cycle gene regulation of Epstein-Barr virus. *Journal of virology* 78(24): 13460-13469.
- Smirnov DA, Morley M, Shin E, Spielman RS, Cheung VG. 2009. Genetic analysis of radiation-induced changes in human gene expression. *Nature* 459(7246): 587-591.
- Niu N, Qin Y, Fridley BL, Hou J, Kalari KR, Zhu M, et al. 2010. Radiation pharmacogenomics: a genome-wide association approach to identify radiation response biomarkers using human lymphoblastoid cell lines. *Genome Res* 20(11): 1482-1492.
- Simon JA, Lin F, Hulley SB, Blanche PJ, Waters D, Shiboski S, et al. 2006. Phenotypic predictors of response to simvastatin therapy among African-Americans and Caucasians: the Cholesterol and Pharmacogenetics (CAP) Study. *The American journal of cardiology* 97(6): 843-850.
- Mangravite LM, Wilke RA, Zhang J, Krauss RM. 2008. Pharmacogenomics of statin response. *Current opinion in molecular therapeutics* 10(6): 555-561.
- Wilke RA, Berg RL, Linneman JG, Zhao C, McCarty CA, Krauss RM. 2008. Characterization of low-density lipoprotein cholesterol-lowering efficacy for atorvastatin in a population-based DNA biorepository. *Basic Clin Pharmacol Toxicol* 103(4): 354-359.
- Medina RJ, O'Neill CL, Devine AB, Gardiner TA, Stitt AW. 2008. The pleiotropic effects of simvastatin on retinal microvascular endothelium has important implications for ischaemic retinopathies. *PloS one* 3(7): e2584.
- Morag A, Pasmanik-Chor M, Oron-Karni V, Rehavi M, Stingl JC, Gurwitz D. 2011. Genome-wide expression profiling of human lymphoblastoid cell lines identifies CHL1 as a putative SSRI antidepressant response biomarker. *Pharmacogenomics* 12(2): 171-184.
- Li L, Fridley B, Kalari K, Jenkins G, Batzler A, Safgren S, et al. 2008. Gemcitabine and cytosine arabinoside cytotoxicity: association with lymphoblastoid cell expression. *Cancer Res* 68(17): 7050-7058.
- Mitra AK, Crews K, Pounds S, Cao X, Downing JR, Raimondi S, et al. 2011. Impact of genetic variation in FKBP5 on clinical response in pediatric acute myeloid leukemia patients: a pilot study. *Leukemia* 25(8): 1354-1356.

The SNP Consortium Database. <http://snpdata.cshl.edu>.

International HapMap Project. www.hapmap.org.

The 1000 Genomes Project. www.1000genomes.org.

Human Variation Panel dbGaP Access. www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000211.v1.p1.

SCAN Database. www.scandb.org.

Gene Expression Omnibus (CEPH pedigrees GEO accession: GSE1485)
www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE1485.

Gene Expression Omnibus (Human Variation Panel GEO accession: GSE24277)
www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE24277.

CEU RNA-Seq Data. http://jungle.unige.ch/maseq_CEU60.
eQTL resources at the Pritchard Lab. <http://eqtl.uchicago.edu/Home.html>.

IOM (Institute of Medicine). 2007. *The Future of Drug Safety: Promoting and Protecting the Health of the Public*. Washington, DC: The National Academies Press.

Wheeler HE, Gamazon ER, Stark AL, O'Donnell PH, Gorsic LK, Huang RS, et al. 2013. Genome-wide meta-analysis identifies variants associated with platinum agent susceptibility across populations. *Pharmacogenomics J* 13(1): 35-43.

Innocenti F, Mirkov S, Nagasubramanian R, Ramirez J, Liu W, Bleibel WK, et al. 2009. The Werner's syndrome 4330T>C (Cys1367Arg) gene variant does not affect the in vitro cytotoxicity of topoisomerase inhibitors and platinum compounds. *Cancer chemotherapy and pharmacology* 63(5): 881-887.

Brown CC, Havener TM, Medina MW, Jack JR, Krauss RM, McLeod HL, et al. 2014. Genome-wide association and pharmacological profiling of 29 anticancer agents using lymphoblastoid cell lines. *Pharmacogenomics* 15(2): 137-146.

Gamazon ER, Huang RS, Cox NJ, Dolan ME. 2010. Chemotherapeutic drug susceptibility associated SNPs are enriched in expression quantitative trait loci. *Proc Natl Acad Sci USA* 107(20): 9287-9292.

Stark AL, Zhang W, Mi S, Duan S, O'Donnell PH, Huang RS, et al. 2010. Heritable and non-genetic factors as variables of pharmacologic phenotypes in lymphoblastoid cell lines. *Pharmacogenomics J* 10(6): 505-512.

Aksoy P, Zhu MJ, Kalari KR, Moon I, Pelleymounter LL, Eckloff BW, et al. 2009. Cytosolic 5'-nucleotidase III (NT5C3): gene sequence variation and functional genomics. *Pharmacogenetics and genomics* 19(8): 567-576.

- O'Donnell PH, Gamazon E, Zhang W, Stark AL, Kistner-Griffin EO, Stephanie Huang R, et al. 2010. Population differences in platinum toxicity as a means to identify novel genetic susceptibility variants. *Pharmacogenetics and genomics* 20(5): 327-337.
- Li L, Fridley BL, Kalari K, Jenkins G, Batzler A, Weinshilboum RM, et al. 2009. Gemcitabine and arabinosylcytosin pharmacogenomics: genome-wide association and drug response biomarkers. *PloS one* 4(11): e7765.
- Li F, Fridley BL, Matimba A, Kalari KR, Pellemounter L, Moon I, et al. 2010. Ecto-5'-nucleotidase and thiopurine cellular circulation: association with cytotoxicity. *Drug Metab Dispos* 38(12): 2329-2338.
- Huang RS, Kistner EO, Bleibel WK, Shukla SJ, Dolan ME. 2007a. Effect of population and gender on chemotherapeutic agent-induced cytotoxicity. *Mol Cancer Ther* 6(1): 31-36.
- Huang RS, Duan S, Bleibel WK, Kistner EO, Zhang W, Clark TA, et al. 2007b. A genome-wide approach to identify genetic variants that contribute to etoposide-induced cytotoxicity. *Proc Natl Acad Sci U S A* 104(23): 9758-9763.
- Fridley BL, Batzler A, Li L, Li F, Matimba A, Jenkins GD, et al. 2011. Gene set analysis of purine and pyrimidine antimetabolites cancer therapies. *Pharmacogenetics and genomics* 21(11): 701-712.
- Peters EJ, Motsinger-Reif A, Havener TM, Everitt L, Hardison NE, Watson VG, et al. 2011. Pharmacogenomic characterization of US FDA-approved cytotoxic drugs. *Pharmacogenomics* 12(10): 1407-1415.
- Huang RS, Johnatty SE, Gamazon ER, Im HK, Ziliak D, Duan S, et al. 2011. Platinum sensitivity-related germline polymorphism discovered via a cell-based approach and analysis of its association with outcome in ovarian cancer patients. *Clin Cancer Res* 17(16): 5490-5500.
- Wheeler HE, Gorsic LK, Welsh M, Stark AL, Gamazon ER, Cox NJ, et al. 2011. Genome-wide local ancestry approach identifies genes and variants associated with chemotherapeutic susceptibility in African Americans. *PloS one* 6(7): e21920.
- Kulkarni H, Goring HH, Diego V, Cole S, Walder KR, Collier GR, et al. 2012. Association of differential gene expression with imatinib mesylate and omacetaxine mepesuccinate toxicity in lymphoblastoid cell lines. *BMC medical genomics* 5: 37.
- Brown CC, Havener TM, Medina MW, Auman JT, Mangravite LM, Krauss RM, et al. 2012. A genome-wide association analysis of temozolomide response using lymphoblastoid cell lines shows a clinically relevant association with MGMT. *Pharmacogenetics and genomics* 22(11): 796-802.
- 1000 Genomes Project Consortium, Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, et al. 2012. An integrated map of genetic variation from 1,092 human genomes. *Nature* 491(7422): 56-65.

- Lerapetritou MG, Georgopoulos PG, Roth CM, Androulakis LP. 2009. Tissue-level modeling of xenobiotic metabolism in liver: An emerging tool for enabling clinical translational research. *Clin Transl Sci.* 2(3):228-37.
- Siest G, Jeannesson E, Marteau JB, Samara A, Marie B, Pfister M, et al. 2008. Transcription factor and drug-metabolizing enzyme gene expression in lymphocytes from healthy human subjects. *Drug Metab Dispos* 36(1): 182-189.
- Schadt EE, Molony C, Chudin E, Hao K, Yang X, Lum PY, et al. 2008. Mapping the genetic architecture of gene expression in human liver. *PLoS Biol* 6(5): e107.
- Stranger BE, Nica AC, Forrest MS, Dimas A, Bird CP, Beazley C, et al. 2007. Population genomics of human gene expression. *Nat Genet* 39(10): 1217-1224.
- Choy E, Yelensky R, Bonakdar S, Plenge RM, Saxena R, De Jager PL, et al. 2008. Genetic analysis of human traits in vitro: drug response and gene expression in lymphoblastoid cell lines. *PLoS Genet* 4(11): e1000287.
- Lock EF, Abdo N, Huang R, Xia M, Kosyk O, O'Shea SH, et al. 2012. Quantitative high-throughput screening for chemical toxicity in a population-based in vitro model. *Toxicol Sci* 126(2): 578-588.
- Liu MT, Chen YR, Chen SC, Hu CY, Lin CS, Chang YT, et al. 2004. Epstein-Barr virus latent membrane protein 1 induces micronucleus formation, represses DNA repair and enhances sensitivity to DNA-damaging agents in human epithelial cells. *Oncogene* 23(14): 2531-2539.
- Caliskan M, Cusanovich DA, Ober C, Gilad Y. 2011. The effects of EBV transformation on gene expression levels and methylation profiles. *Hum Mol Genet* 20(8): 1643-1652.
- Bullaughay K, Chavarria CI, Coop G, Gilad Y. 2009. Expression quantitative trait loci detected in cell-lines are often present in primary tissues. *Hum Mol Genet.*
- Ding J, Gudjonsson JE, Liang L, Stuart PE, Li Y, Chen W, et al. 2010. Gene expression in skin and lymphoblastoid cells: Refined statistical method reveals extensive overlap in cis-eQTL signals. *Am J Hum Genet* 87(6): 779-789.
- Judson R, Richard A, Dix DJ, Houck K, Martin M, Kavlock R, et al. 2009. The toxicity data landscape for environmental chemicals. *Environ Health Perspect* 117(5): 685-695.
- Knudsen T, Martin M, Chandler K, Kleinstreuer N, Judson R, Sipes N. 2013. Predictive models and computational toxicology. *Methods Mol Biol* 947: 343-374.
- Rotroff DM, Wetmore BA, Dix DJ, Ferguson SS, Clewell HJ, Houck KA, et al. 2010. Incorporating human dosimetry and exposure into high-throughput in vitro toxicity screening. *Toxicol Sci* 117(2): 348-358.

Judson RS, Kavlock RJ, Setzer RW, Hubal EA, Martin MT, Knudsen TB, et al. 2011. Estimating toxicity-related biological pathway altering doses for high-throughput chemical risk assessment. *Chem Res Toxicol* 24(4): 451-462.

Fan F, Wood KV. 2007. Bioluminescent assays for high-throughput screening. *Assay Drug Dev Technol* 5(1): 127-136.

Fraga H, Fernandes D, Novotny J, Fontes R, Esteves da Silva JC. 2006. Firefly luciferase produces hydrogen peroxide as a coproduct in dehydroluciferyl adenylate formation. *Chembiochem : a European journal of chemical biology* 7(6): 929-935.

Hall MP, Gruber MG, Hannah RR, Jennens-Clough ML, Wood KV. 1998. Stabilization of firefly luciferase using directed evolution. In: *Bioluminescence and Chemiluminescence, Perspectives for the 21st Century* (Roda A, Pazzagli M, Kricka LJ, Stanley PE, eds.), pp. 392–395. John Wiley & Sons, New York, 1998.

Riss TL, Moravec RA, O'Brien MA, Hawkins EM, Niles. 2006. A: Homogeneous multiwell assays for measuring cell viability, cytotoxicity, and apoptosis. In: *Handbook of Assay Development in Drug Discovery* (Minor LK, ed.), pp. 385–406. CRC Taylor & Francis, Boca Raton, FL, 2006.

Fan F, Butler B, Wood KV. 2005. A single-step bioluminescent assay for rapid detection and quantitation of viable microbial cells. In: *Bioluminescence and Chemiluminescence, Progress and Perspectives* (Tsuji A, Maeda M, Matsumoto M, Kricka LJ, Stanley PE, eds.), pp. 381–384. World Scientific Publishing Co. Pte. Ltd., Singapore, 2005.

Melnick JS, Janes J, Kim S, Chang JY, Sipes DG, Gunderson D, et al. 2006. An efficient rapid system for profiling the cellular activities of molecular libraries. *Proc Natl Acad Sci U S A* 103(9): 3153-3158..

Xia M, Huang R, Witt KL, Southall N, Fostel J, Cho MH, et al. 2008b. Compound cytotoxicity profiling using quantitative high-throughput screening. *Environ Health Perspect* 116(3): 284-291.

Tice RR, Austin CP, Kavlock RJ, Bucher JR. 2013. Improving the human hazard characterization of chemicals: a Tox21 update. *Environ Health Perspect* 121(7): 756-765.

Huang R, Southall N, Cho MH, Xia M, Inglese J, Austin CP. 2008. Characterization of diversity in toxicity mechanism using in vitro cytotoxicity assays in quantitative high throughput screening. *Chem Res Toxicol* 21(3): 659-667.

Thomas RS, Philbert MA, Auerbach SS, Wetmore BA, Devito MJ, Cote I, et al. 2013. Incorporating new technologies into toxicity testing and risk assessment: moving from 21st century vision to a data-driven framework. *Toxicol Sci* 136(1): 4-18.

Wetmore BA, Wambaugh JF, Ferguson SS, Li L, Clewell HJ, 3rd, Judson RS, et al. 2013. Relative impact of incorporating pharmacokinetics on predicting in vivo hazard and mode of action from high-throughput in vitro toxicity assays. *Toxicol Sci* 132(2): 327-346.

Wetmore BA, Wambaugh JF, Ferguson SS, Sochaski MA, Rotroff DM, Freeman K, et al. 2012. Integration of dosimetry, exposure, and high-throughput screening data in chemical toxicity assessment. *Toxicol Sci* 125(1): 157-174.

CHAPTER 2: QUANTITATIVE HIGH-THROUGHOUT SCREENING FOR CHEMICAL TOXICITY IN A POPULATION-BASED IN VITRO MODEL

Eric F. Lock¹, Nour Abdo¹, Ruili Huang², Menghang Xia², Oksana Kosyk¹,
Shannon H. O'Shea¹, Yi-Hui Zhou¹, Alexander Sedykh¹, Alexander Tropsha¹,
Christopher P. Austin², Raymond R. Tice³, Fred A. Wright^{1,*}, and Ivan Rusyn^{1,*}

¹University of North Carolina, Chapel Hill, NC 27599;

²NIH Chemical Genomics Center, National Human Genome Research Institute, National
Institutes of Health, Bethesda, MD 20892;

³Division of the National Toxicology Program, National Institute of Environmental Health
Sciences, Research Triangle Park, NC 27711

Mostly reproduced from Oxford Journals.

This is a pre-copy-editing, author-produced PDF of an article accepted for publication in *Toxicol Sci* following peer review. The definitive publisher-authenticated version *Toxicol Sci*. Apr 2012;

126(2): 578–588. 10.1093/toxsci/kfs023. PMID: PMC3307611 is available online at:

<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3307611/>

A. ABSTRACT

A shift in toxicity testing from *in vivo* to *in vitro* may efficiently prioritize compounds, reveal new mechanisms, and enable predictive modeling. Quantitative high-throughput screening (qHTS) is a major source of data for computational toxicology, and our goal in this study was to aid in the development of predictive *in vitro* models of chemical-induced toxicity, anchored on inter-individual genetic variability. Eighty-one human lymphoblast cell lines from 27 Centre d'Etude du Polymorphisme Humain (CEPH) trios were exposed to 240 chemical substances (12 concentrations, 0.26 nM-46.0 uM) and evaluated for cytotoxicity and apoptosis. qHTS screening in the genetically-defined population produced robust and reproducible results, which allowed for cross-compound, -assay and -individual comparisons. Some compounds were cytotoxic to all cell types at similar concentrations, whereas others exhibited inter-individual differences in cytotoxicity. Specifically, the quantitative high-throughput screening in a population-based human *in vitro* model system has several unique aspects that are of utility for toxicity testing, chemical prioritization, and high-throughput risk assessment. First, standardized and high-quality concentration-response profiling, with reproducibility confirmed by comparison with previous experiments, enables prioritization of chemicals for variability in inter-individual range in cytotoxicity. Second, genome-wide association analysis of cytotoxicity phenotypes allows exploration of the potential genetic determinants of inter-individual variability in toxicity. Furthermore, highly significant associations identified through the analysis of population-level correlations between basal gene expression variability and chemical-induced toxicity suggest plausible mode of action hypotheses for follow up analyses. We conclude that as the improved resolution of genetic profiling can now be matched with high-quality *in vitro* screening data, the evaluation of the toxicity pathways and the effects of genetic diversity are now feasible through the use of human lymphoblast cell lines.

B. INTRODUCTION

The “Registration, Evaluation, Authorisation and Restriction of Chemicals” (REACH) regulations in Europe and Toxic Substances Control Act reform activities in the US are creating substantial pressure to develop improved methods for evaluating potential chemical hazards (Plunkett et al., 2010). Current chemical safety evaluation (National Research Council, 2007) relies on *in vivo* animal testing. In Europe alone, it is expected that 100,000+ chemicals will require new safety data, yet the worldwide capacity to evaluate chemicals for the most animal-intensive *in vivo* tests is 200–300 chemicals each year (Hartung and Rovida, 2009).

In the US, the Tox21 program (Collins et al., 2008) is a collaborative initiative of four government agencies. This effort leads the field in its use of a broad spectrum of *in vitro* assays, many in quantitative High Throughput Screening (qHTS) format (Inglese et al., 2006), to screen thousands of environmental chemicals for their potential to affect biological pathways that may result in human disease (Xia et al., 2008). Such data on toxicologically relevant *in vitro* endpoints can assist in decision-making (Reif et al., 2010), serve as predictive surrogates for *in vivo* toxicity (Martin et al., 2010; Zhu et al., 2008), and generate testable hypotheses on the mechanisms (Xia et al., 2009).

Another important consideration in assessing the potential human health hazard is the degree of inter-individual biological variability in the human population (National Research Council, 2008). A comprehensive characterization of human genome sequence variation is important for understanding observed inherited variation in toxicity phenotypes. Indeed, genetic polymorphisms can have a profound influence on disease risk after drug or toxicant exposure (Harrill et al., 2009), yet these factors are difficult to quantitatively evaluate using current *in vivo* animal test systems or established cell lines (Rusyn et al., 2010). The availability of genetically-

diverse, genetically-defined renewable sources of human cells, such as lymphoblasts from the International HapMap (International HapMap Consortium, 2005) and 1000 Genomes (Durbin et al., 2010) projects, enables in vitro testing at the population scale. As the risk assessment process shifts towards in vitro data, the quantitative assessment of inter-individual variability in responses to chemicals, as well as an understanding of the underlying genetic causes, are needed so that regulatory decisions can be based on data rather than default assumptions.

To demonstrate the feasibility of an in vitro model system to assess inter-individual and population-wide variability of chemical-induced toxicity phenotypes, we exposed cells from over 80 CEPH cell lines (O'Shea et al., 2011) to 3 concentrations of 14 environmental chemicals, and assessed induction of caspase-3/7, indicative of apoptosis, and cytotoxicity, based on measuring intracellular levels of adenosine triphosphate (ATP) as a surrogate for cell number. This study showed that an in vitro genetics-anchored human model system can be utilized in a population-level screen for chemical toxicity, with the potential to identify candidate genetic susceptibility factors for further study. As a next step, we report here on a larger-scale population-based qHTS screening using hundreds of compounds and covering a more comprehensive range of concentrations. The quantitative assessment of inter-individual variability in response at this scale demonstrates the potential of this methodology for toxicity screening, hazard evaluation and exploration of genetic determinants of susceptibility.

C. MATERIALS AND METHODS

Experimental Design

Chemicals. A sub-set (240 compounds) of the National Toxicology Program's 1,408 chemical library (Xia et al., 2008) was used in these experiments. See Table 2.1 for a complete list of chemicals used in these experiments. Chemicals were dissolved with dimethyl sulfoxide (DMSO) into 12 different stock concentrations ranging from 56.5nM to 10 mM and were aliquoted to 1536-well plate format via pin tool (Kalypsys, San Diego, CA, USA). The final concentration ranges from 0.26 nM to 46.08 uM in the assay plates. The negative control was DMSO at 0.5% v/v; the positive control was staurosporine at the tested concentration range.

Cell lines. A set of 81 immortalized lymphoblastoid cell lines was acquired from Coriell Cell Repositories (Camden, NJ, USA). The 81 cell lines were from HapMap Consortium's Centre d'Etude du Polymorphisme Humain (CEPH) panel and consisted of 27 trios (father, mother and a child). Screening was conducted in 3 batches and cell lines were randomly divided into batches without regard to family structure. Cells were cultured at 37°C with 5% CO₂ in suspension in flasks with upright position in RPMI 1640 media (Gibco, Carlsbad, CA, USA) supplemented with 15% fetal bovine serum (HyClone, South Logan, UT, USA) and 0.1% penicillin-streptomycin (Gibco). Media was changed every 3 days. Cell counts and viability were assessed prior to chemical treatment using Cellometer Auto T4 Plus (Nexcelom Bioscience, Lawrence, MA, USA). Cells were grown to a concentration up to 10⁶ cells/mL, volume of at least 100 mL, and viability of >85% before treatment. After centrifugation, the cells were re-suspended in fresh media. The cell suspension was filtered through a 40 um nylon cell strainer (BD Biosciences, Durham, NC, USA). Cell stock was diluted with fresh media to final concentrations of 3-4x10⁵cells/mL, and plated into a tissue-culture treated 1,536-well white/solid

bottom assay plates (Grenier Bio-One North America, Monroe, NC, USA) at 2000 cells/5 μ L/well using a flying reagent dispenser (Aurora Discovery, Carlsbad, CA, USA). To increase the robustness of the data and evaluate reproducibility, each cell line was seeded on multiple plates (6 plates except for 2 cell lines where 5 plates were seeded) so that each compound was screened in each cell line on 2-3 plates (chemicals were randomly divided in half to enable screening of 120 compound x 12 concentrations on each plate).

Cytotoxicity and Caspase-3/7 assays. Two assays were chosen to evaluate cytotoxicity according to the manufacturer's protocols. Cell-Titer-Glo[®] Luminescent Cell Viability (Promega Corporation, Madison, WI, USA) assay was used to assess intra-cellular ATP concentration, a marker for cytotoxicity, 40 hours post treatment. Caspase-Glo[®] 3/7 (Promega) was used to assess activity of caspase-3/7, a marker of apoptosis, 16 hours post treatment. These assays were selected based on their utility for in vitro screening of cytotoxicity in cell type- (Xia et al., 2008) and individual-independent (Choy et al., 2008) manner. Time points were selected based on previous experiments at NCGC (Xia et al., 2008). A ViewLux plate reader (PerkinElmer, Shelton, CT, USA) was used to detect luminescent intensity in each well for both assays. Data is publicly available from PubChem (AIDs: 588812 and 588813).

Data Processing

Response Normalization & Curve Fitting. Data was normalized relative to the positive/negative controls and corrected as detailed elsewhere (Xia et al., 2008). Concentration-response titration points were fitted to a Hill equation for each chemical. Chemicals were classified into 3 categories based on their concentration-response curves: active, non-active, and inconclusive (Huang et al., 2008; Xia et al., 2008). Specifically, in data from cytotoxicity assay

the curve classes -1.1, -1.2,-2.1 were classified as “active,” any positive curve class as “non-active,” and others as “inconclusive.” For data from caspase-3/7 assay, curve classes 1.1, 1.2, 2.1 were classified as “active,” any negative curve class as “non-active,” and others as “inconclusive.”

CurveP. To evaluate the cytotoxic potency of each compound, we calculated a “curve P” value for each compound-cell line pair. Curve P is defined as the lowest concentration which showed a consistent deviation from the baseline response and derived as detailed in (Sedykh et al., 2011). It can be regarded as a close approximation for the point of departure. Curve P was derived for all compounds even if little or no toxicity was observed. For the latter compounds, to enable the follow-up statistical analyses, the curve P was assigned to a concentration of 50 μ M. Batch effects were adjusted using the ComBat method (Johnson et al., 2007).

Data Analysis

Assessing variability across individual, chemical, and assay. The Pearson correlation coefficient (r) between pairs of replicate plates was used to assess experimental reproducibility. For this analysis, two replicate plates were randomly selected for each chemical and cell line pair (240 chemicals x 81 cell lines = 19,440 total replicate pairs sampled).

Kruskal-Wallis ANOVA (Kruskal and Wallis, 1952) was used to assess the significance of a cell line effect (versus experimental effect) in curve P for each chemical. The Benjamini-Hochberg FDR (Johnson et al., 2007) was used to correct for multiple comparisons. To measure potential confounding with basal metabolic rate, the Spearman rank correlation coefficient between curve P and the average ATP level in DMSO-treated cells was computed for each chemical. The Spearman (rank) correlation between the average curve P value for the

cytotoxicity assay and the average curve P value for the apoptosis assay for each chemical was computed to measure an overall relationship between the two assays. Furthermore, within each chemical, the correlation between the two assays across cell lines (averaged over replications) was computed separately. For both assays, chemical-by-chemical correlation heatmaps were used to identify clusters of chemicals with similar response across cell lines. The order of the chemicals in these heatmaps was determined by complete-linkage distance clustering.

All computations, graphs and heatmaps used the R programming environment for statistical computing and graphics (2.10.0, R Development Core Team, Vienna, Austria).

Dose-response for populations and individuals. For the ATP assay data for progesterone, a four-parameter logistic model was fit to the assay vs. dose data for each cell line, using maximum likelihood and the `optim` routine in R. The model can be written

$assay = f(dose) + \varepsilon$, where

$f(dose) = \min + (\max - \min) \left(\frac{\exp(\beta_0 + \beta_1 dose)}{1 + \exp(\beta_0 + \beta_1 dose)} \right)$, $\varepsilon \sim N(0, \sigma^2)$, and $\{\min,$

$\max, \beta_0, \beta_1, \sigma^2\}$ are cell line-specific parameter vectors. For a negative dose-response relationship, EC_{10} is the dose for which

$$\exp(\beta_0 + \beta_1 dose) / (1 + \exp(\beta_0 + \beta_1 dose)) = 0.9.$$

The variation in the EC_{10} estimates was used as illustrative of population variation in true EC_{10} values, although additional sampling variation underlies each EC_{10} estimate. An overall logistic dose-response curve was fit to the aggregated data across all individuals.

Assessing heritability and genetic associations. Heritability calculations were used to determine overall familial effects among the 27 CEPH trios for each chemical, on both assays. Calculations were motivated by the mid-parent regression model $y = \beta_0 + \beta_1(a_p + a_m) + \varepsilon$,

where y is the child's response, a_p is the father's response, a_m is the mother's response and ϵ is an error term. A likelihood ratio significance test is then based on the heritability h^2 : the variability in response due to shared genetics as a proportion of total variability in response. For this analysis, curve P values for each chemical were quantile-normalized to the standard Gaussian distribution.

To measure genotype-toxicity relationships, genome-wide association studies (GWAS) were performed in R using the GenABEL package (Aulchenko et al., 2007). Phase III genotype data, on approximately 1.4×10^6 single nucleotide polymorphisms (SNPs), was obtained for each cell line from the International HapMap Project (International HapMap Consortium, 2005). GWAS was performed for each chemical on both assays, with quantile normalized curve P values as the response phenotype. The significance of an association between a given SNP and the response was measured using a likelihood-based score test (Schaid et al., 2002) (qtscore in genABEL). For our initial screen, the familial trio relationships were not used for the analysis, due to the low evidence for overall heritability, on the grounds that methods such as transmission disequilibrium testing would reduce power, and with the intent to follow any significant findings with further testing. LocusZoom (Pruim et al., 2010) was used to visualize the genomic context for suggestive loci determined by GWAS.

RNA-Seq expression vs. toxicity assays. The 42 cell lines in common between Montgomery et al. (Montgomery et al., 2010) and the present study were matched with HapMap IDs, using RNA-Seq tag counts mapped to the genome as previously described for 20,000 genes (Zhou et al., 2011). For computational efficiency, simple read proportions consisting of number of tag counts per gene divided by the mapped library size (Zhou et al., 2011) were used in linear regression as predictors for the cytotoxicity assays. FDR q-values were then obtained for the

entire set of genes and chemicals, using `p.adjust` in R. For the caspase assay, ~5,000 genes were determined to have at least one chemical with $q < 0.01$, and these genes were retained for clustering. Hierarchical clustering with average linkage was performed directly on the FDR q -values using the `heatmap` function in R.

D. RESULTS

qHTS screening in a population of human lymphoblasts yields robust and reproducible data

Screening was conducted in a 1,536-well plate format using a robotic system. The 81 cell lines were randomly sub-divided into 3 batches and each line was screened against 240 chemical substances (see Table 2.1 for a complete list) at 12 concentrations (0.26 nM-46.0 uM). Each 1,536-well plate contained one cell line exposed to 120 chemicals accompanied by concurrent vehicle (DMSO) and positive controls. To increase the robustness of the data, duplicates or triplicates of each plate were run. Assays for intracellular ATP content and caspase-3/7 activity were used based on their utility for in vitro screening of cytotoxicity and apoptosis, respectively, in cell type- and individual-independent manner (Choy et al., 2008; Xia et al., 2008). A combination of the two assays allows for the role of apoptosis in the cytotoxicity response to be evaluated (Shi et al., 2010).

Several metrics were used to evaluate the reproducibility of the toxicity phenotypes. First, the concentration-response curve class (Parham et al., 2009) was identical across replicate plates 95.2% of the time for cytotoxicity and 94.1% for apoptosis. Second, the pair-wise Pearson correlation among replicate plate pairs using $\log(AC_{50})$ values for the compounds with active curve classes for the cytotoxicity and apoptosis assays was $r=0.99$ and $r=0.98$, respectively. Third, to evaluate the effects correlation for all compounds, we calculated a “curve P” value, the lowest concentration which showed a consistent deviation from the baseline response (Sedykh et al., 2011), which can be regarded as a close approximation for the Lowest Observed Adverse Effect Level. For chemicals exhibiting no effect across the concentrations tested, the curve P was assigned to 50 uM to enable straightforward statistical analyses. The pair-wise correlation among

replicate plates of the log(curve P) values was equally high ($r[\text{cytotoxicity}] = 0.91$, $r[\text{apoptosis}] = 0.95$) when all compounds were included (Fig. 2.1.a, 2.1.b). Finally, there were 8 duplicates among the compounds screened. High concordance in median and range of responses for these was observed (Fig. 2.1.c, 2.1.d).

Range in cytotoxicity across the chemicals

The chemicals selected for screening were a subset of 1,408 compounds previously tested in one or more traditional toxicological assays, and had been profiled for cytotoxicity and caspase-3/7 induction by NTP and NCGC using qHTS (Xia et al., 2008) in (i) 13 human and rodent cells derived from liver, blood, kidney, nerve, lung, skin; and (ii) 26 human lymphoblast cells (data available from PubChem AID: 963-989). Of these, 240 compounds that were clearly active in those experiments were selected for the current study (iii).

Comparison of the cytotoxicity average log(curve P) from the current study showed high concordance with that in panels (i) and (ii), see above. Pair-wise correlation analysis for the 240 chemicals across three data sets was highly significant ($p < 0.0001$). High correlation ($r = 0.87$; rank correlation = 0.83) was observed between lymphoblast panels (ii) and (iii), while the correlations with the diverse panel (i) were moderately high ($r = 0.74$ or 0.75 ; rank correlation = 0.72 or 0.75 with (ii) and (iii), respectively). Together, the results indicate high external reproducibility for this measurement of cytotoxicity and, importantly, the potential utility of lymphoblast cell lines as a toll for population-based toxicity screening.

Inter-individual variability in response across cell lines

In contrast to the highly invariant reproducible results found within individual cell lines, the chemicals induced a wide range of responses among the lymphoblast lines. The percentage of compounds classified as active in the cytotoxicity assay varied from 28% to 56% (Fig. 2.2a); an equally broad range of activity (i.e., 24 to 45%) was seen in the caspase-3/7 assay (Fig. 2.2b). Among actives, a wide range of potency, assessed from the curve P, was observed for each cell line in both assays (Fig. 2.2c,d).

Some chemicals were classified as active for cytotoxicity and caspase-3/7 induction in all of the lymphoblast lines, while others were not active for either endpoint (Fig. 2.3a,b). In both assays, most chemicals were active in some cell lines while not active in others, indicative of inter-individual (cell line) variability in response. The significant correlation (rank correlation=0.77; $p=2.2E-16$; all compounds tested) between the chemical's average curveP for cytotoxicity and caspase-3/7 (Fig. 2.3c) indicates the primary cause of cell death for these compounds is most likely via apoptosis. A heatmap shows correlations between average $\log(\text{curveP})$ for all chemicals in both assays (Fig. 2.3d). Clusters of chemicals with highly concordant responses across cell lines were evident for cytotoxicity, apoptosis, or both phenotypes. A significant ($FDR < 5\%$) correlation between responses in cytotoxicity and apoptosis assays was observed for most of the compounds screened.

Inter-individual variability in cytotoxicity was visualized using box plots of $\log(\text{curveP})$ for each chemical (Fig. 2.4a,b). Although median cytotoxicity differed between chemicals tested, inter-individual variability was observed even for the most active chemicals. Variance-components heritability testing for each chemical/assay showed that none of the derived h^2

statistics was significant after adjusting for multiple comparisons, an observation which was confirmed using mid-parent assays' values compared to those of the offspring (data not shown).

Inter-individual (between cell lines) vs. experimental (between replicates) variability for each chemical was evaluated using Kruskal-Wallis ANOVA (Kruskal and Wallis, 1952). Most chemicals show a significant ($FDR < 5\%$) cell line effect (Fig. 2.4c,d). It has been suggested that differences in chemical's toxicity among lymphoblast lines could be partly attributed to differences in baseline growth rate and metabolic status (Choy et al., 2008). Correcting for these measurements reduces effect correlation that would otherwise make responses across chemicals appear more similar. We therefore normalized for control levels of intracellular ATP (e.g., metabolic activity) and basal activity of caspase-3/7, as well as for the response of the positive control cytotoxicant. In addition, we directly assessed for each chemical whether the basal metabolic rate, an endpoint which correlates closely with the growth rate (Choy et al., 2008), significantly correlated with cytotoxicity. Approximately 80% and 90% of chemicals (Fig. 2.4c,d; black dots) exhibited no correlation ($FDR > 0.05$) between basal metabolic rate (ATP level in vehicle-treated cells) and cytotoxicity or apoptosis, respectively, across the cell panel.

Assessing relationships between cytotoxicity and genotype

With variability among cells from different individuals demonstrated, we then asked if we could identify genetic loci responsible, utilizing toxicity phenotypes as quantitative traits and publicly available genotypes (International HapMap Consortium, 2005) (Fig. 2.5). The top two plots in Figure 5 show p-values for the most significant SNP associated with cytotoxicity (Fig. 2.5a) or induction of caspase-3/7 (Fig. 2.5b) for each chemical. The inset shows a plot of $-\log_{10}(\text{p-values})$ for SNP-endpoint associations for the selected chemicals. Progesterone had the

lowest p-value SNPs on chromosome 9, while Guggulsterones Z (4,17(20)-pregnadiene-3,16-dione, z-isoform) exhibited many suggestive associations on chromosome 6p. Fig. 2.5c,d provide a zoomed-in view of the genomic context for these suggestive regions.

Progesterone was not highly cytotoxic, yet showed an appreciable degree of inter-individual variability in curve P values (Fig. 2.5c inset). A characteristic pattern of SNPs with low p-values in linkage disequilibrium is evident in a ~300 kb region containing two genes, structural maintenance of chromosomes protein 5 (SMC5) and MAM domain containing 2 (MAMDC2). Guggulsterone Z, a bioactive constituent of resinous sap from *Commiphora mukul*, is a farnesoid X receptor antagonist and is used widely as a nutraceutical. It is known to suppress expression of anti-apoptotic genes, promote apoptosis, and inhibit NF- κ B (Shishodia and Aggarwal, 2004). In our study, it was moderately active in inducing caspase-3/7 (Fig. 2.5d inset) and exhibited inter-individual variability. A narrow 100 Kb region on chromosome 6p, containing the gene human immunodeficiency virus type I enhancer binding protein 1 (HIVEP1), shows association with the apoptosis phenotype.

Dose-response for populations and individuals

The availability of cytotoxicity screens on 80+ individuals, with the assays performed under controlled conditions, enables sensitive investigation of variation in individual dose-response profiles (National Research Council, 2008). This concept is illustrated in Fig.2.6a, in which the progesterone ATP assay values are shown in grey for each concentration for all individuals. Separate logistic curve fits were performed, providing for each individual cell line an “effective concentration 10%” (EC_{10}), the estimated concentration at which the response deviates by at least 10% from the control baseline, and these are shown as a histogram. The

mean of these EC₁₀ values offers a population-wide summary of the activity (e.g., cytotoxicity, caspase-3/7) of a chemical, and is very similar to the EC₁₀ produced when the data are first pooled for all individuals and then fit using a single dose-response curve (red dashed curve in Fig.2.6a). However, aggregation across the population ignores the variability in toxic susceptibility, and the EC₁₀ estimated 5th percentile may be used to illustrate the concept of a “vulnerable” sub-population.

Defining mode-of-action chemical-perturbed pathways

Gene expression data forms another rich source of publicly available data, which can be matched with cytotoxicity profiles to provide further evidence of toxicity pathway activity. Many of the HapMap cell lines have been profiled for expression in a number of studies, including highly sensitive RNA-Seq profiling (Montgomery et al., 2010). For the 42 cell lines for which RNA-Seq data are publicly available, expression values for each of ~20,000 genes were compared to the caspase-3/7 and cytotoxicity assay results, with a number of highly-significant associations. A heatmap of clustering performed on FDR q-values (Fig.2.6b) shows striking patterns of gene-chemical relationships, with much of the structure resolving into distinct sets of genes associated with sets of chemicals. The results for progesterone are shown as a highly specific subgroup, with lymphoblast cytotoxicity for several chemicals being significantly associated with background RNA levels for 6 transcripts and several microRNAs.

Comparing cytotoxicity of LCLs across different cell lines

We wanted to test the hypothesis that if a chemical causes toxicity, it will have a similar toxic effect on different cell lines from different tissues or species. To test this hypothesis we compared the cytotoxicity results obtained from specific aim 1 to another two experiments done by other researchers that utilized the same chemicals with the same assay in different cell lines. The first experiment, labeled twins study (NCGC-U Penn), was a study that exposed lymphoblast cell lines from pairs of twins to the same set of chemicals (unpublished data). The other experiment was collaboration between NCGC and NTP, and exposed 13 human and rodent cell types derived from six common targets of xenobiotic toxicity (liver, blood, kidney, nerve, lung, skin) to the same set of chemicals (Xia et al. 2008a).

A pairwise correlation between the cytotoxic response (curve P) of common screened chemicals across our 81 cell lines in our experiment, paired twin lymphoblast cell lines in U Penn study, and the 13 cell lines in the NTP study showed significant correlation (Spearman rank correlation of 0.83 and 0.75 respectively) (Fig 2.7).

We picked the 30 most toxic chemicals in our panel and plotted the range of toxicity expressed by Curve P across cell lines for each study in a box and whisker plot shown in (Fig 2.8). While the median of cytotoxicity was similar in all the three studies, the range across tested cell lines was largest among the 13 cell type study coming from rodents and humans. Both LCL studies had similar range of cytotoxicity (Fig 2.8).

E. DISCUSSION

New paradigms for the rapid and accurate evaluation of the potential health hazard from environmental chemicals are needed, given the large number of environmental chemicals to be evaluated, and the high cost and low throughput of traditional toxicity testing approaches (Collins et al., 2008). Development of *in vitro* toxicity tests that can be utilized in a tiered framework is necessary, feasible and consistent with the needs of scientifically-rigorous high-throughput risk assessment (Kavlock et al., 2009). A particular challenge in developing such next-generation toxicity testing schemata is the assessment of differential susceptibility among individuals. The results presented here provide proof of principle of such a testing system, demonstrating the feasibility and utility of screening a panel of cells from genetically diverse individuals, whereby both population-wide and individual responses can be evaluated.

The *in vitro* toxicity screening paradigm detailed here has focused on a population-based cell culture model, an approach that affords several key benefits compared to collections of unrelated cell lines from different species and tissues (Xia et al., 2008). Our results show that many chemicals exhibit inter-individual variation in induction of toxicity and this information is crucial for chemical testing prioritization. This screening paradigm also provides quantitative data on population-wide variability in toxicity which may be used to establish data-driven uncertainty estimates when extrapolating from *in vitro* data to potential *in vivo* toxicity (Judson et al., 2011). Even though the data collected herein is on a limited population (81 individuals), it is immediately interpretable for ranking and prioritizing chemicals. For example, a population-based view of dose-response is an important concept that directly addresses the issue of sub-populations (National Research Council, 2008); however, actual experimental data-driven implementation has been limited. We reason that the population-based concentration-response in

in vitro qHTS data allows for the development of models to estimate in vitro point-of-departure and safety/uncertainty factors (Crump et al., 2010), because variation between genetically-defined/diverse cell lines may be treated as reflective of that among individuals. The recognition of underlying genetic causes may further enhance extrapolation and understanding of the shape of the dose-response relationships. In addition, the data may be used to explore potential differences/similarities in modes of action between chemicals on the population-wide level.

By combining toxicity data with publicly available genetic information, such as that provided by the HapMap (International HapMap Consortium, 2005), 1000 Genomes (Durbin et al., 2010), and public RNA sequencing projects (Montgomery et al., 2011), it is possible to probe the contribution of genomics to toxicity phenotypes. Such an approach represents a substantial savings of cost and time, capitalizing on the extensive prior characterization of these samples. Accordingly, we have begun to explore variation in toxicity susceptibility as a function of genotype, as well as the relationship between toxic response and basal expression profiles.

Genotype-phenotype relationships are likely to reflect causal action of underlying physiological variation, and are thus of great interest to epidemiologists for understanding the ultimate sources of population variation. However, the effect sizes are typically small, as has been the source of considerable discussion in the genomics community (Manolio et al., 2009). Variation in basal mRNA expression, in contrast, may reflect cascades of responses controlled by the underlying genotype, and typically involves a smaller multiple testing penalty. Thus, we likely have more power to detect association of expression with toxicity response phenotypes, even though the underlying causality relationships may remain elusive. The highly significant associations identified through the analysis of population-level correlations between basal gene expression variability and chemical-induced toxicity have revealed several reasonable mode of

action hypotheses. For example, the in vitro toxicity of 1,3-indandione-containing rodenticides has been shown to occur through the inhibition of the pyrimidine synthetic pathway (Hall et al., 1994), and thioredoxin reductase (e.g., TXNRD3IT1) is required for dNTP pool maintenance during S phase (Koc et al., 2006). Expression of somatostatin receptor (SSTR)4 correlates with progesterone receptor levels in human breast tumors (Kumar et al., 2005). Thioredoxin reductase affects expression of progesterone receptor-controlled genes in MCF-7 cells (Rao et al., 2009).

Similarly, the quantitative assessment of inter-individual genetic variability in responses to environmental agents in vitro demonstrates the potential of this approach to explore the genetic basis for susceptibility through genome-wide association analysis. The genes SMC5 and MAMDC2 implicated in this study as associated with progesterone-induced toxicity are highly plausible and belong to pathways critical for development. The same locus was reported as associated with developmental abnormalities cleft palate and Kabuki syndrome (Kuniba et al., 2009; Marazita et al., 2004), and exposure to progesterone during gestation is known to cause cleft palate in rabbits (Andrew and Staples, 1977). Likewise, the association between Guggulsterone Z and polymorphisms in HIVEP1 is highly credible, given the known effects of Guggulsterone Z on apoptosis through NF- κ B-related signaling (Shishodia and Aggarwal, 2004). HIVEP1 belongs to a family of large zinc finger-containing transcription factors that bind specifically to the NF- κ B motif and related sequences (Yu et al., 2009). The alternative splice variant of HIVEP1, the GAAP-1 protein, can regulate p53 and IRF-1 dependent cell proliferation and apoptosis (Lallemand et al., 2002).

Important limitations to in vitro toxicity profiling using lymphoblasts, as compared to primary cells that may be obtained from other tissues of interest, include inability to assess target organ adverse effects, or a potential role of other environmental factors such as lifestyle, diet, or

co-exposures. In addition, the challenge of assessing the potential toxicity of chemical's metabolites, or the potential lack of the receptor-mediated signaling that may be critical for the downstream adverse molecular events, in lymphoblast cell lines also should be taken into consideration when interpreting the data. Still, whereas lymphocytes do not have the metabolic capacity of the liver, or even that of freshly isolated hepatocytes, they do express a number of nuclear receptors, as well as most genes of the phase I and II metabolism, and transporters (Siest et al., 2008). A comparison of the population-wide (250+ individuals of various races, ages and gender) variability in mRNA levels for several dozen liver-specific thyroid hormone-related genes between human liver (Schadt et al., 2008) and lymphoblast cell lines (Stranger et al., 2007) shows that most of the nuclear receptors and metabolism genes are expressed in lymphoblasts, albeit at 10- to 100-times lower quantity. Importantly, the between-subject variability in expression of these genes in either human liver or lymphoblasts is also of appreciable magnitude (4- to 10-fold). To overcome these limitations, both higher concentrations and known metabolites can be tested in vitro because of high throughput. Correcting for the cell growth rate and baseline metabolic rate also reduces effect correlation that may make responses across chemicals appear more similar (Choy et al., 2008).

Based on these results, we reason that a full and sensitive analysis of genomic predictors of toxicity response will be feasible through the joint use of toxicity phenotypes, genotype and expression information, though considerably larger sample sizes—likely on the order of several hundred or 1000s of individual cell lines—will be necessary. Such a population-based in vitro survey would greatly advance our understanding of the genetic underpinnings of susceptibility-related regulatory networks, and is ongoing in our laboratories.

FIGURES

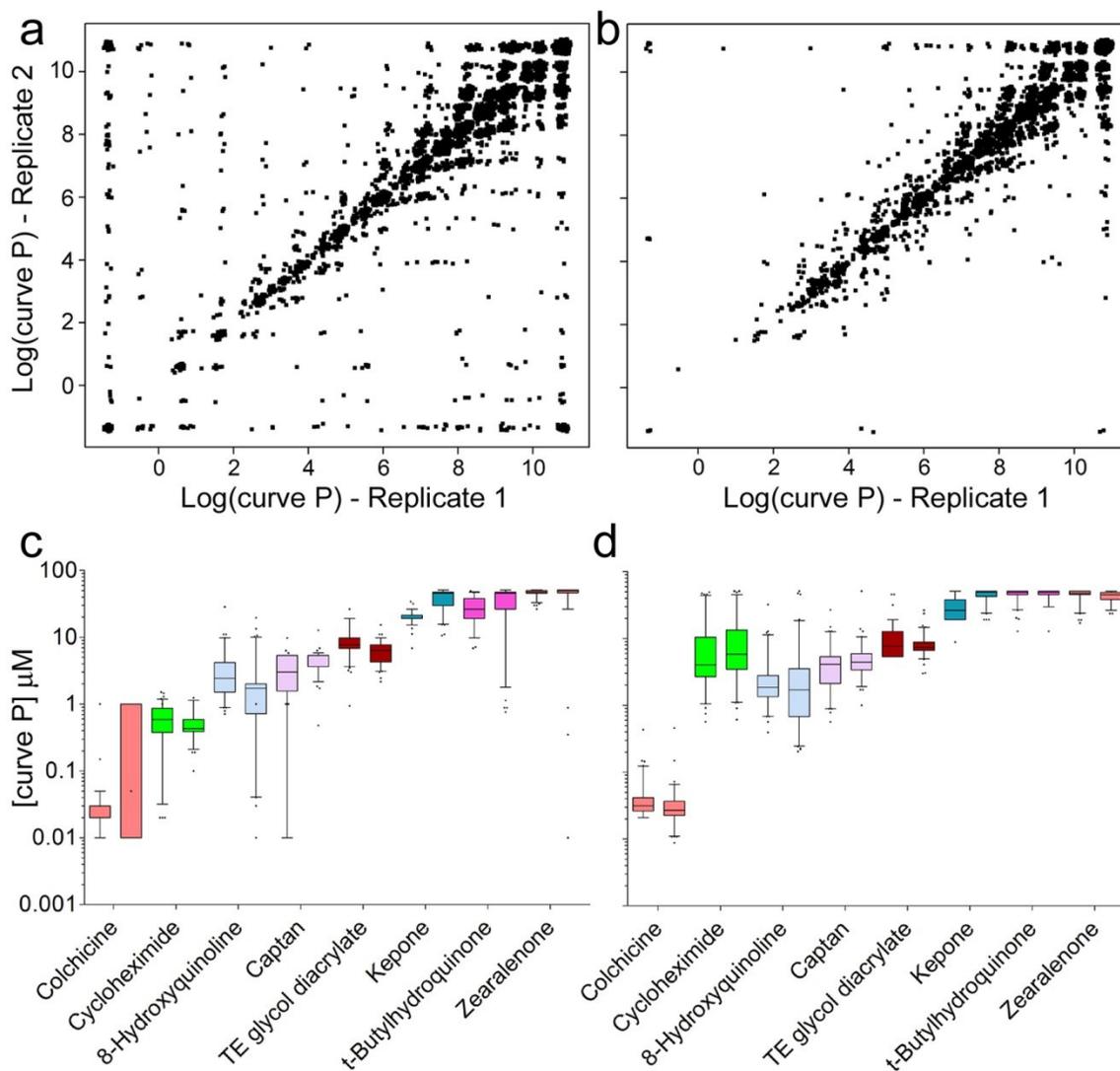


Figure 2.1. Experimental Reproducibility. Intra-experimental reproducibility for cytotoxicity (panels a and c) and caspase-3/7 (panels b and d) assays. Panels a and b show $\log(\text{curve P})$ values for randomly selected pairs of replicate plates within each chemical and cell line (240 chemicals x 81 cell lines=19,440 replicate pairs displayed). Panels c and d show side-by-side boxplots for eight duplicate compounds that were tested in 2 independent wells on each plate.

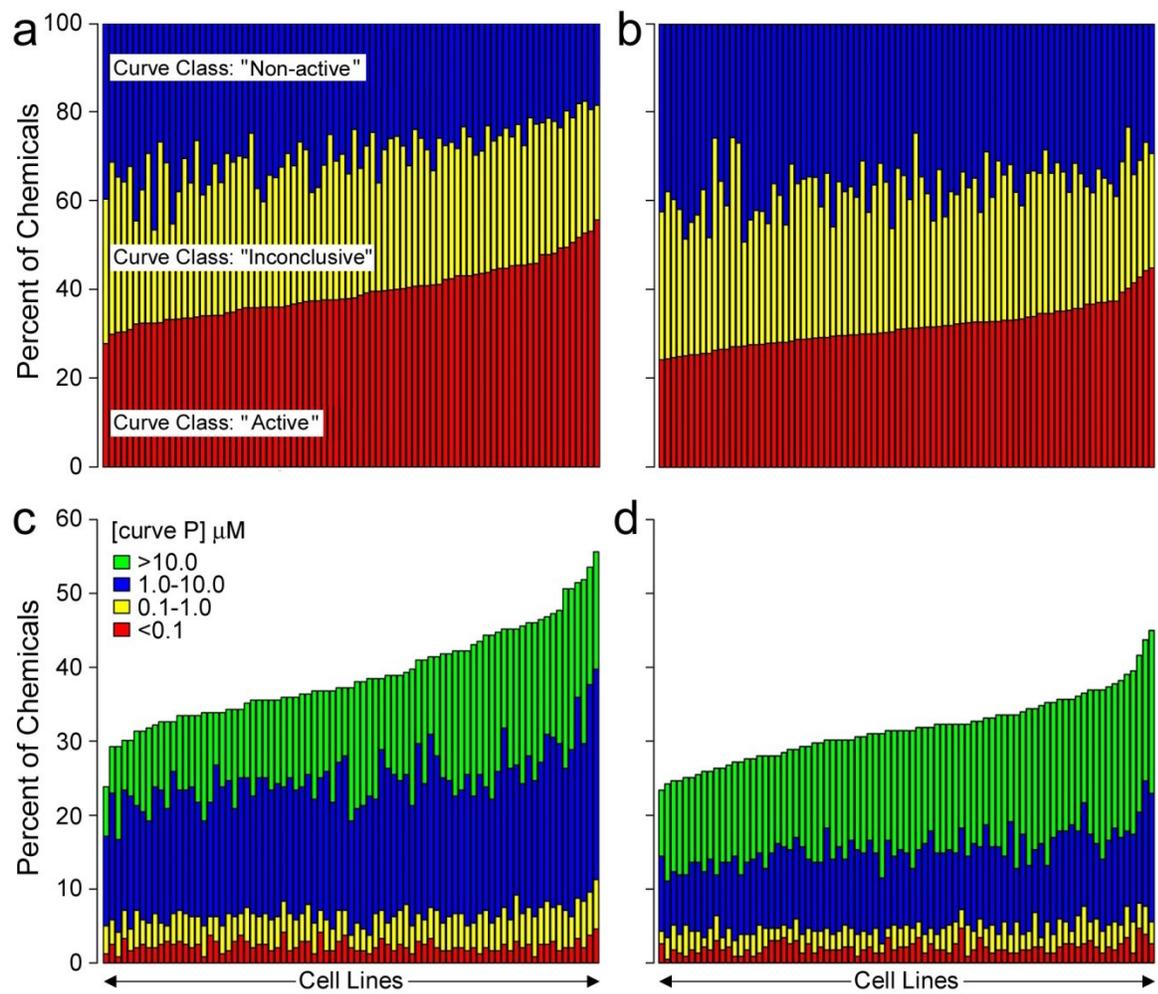


Figure 2.2. Cytotoxicity Distributions across Chemicals. Distribution of cytotoxicity across chemicals for cytotoxicity (panels a and c) and caspase-3/7 (panels b and d) assays. Panels a and b give the percentage of chemicals classified as ‘active’, ‘non-active’, or ‘inconclusive’ for each cell line. Panels c and d give the range of potency (curve P) for active chemicals in each cell line.

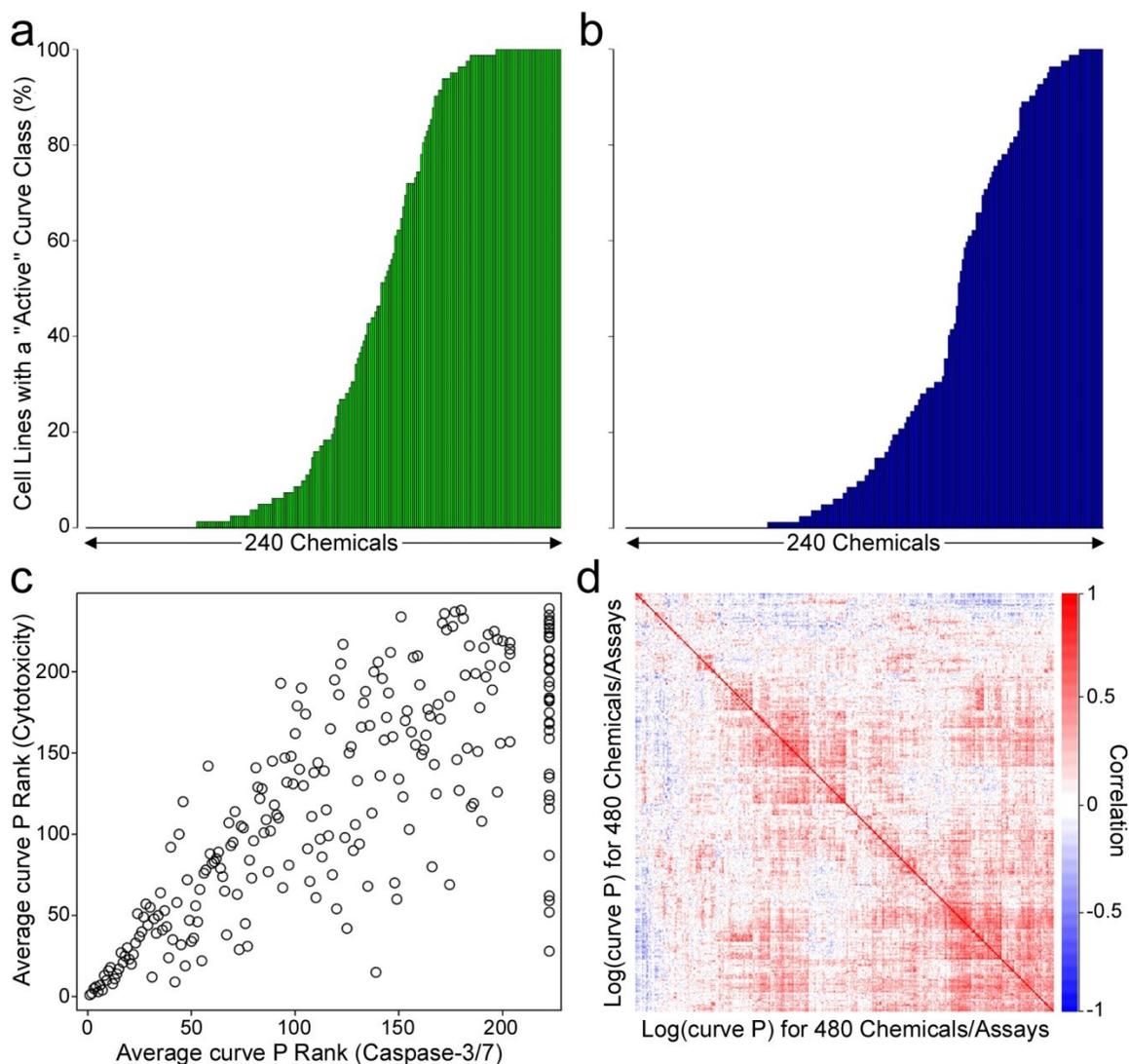


Figure 2.3. Cytotoxicity Distributions across Cell Lines. The percent of cell lines exhibiting activity for each chemical for cytotoxicity (panel a) and caspase-3/7 (panel b) assays. Panel c displays the rank of the mean ATP curve P value versus the mean caspase curve P value for each chemical. Panel d shows a heatmap of the correlations between log(curve P) values for all chemical-assay combinations.

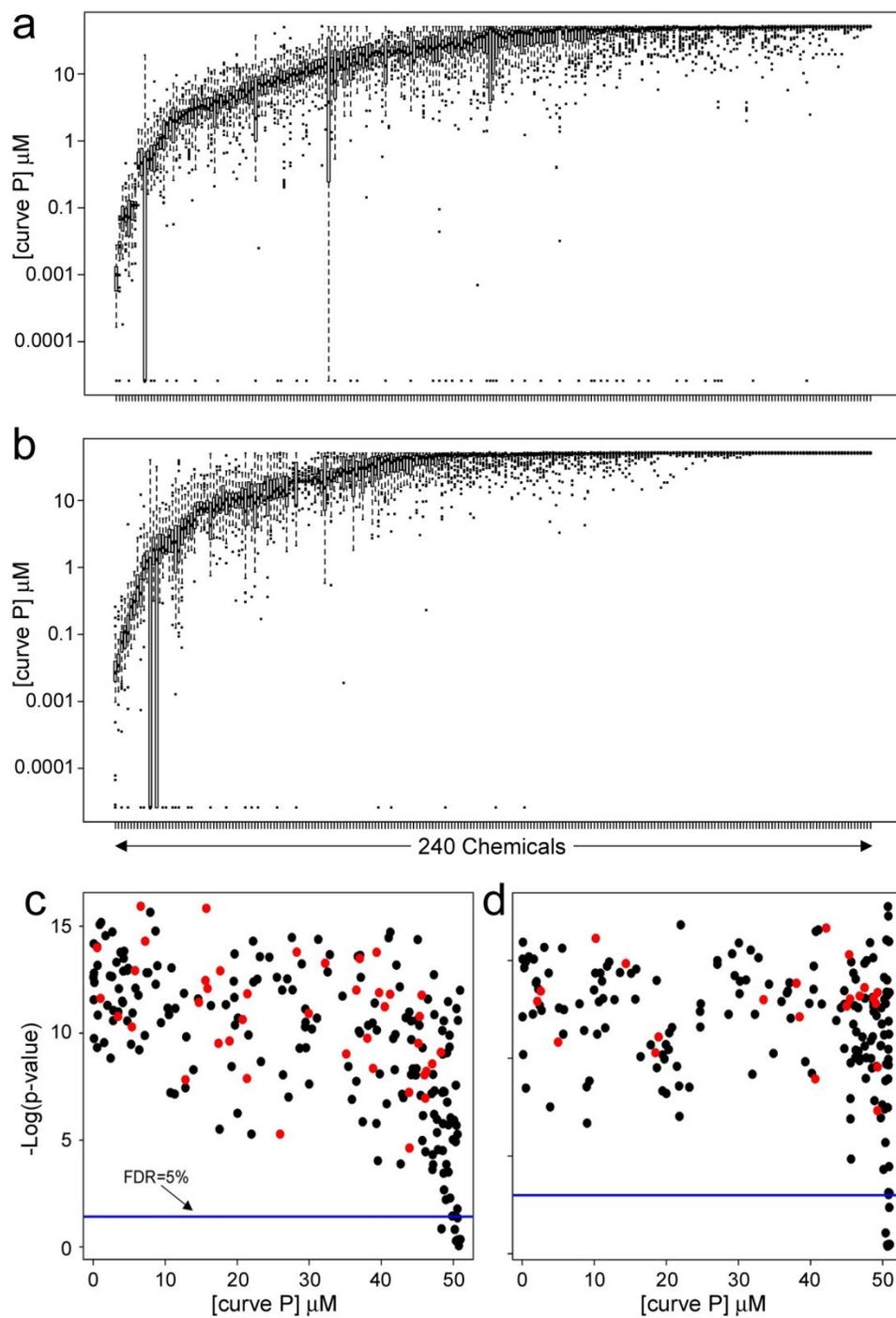


Figure 2.4. Inter-individual Variability Range for the 240 Chemicals. Boxplots of curve P values for each of the 240 chemicals (arranged by mean activity) across the 81 cell lines are shown for cytotoxicity (panel a) and caspase-3/7 (panel b) assays. For cytotoxicity (panel c) and caspase-3/7 (panel d) assays, $-\log(\text{p-values, Kruskal-Wallis test})$ were plotted against mean

curveP (μM). The blue line gives a False Discovery Rate-adjusted significance threshold (FDR=0.05). Chemicals colored in red had a significant correlation between activity and basal metabolic rate (ATP level in vehicle-treated cells) across the panel of cell lines (Spearman rank correlation; FDR<0.05).

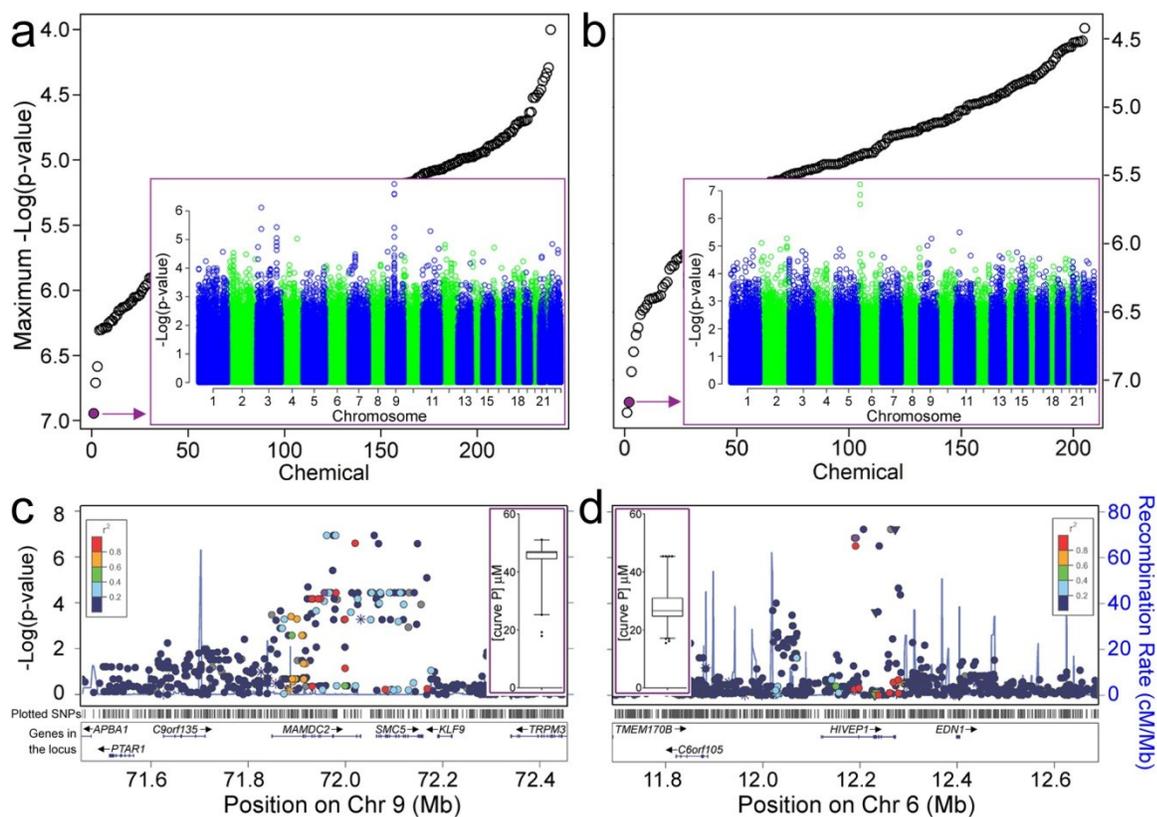


Figure 2.5. Toxicity-Genotype Associations. Toxicity-genotype relationships were assessed using GWAS analysis for the 240 chemicals on both cytotoxicity (panels a and c) and caspase-3/7 (panels b and d) assays. Panels a and b give p-values for the most significant SNP associated with toxicity for each chemical. The inset in the diagram gives $-\log(p\text{-values})$ for SNP-toxicity associations across the entire genome, for progesterone (cytotoxicity assay, inset in panel a) and Guggulsterones Z (caspase-3/7 assay, inset in panel b). Panels c and d provide a zoomed-in look at the locus with the most significant p-value for each of the two compounds, respectively. Correlation between SNPs is identified with colors. SNP and gene tracks are also shown. Inset: box & whiskers plots for each compound's curve P.

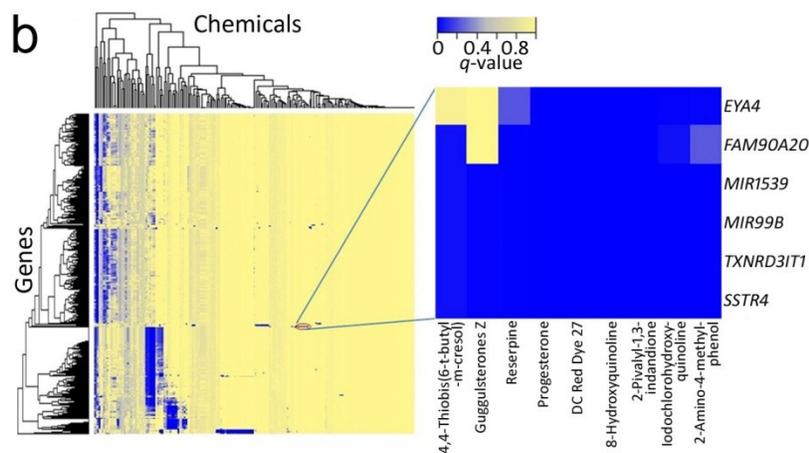
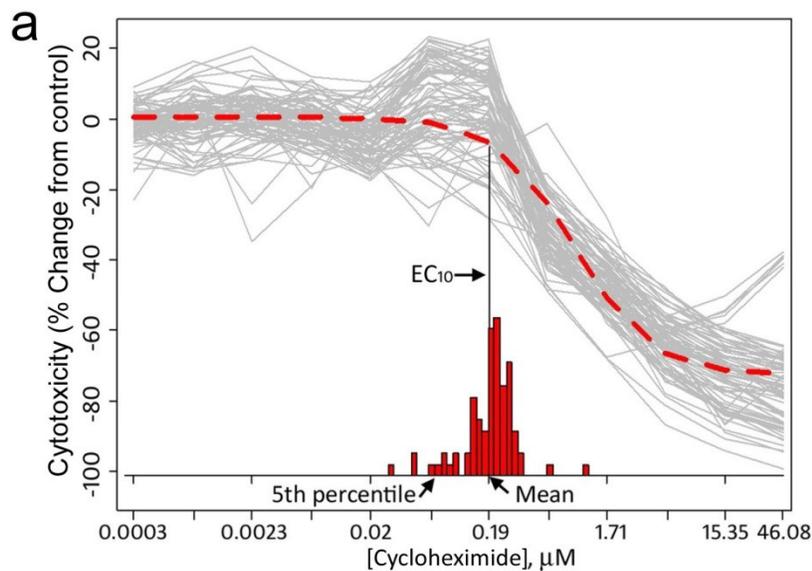


Figure 2.6. Toxicity-RNA expression Associations. Panel a, a population concentration-response was modeled using in vitro qHTS data using progesterone data (cytotoxicity assay) as an example. Logistic dose-response modeling was performed for each individual to the values shown in grey, providing individual 10% effect effective dose values (EC₁₀). The EC₁₀ obtained by performing the modeling on average assay values for each dose (see frequency distribution) are shown in the inset. Panel b, a heatmap of clustered false discovery rates (q-values, see color bar) for association of the data from caspase-3/7 assay with publicly-available

RNA-Seq expression data on a subset of cell lines. A sample subcluster containing progesterone is also shown.

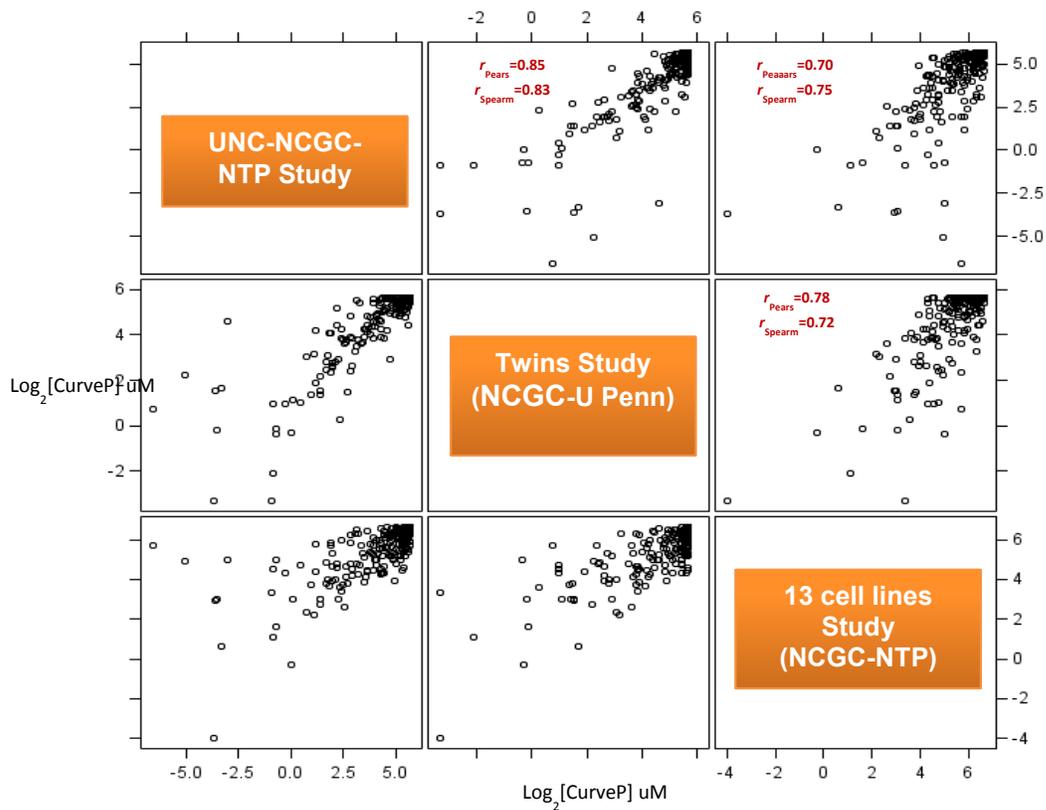


Figure 2.7. Comparison of Cytotoxicity across Different Studies. Cytotoxicity across different cells from different organs and species. A scatter plot of pairwise correlation between the cytotoxic response (curve P: the lowest concentration, which showed a consistent deviation from the baseline response and derived as detailed in Sedykh et al. (Sedykh et al. 2011) of common screened chemicals across our 81 cell lines in our experiment, paired twin lymphoblast cell lines in U Penn study, and the 13 cell lines in the NTP study.

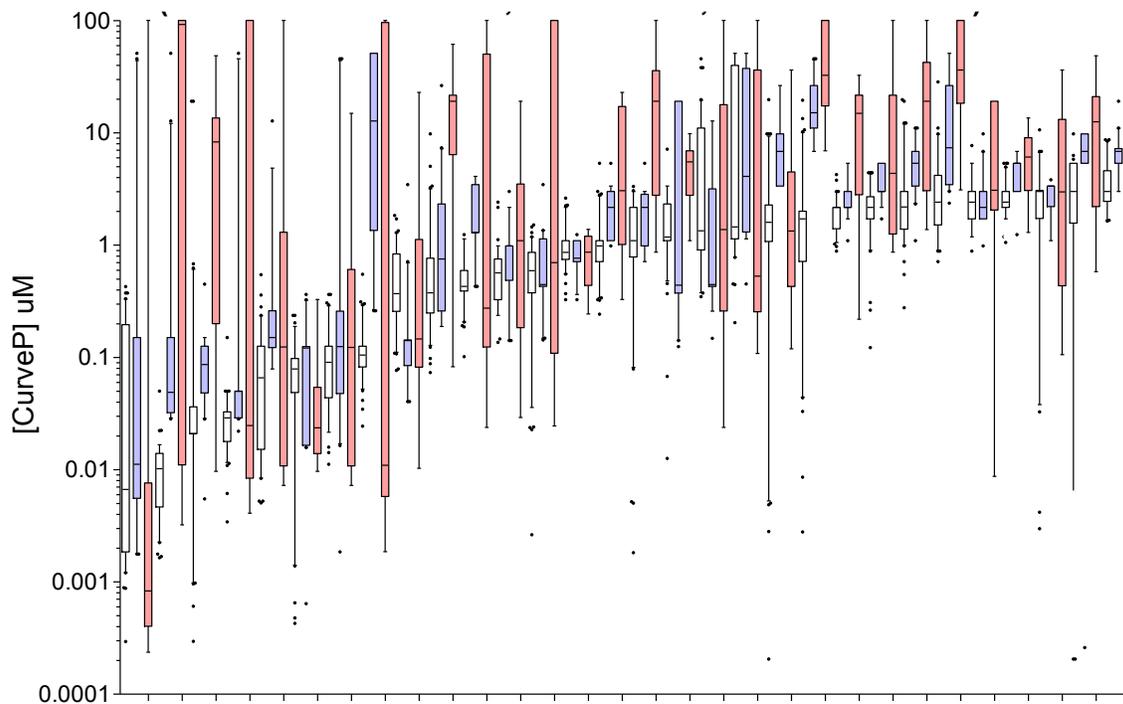


Figure 2.8. Comparison of Cytotoxicity Range across Different Studies. Boxplot of 30 most toxic chemicals in our experiments across cell lines in each study. White represents UNC-NCGC-NTP study, blue represents U-Penn-NCGC twins study, and red represents NCGC-NTP 13 cell lines study. We picked the 30 most toxic chemicals in our panel and plotted the range of toxicity expressed by CurveP across cell lines for each study in a box and whisker plot shown. This is the blue color coded box, and is cited as the first experiment (see comment above). The first experiment, labeled twins study (NCGC-U Penn), was a study that exposed lymphoblast cell lines from pairs of twins to the same set of chemicals (unpublished data). The other experiment was collaboration between NCGC and NTP, and exposed 13 cell lines coming from humans and rodents to the same set of chemicals (Xia et al. 2008a).

Table 2.1 List of 240 Screened Chemicals

Sequence ID	Screening ID (CID)	CASRN	Chemical Name
1	AB02509330-01	59870	Nitrofurazone
2	AB02509332-01	25152845	2,4-Decadienal
3	AB02509368-01	148243	8-Hydroxyquinoline
4	AB02509400-01	77439760	3-Chloro-4-(dichloromethyl)-5-hydroxy-2(5H)-furanone, (MX)
5	AB02509551-01	126987	Methacrylonitrile
6	AB02540517-01	65646686	4-Hydroxyphenyl retinamide
7	AB02540519-01	7789120	Sodium dichromate dihydrate (VI)
8	AB02540526-01	34256821	Acetochlor
9	AB02540527-01	7778509	Potassium dichromate
10	AB02540528-01	127479	all-trans-Retinyl acetate
11	AB02540529-01	75330755	Lovastatin
12	AB02540531-01	148823	Melphalan
13	AB02540538-01	15972608	Alachlor
14	AB02540539-01	15663271	cis-Dichlorodiamine platinum
15	AB02540542-01	72559069	Rifabutin
16	AB02540564-01	458377	Curcumin
17	AB02540565-01	54965241	Tamoxifen, citrate salt
18	AB02540575-01	55981094	Nitazoxanide
19	AB02540576-01	7220793	Methylene blue trihydrate
20	AB02540577-01	25316409	Adriamycin, hydrochloride
21	AB02540586-01	10026241	Cobalt sulfate heptahydrate
22	AB02540589-01	140647	Pentamidine isethionate
23	AB02540601-01	1.6E+08	Nelfinavir mesylate
24	AB02911345-01	142836	2,4-Hexadienal
25	AB02911369-01	1948330	tert-Butylhydroquinone
26	AB02911410-01	91532	Ethoxyquin
27	AB02911423-01	116063	Aldicarb
28	AB02914401-01	50760	Actinomycin D
29	AB07859937-01	20830755	Digoxin
30	AB07930213-01	57830	Progesterone
31	AB07930229-01	17924924	Zearalenone
32	AB07930241-01	123308	p-Aminophenol
33	AB07930243-01	97187	2,2'-Thiobis(4,6-dichlorophenol)
34	AB07930251-01	472866	13-cis-Retinal (Vitamin A aldehyde)
35	AB07930253-01	61702441	2-Chloro-p-phenylenediamine SO4
36	AB07930263-01	95067	Sulfallate
37	AB07930274-01	446866	Azathioprine
38	AB07930275-01	533744	Dazomet
39	AB07930277-01	879390	2,3,4,5-Tetrachloronitrobenzene

Sequence ID	Screening ID	CASRN	Chemical Name
40	AB07930285-01	2243767	Alizarin Yellow R, sodium salt
41	AB07930286-01	520365	Apigenin
42	AB07930288-01	961115	Tetrachlorvinphos (Gardona)
43	AB07930292-01	115093	Methyl mercuric (II) chloride
44	AB07930298-01	320672	5-Azacytidine
45	AB07931485-01	83261	2-Pivalyl-1,3-indandione
46	AB07934770-01	74317	N,N'-Diphenyl-p-phenylenediamine
47	AB07935027-01	396010	Triamterene
48	AB07935479-01	24169026	Econazole nitrate
49	AB07935653-01	154427	6-Thioguanine (6-TG)
50	AB07935671-01	130267	Iodochlorohydroxyquinoline
51	AB07935672-01	123319	Hydroquinone
52	AB07935817-01	612839	3,3'-Dichlorobenzidine 2HCl
53	AB07935832-01	95545	o-Phenylenediamine
54	AB07944034-01	104405	p -n -Nonylphenol
55	AB07944048-01	10108642	Cadmium II chloride
56	AB07944059-01	65558692	1,3-Diiminobenz (f)-isoindoline
57	AB07944071-01	77929	Citric Acid
58	AB07944077-01	818611	Hydroxyethyl acrylate
59	AB07944095-01	102363	3,4-Dichlorophenyl isocyanate
60	AB07944097-01	10563298	dimethyldipropylene-triamine
61	AB07944109-01	446355	2,4-Difluoronitrobenzene
62	AB07944128-01	66819	Cycloheximide
63	AB07944329-01	143500	Kepon
64	AB07944331-01	538716	Domiphen bromide
65	AB07944340-01	9002931	Triton X-100
66	AB07944343-01	64868	Colchicine
67	AB07944355-01	50022	Dexamethasone
68	AB07944366-01	14866332	tetra-N-Octylammonium bromide
69	AB07944367-01	140727	Cetylpyridinium bromide
70	AB07944368-01	57830	Progesterone
71	AB07944378-01	55561	Chlorhexidine
72	AB07944380-01	549188	Amitriptyline HCl
73	AB07944388-01	68047063	4-Hydroxytamoxifen
74	AB07944390-01	152114	Verapamil HCl
75	AB07944392-01	17831719	Tetraethylene glycol diacrylate
76	AB07944400-01	17924924	Zearalenone
77	AB07944404-01	133062	Captan 90-concentrate (solid)
78	AB07944715-01	62384	Phenyl mercuric acetate
79	AB07944716-01	13463417	Zinc pyrithione
80	AB07944719-01	66819	Cycloheximide

Sequence ID	Screening ID	CASRN	Chemical Name
81	AB07944721-01	106514	p-Quinone
82	AB07944724-01	100425	Styrene
83	AB07944726-01	463401	Linolenic acid
84	AB07944736-01	110576	trans-1,4-Dichloro-2-butene
85	AB07944738-01	1191419	Ethyl linolenate
86	AB07944741-01	67970	Vitamin D3
87	AB07944749-01	116314	trans-Retinal (Vitamin A aldehyde)
88	AB07944762-01	134327	1-Naphthylamine
89	AB07944765-01	5153673	beta-Nitrostyrene
90	AB07944785-01	111308	Glutaraldehyde (Glutaric dialdehyde)
91	AB07944787-01	548629	Hexamethyl-p-rosaniline chloride (Gentian violet)
92	AB07944788-01	379793	Ergotamine tartrate
93	AB07944790-01	305033	Chlorambucil
94	AB07944791-01	485472	Ninhydrin
95	AB07944800-01	3018120	Dichloroacetonitrile
96	AB07949036-01	105113	p-Quinone dioxime (p-Benzoquinone dioxime)
97	AB07949142-01	72140	Sulfathiazole
98	AB07949247-01	90415	2-Aminobiphenyl
99	AB07949372-01	100221	N,N,N',N'-Tetramethyl-p-phenylenediamine
100	AB07975480-01	517282	Hematoxylin (C.I. 75290)
101	AB07976081-01	2451629	Tris(2,3-epoxypropyl)isocyanurate
102	AB07980848-01	95556	o-Aminophenol
103	AB07980929-01	70304	Hexachlorophene
104	AB07980947-01	2016888	Amiloride HCl
105	AB07980958-01	66717	o-Phenanthroline
106	AB07980986-01	532274	Chloroacetophenone
107	AB07981003-01	2425798	1,4-Butanediol diglycidyl ether
108	AB07981015-01	4074888	Diethylene glycol diacrylate
109	AB07981053-01	81550	1,8-Dihydroxy-4,5-dinitroanthraquinone
110	AB07981054-01	62737	Dichlorvos (Vapona)
111	AB07981076-01	58140	Pyrimethamine
112	AB07981077-01	610399	3,4-Dinitrotoluene
113	AB07981167-01	2465272	Auramine
114	AB07981200-01	74975	Bromochloromethane
115	AB07981244-01	99989	N,N-Dimethyl-p-phenylenediamine
116	AB07981253-01	764421	Fumaronitrile
117	AB07981260-01	71589	Medroxyprogesterone acetate
118	AB07981274-01	33229344	HC blue 2
119	AB07981291-01	50555	Reserpine
120	AB07981301-01	118752	Chloranil
121	AB07981317-01	13048334	1,6-Hexanediol diacrylate

Sequence ID	Screening ID	CASRN	Chemical Name
122	AB07981330-01	4252782	2,2',4'-Trichloroacetophenone
123	AB07981368-01	120809	Catechol (1,2-Benzenediol)
124	AB07981464-01	496720	3,4-Diaminotoluene
125	AB07981614-01	1465254	N-(1-Naphthyl)ethylenediamine 2HCl
126	AB07981816-01	569619	Basic red 9 (p-Rosaniline HCl) (C.I. 42500)
127	AB07981842-01	87683	Hexachloro-1,3-butadiene
128	AB07981866-01	113928	Chlorpheniramine maleate
129	AB07985231-01	6459945	Acid red 114 (C.I. 23635)
130	AB07985248-01	97234	2,2'-Methylenebis-(4-chlorophenol)
131	AB07985261-01	393759	4-Chloro-3,5-dinitro-a,a,a-trifluorotoluene
132	AB07985289-01	1675543	Bisphenol A diglycidyl ether
133	AB07985291-01	156105	p-Nitrosodiphenylamine
134	AB07985305-01	992596	Direct red 2 (C.I. 23500)
135	AB07985346-01	93050	N,N-Diethyl-p-phenylenediamine
136	AB07985386-01	133062	Captan
137	AB07990164-01	51218	5-Fluorouracil
138	AB07990179-01	60571	Dieldrin
139	AB07990183-01	2213630	2,3-Dichloroquinoxaline
140	AB07990199-01	834286	Phenformin HCl
141	AB07990229-01	84695	Diisobutyl phthalate
142	AB07990272-01	117102	Danthron
143	AB07990332-01	64868	Colchicine
144	AB07990339-01	137304	Ziram
145	AB07990348-01	55550	N-Methyl-p-aminophenol sulfate
146	AB07990352-01	140498	4'-(Chloroacetyl)acetanilide
147	AB07990366-01	1239458	Ethidium bromide
148	AB08000887-01	3252435	Dibromoacetonitrile
149	AB08001062-01	434071	Oxymetholone
150	AB08001092-01	3524683	Pentaerythritol Triacrylate
151	AB08001212-01	12789036	Chlordane, technical grade
152	AB08001301-01	20562021	a-Solanine
153	AB08001481-01	70257	N-Methyl-N'-nitro-N-nitrosoguanidine
154	AB08001574-01	630160	1,1,1,2-Tetrabromoethane
155	AB08002009-01	13311847	Flutamide
156	AB08002746-01	6983795	Bixin
157	AB08002778-01	148243	8-Hydroxyquinoline
158	AB08002816-01	2437298	Malachite Green Oxalate
159	AB08002895-01	1.5E+08	Saquinavir Mesylate
160	AB08002899-01	7789120	Sodium Dichromate Dihydrate
161	AB08003367-01	73314	Melatonin
162	AB08003603-01	930687	2-Cyclohexen-1-one

Sequence ID	Screening ID	CASRN	Chemical Name
163	AB08003789-01	98293	p-tert Butylcatechol
164	AB08003879-01	598914	Dibromonitromethane
165	AB08003920-01	13473262	D&C Red Dye 27
166	AB08006114-01	538750	1,3-Dicyclohexylcarbodiimide
167	AB08006121-01	115297	Endosulfan
168	AB08006152-01	96695	4,4-Thiobis(6-t-butyl-m-cresol)
169	AB08006174-01	1948330	t-Butylhydroquinone
170	AB08006184-01	67209	Nitrofurantoin
171	AB08006209-01	87661	Pyrogallol
172	AB08006212-01	481403	1,4,5-Trihydroxynaphthalene
173	AB08006314-01	76448	Heptachlor
174	AB08006839-01	57976	7,12-Dimethylbenz(a)anthracene
175	AB08006849-01	2373980	3,3'-Dihydroxybenzidine
176	AB08007015-01	6317186	Methylene bis(thiocyanate)
177	AB08007268-01	121540	Benzethonium chloride
178	AB08007382-01	1402260	Cadmium acetate, dihydrate
179	AB08007516-01	518821	Emodin (bulk)
180	AB08007549-01	23541506	Daunomycin HCL
181	AB08007607-01	50282	17-beta-Estradiol
182	AB08007613-01	77474	Hexachlorocyclopentadiene
183	AB08007614-01	6338416	5-(Hydroxymethyl)-2-furoic acid
184	AB08007627-01	793248	n-(1,3-Dimethylbutyl)-n'-phenyl-p-phenylenediamine
185	AB08080834-01	137268	Tetramethylthiuram disulfide
186	AB08080836-01	95852	2-Amino-4-chlorophenol
187	AB08080842-01	6088513	6-Hydroxy-2-naphthyl disulfide (DDD)
188	AB08080844-01	7487947	Mercuric chloride
189	AB08080867-01	95841	2-Amino-4-methylphenol
190	AB08080883-01	117793	2-Aminoanthraquinone
191	AB08080892-01	1074120	Phenyl glyoxal
192	AB08080909-01	1271198	Titanocene dichloride
193	AB08080916-01	6112761	6-Mercaptopurine monohydrate
194	AB08082658-01	100276	p-Nitrophenethyl alcohol
195	AB08082669-01	95830	4-Chloro-o-phenylenediamine
196	AB08082688-01	762754	t-Butyl formate
197	AB08082706-01	138896	N,N-Dimethyl-p-nitrosoaniline
198	AB08082716-01	21829254	Nifedipine
199	AB08082717-01	1260179	Carminic acid
200	AB08082719-01	619238	m-Nitrobenzyl chloride
201	AB08082743-01	612237	o-Nitrobenzyl chloride
202	AB08097184-01	58548	Ethacrynic acid
203	AB08548260-01	12083486	Bis(cyclopentadienyl)vanadium chloride

Sequence ID	Screening ID	CASRN	Chemical Name
204	AB08548266-01	80159	Cumene hydroperoxide
205	AB08548274-01	1338234	Methyl ethyl ketone peroxide
206	AB08548279-01	1271289	Nickelocene
207	AB08548292-01	933788	2,3,5-Trichlorophenol
208	AB08548310-01	68268	trans-Retinol acid (Vitamin A)
209	AB08548314-01	17831719	Tetraethylene glycol diacrylate
210	AB08548322-01	106898	1,2-Epoxy-3-chloropropane (Epichlorohydrin)
211	AB08548327-01	143500	Chlordecone (Kepone)
212	AB08548342-01	101906	Diglycidyl resorcinol ether
213	AB08548344-01	70348	1-Fluoro-2,4-dinitrobenzene
214	AB08548348-01	101962	N,N'-Di-sec-butyl-p-phenylenediamine
215	AB08582918-01	11024241	Digitonin
216	AB08582919-01	6219892	4-Amino-4'-hydroxy-3-methyldiphenylamine
217	AB08582920-01	101724	N-Isopropyl-N'-phenyl-p-phenylenediamine
218	AB08582927-01	4901513	2,3,4,5-Tetrachlorophenol
219	AB08582929-01	55867	Nitrogen mustard HCl
220	AB08582931-01	97325	4-Methoxy-3-nitro-N-phenylbenzamide
221	AB08582933-01	1421632	2'4'5'-Trihydroxybutyrophenone
222	AB08582939-01	123773	Azodicarbonamide
223	AB08582940-01	52417228	9-Aminoacridine HCl H2O
224	AB08582944-01	7779308	1-(2,6,6-Trimethyl-2-cyclohexen-1-yl)-1-penten-3-one
225	AB08582948-01	63923	Phenoxybenzamine HCl
226	AB08582953-01	55566308	Tetrakis(hydroxymethyl)phosphonium SO4
227	AB08582961-01	5493458	1,2-Cyclohexanedicarboxylic acid, bis(oxiranymethyl) ester
228	AB08582965-01	609198	3,4,5-Trichlorophenol
229	AB08582970-01	103333	Azobenzene
230	AB08582976-01	26530201	2-Octyl-3-isothiazolone
231	AB08582980-01	97245	2,2'-Thiobis(4-chlorophenol)
232	AB08582982-01	8001352	Toxaphene
233	AB08582989-01	309002	Aldrin
234	AB08582998-01	478433	Rhein (1,8-Dihydroxy-3-carboxylantraquinone)
235	AB08583000-01	633658	Berberine chloride
236	AB13681039-01	39025235	Guggulsterones Z
237	AB13681051-01	39025246	Guggulsterones E
238	AB13681075-01	3155575	Dihydromethysticin
239	AB13681076-01	18642449	Actein
240	CONTROL	62996741	Staurosporine

F. REFERENCES

- Andrew, F. D. and Staples, R. E. (1977). Prenatal toxicity of medroxyprogesterone acetate in rabbits, rats, and mice. *Teratology* 15, 25-32.
- Aulchenko, Y. S., Ripke, S., Isaacs, A., and van Duijn, C. M. (2007). GenABEL: an R library for genome-wide association analysis. *Bioinformatics* 23, 1294-1296.
- Choy, E., Yelensky, R., Bonakdar, S., Plenge, R. M., Saxena, R., De Jager, P. L., Shaw, S. Y., Wolfish, C. S., Slavik, J. M., Cotsapas, C., Rivas, M., Dermitzakis, E. T., Cahir-McFarland, E., Kieff, E., Hafler, D., Daly, M. J., and Altshuler, D. (2008). Genetic analysis of human traits in vitro: drug response and gene expression in lymphoblastoid cell lines. *PLoS. Genet* 4, e1000287.
- Collins, F. S., Gray, G. M., and Bucher, J. R. (2008). Toxicology. Transforming environmental health protection. *Science* 319, 906-907.
- Crump, K. S., Chen, C., and Louis, T. A. (2010). The future use of in vitro data in risk assessment to set human exposure standards: challenging problems and familiar solutions. *Environ Health Perspect.* 118, 1350-1354.
- Durbin, R. M., Abecasis, G. R., Altshuler, D. L., Auton, A., Brooks, L. D., Durbin, R. M., Gibbs, R. A., Hurles, M. E., and McVean, G. A. (2010). A map of human genome variation from population-scale sequencing. *Nature* 467, 1061-1073.
- Hall, I. H., Wong, O. T., Chi, L. K., and Chen, S. Y. (1994). Cytotoxicity and mode of action of substituted indan-1, 3-diones in murine and human tissue cultured cells. *Anticancer Res* 14, 2053-2058.
- Harrill, A. H., Watkins, P. B., Su, S., Ross, P. K., Harbourt, D. E., Stylianou, I. M., Boorman, G. A., Russo, M. W., Sackler, R. S., Harris, S. C., Smith, P. C., Tennant, R., Bogue, M., Paigen, K., Harris, C., Contractor, T., Wiltshire, T., Rusyn, I., and Threadgill, D. W. (2009). Mouse population-guided resequencing reveals that variants in CD44 contribute to acetaminophen-induced liver injury in humans. *Genome Res* 19, 1507-1515.
- Hartung, T. and Rovida, C. (2009). Chemical regulators have overreached. *Nature* 460, 1080-1081.
- Huang, R., Southall, N., Cho, M. H., Xia, M., Inglese, J., and Austin, C. P. (2008). Characterization of diversity in toxicity mechanism using in vitro cytotoxicity assays in quantitative high throughput screening. *Chem Res Toxicol* 21, 659-667.
- Inglese, J., Auld, D. S., Jadhav, A., Johnson, R. L., Simeonov, A., Yasgar, A., Zheng, W., and Austin, C. P. (2006). Quantitative high-throughput screening: a titration-based approach that efficiently identifies biological activities in large chemical libraries. *Proc. Natl. Acad Sci U. S. A* 103, 11473-11478.
- International HapMap Consortium (2005). A haplotype map of the human genome. *Nature* 437, 1299-1320.
- Johnson, W. E., Li, C., and Rabinovic, A. (2007). Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics.* 8, 118-127.

- Judson, R. S., Kavlock, R. J., Setzer, R. W., Cohen Hubal, E. A., Martin, M. T., Knudsen, T. B., Houck, K. A., Thomas, R. S., Wetmore, B. A., and Dix, D. J. (2011). Estimating Toxicity-Related Biological Pathway Altering Doses for High-Throughput Chemical Risk Assessment. *Chem Res Toxicol* 24, 451-462.
- Kavlock, R. J., Austin, C. P., and Tice, R. R. (2009). Toxicity testing in the 21st century: implications for human health risk assessment. *Risk Anal* 29, 485-487.
- Koc, A., Mathews, C. K., Wheeler, L. J., Gross, M. K., and Merrill, G. F. (2006). Thioredoxin is required for deoxyribonucleotide pool maintenance during S phase. *J Biol Chem* 281, 15058-15063.
- Kruskal, W. H. and Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. *J. Am. Stat. Assoc.* 47, 583-621.
- Kumar, U., Grigorakis, S. I., Watt, H. L., Sasi, R., Snell, L., Watson, P., and Chaudhari, S. (2005). Somatostatin receptors in primary human breast cancer: quantitative analysis of mRNA for subtypes 1--5 and correlation with receptor protein expression and tumor pathology. *Breast Cancer Res Treat.* 92, 175-186.
- Kuniba, H., Yoshiura, K., Kondoh, T., Ohashi, H., Kurosawa, K., Tonoki, H., Nagai, T., Okamoto, N., Kato, M., Fukushima, Y., Kaname, T., Naritomi, K., Matsumoto, T., Moriuchi, H., Kishino, T., Kinoshita, A., Miyake, N., Matsumoto, N., and Niikawa, N. (2009). Molecular karyotyping in 17 patients and mutation screening in 41 patients with Kabuki syndrome. *J Hum. Genet* 54, 304-309.
- Lallemant, C., Plamieri, M., Blanchard, B., Meritet, J. F., and Tovey, M. G. (2002). GAAP-1: a transcriptional activator of p53 and IRF-1 possesses pro-apoptotic activity. *EMBO Rep.* 3, 153-158.
- Manolio, T. A., Collins, F. S., Cox, N. J., Goldstein, D. B., Hindorff, L. A., Hunter, D. J., McCarthy, M. I., Ramos, E. M., Cardon, L. R., Chakravarti, A., Cho, J. H., Guttmacher, A. E., Kong, A., Kruglyak, L., Mardis, E., Rotimi, C. N., Slatkin, M., Valle, D., Whittemore, A. S., Boehnke, M., Clark, A. G., Eichler, E. E., Gibson, G., Haines, J. L., Mackay, T. F., McCarroll, S. A., and Visscher, P. M. (2009). Finding the missing heritability of complex diseases. *Nature* 461, 747-753.
- Marazita, M. L., Murray, J. C., Lidral, A. C., Arcos-Burgos, M., Cooper, M. E., Goldstein, T., Maher, B. S., Daack-Hirsch, S., Schultz, R., Mansilla, M. A., Field, L. L., Liu, Y. E., Prescott, N., Malcolm, S., Winter, R., Ray, A., Moreno, L., Valencia, C., Neiswanger, K., Wyszynski, D. F., Bailey-Wilson, J. E., Albacha-Hejazi, H., Beaty, T. H., McIntosh, I., Hetmanski, J. B., Tuncbilek, G., Edwards, M., Harkin, L., Scott, R., and Roddick, L. G. (2004). Meta-analysis of 13 genome scans reveals multiple cleft lip/palate genes with novel loci on 9q21 and 2q32-35. *Am J Hum. Genet* 75, 161-173.
- Martin, M. T., Dix, D. J., Judson, R. S., Kavlock, R. J., Reif, D. M., Richard, A. M., Rotroff, D. M., Romanov, S., Medvedev, A., Poltoratskaya, N., Gambarian, M., Moeser, M., Makarov, S. S., and Houck, K. A. (2010). Impact of environmental chemicals on key transcription regulators and correlation to toxicity end points within EPA's ToxCast program. *Chem Res Toxicol* 23, 578-590.

- Montgomery, S. B., Lappalainen, T., Gutierrez-Arcelus, M., and Dermitzakis, E. T. (2011). Rare and common regulatory variation in population-scale sequenced human genomes. *PLoS Genet* 7, e1002144.
- Montgomery, S. B., Sammeth, M., Gutierrez-Arcelus, M., Lach, R. P., Ingle, C., Nisbett, J., Guigo, R., and Dermitzakis, E. T. (2010). Transcriptome genetics using second generation sequencing in a Caucasian population. *Nature* 464, 773-777.
- National Research Council (2007). *Toxicity Testing in the 21st Century: A Vision and a Strategy*. National Academies Press, Washington, DC.
- National Research Council (2008). *Science and Decisions: Advancing Risk Assessment*. The National Academies Press, Washington, D.C.
- O'Shea, S. H., Schwarz, J., Kosyk, O., Ross, P. K., Ha, M. J., Wright, F. A., and Rusyn, I. (2011). In vitro screening for population variability in chemical toxicity. *Toxicol Sci* 119, 398-407.
- Parham, F., Austin, C., Southall, N., Huang, R., Tice, R., and Portier, C. (2009). Dose-response modeling of high-throughput screening data. *J Biomol. Screen.* 14, 1216-1227.
- Plunkett, L. M., Kaplan, A. M., and Becker, R. A. (2010). An enhanced tiered toxicity testing framework with triggers for assessing hazards and risks of commodity chemicals. *Regul. Toxicol Pharmacol* 58, 382-394.
- Pruim, R. J., Welch, R. P., Sanna, S., Teslovich, T. M., Chines, P. S., Gliedt, T. P., Boehnke, M., Abecasis, G. R., and Willer, C. J. (2010). LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* 26, 2336-2337.
- Rao, A. K., Ziegler, Y. S., McLeod, I. X., Yates, J. R., and Nardulli, A. M. (2009). Thioredoxin and thioredoxin reductase influence estrogen receptor alpha-mediated gene expression in human breast cancer cells. *J Mol Endocrinol.* 43, 251-261.
- Reif, D. M., Martin, M. T., Tan, S. W., Houck, K. A., Judson, R. S., Richard, A. M., Knudsen, T. B., Dix, D. J., and Kavlock, R. J. (2010). Endocrine profiling and prioritization of environmental chemicals using ToxCast data. *Environ Health Perspect.* 118, 1714-1720.
- Rusyn, I., Gatti, D. M., Wiltshire, T., Kleeberger, S. R., and Threadgill, D. W. (2010). Toxicogenetics: population-based testing of drug and chemical safety in mouse models. *Pharmacogenomics* 11, 1127-1136.
- Schadt, E. E., Molony, C., Chudin, E., Hao, K., Yang, X., Lum, P. Y., Kasarskis, A., Zhang, B., Wang, S., Suver, C., Zhu, J., Millstein, J., Sieberts, S., Lamb, J., Guhathakurta, D., Derry, J., Storey, J. D., Avila-Campillo, I., Kruger, M. J., Johnson, J. M., Rohl, C. A., van Nas, A., Mehrabian, M., Drake, T. A., Luskis, A. J., Smith, R. C., Guengerich, F. P., Strom, S. C., Schuetz, E., Rushmore, T. H., and Ulrich, R. (2008). Mapping the genetic architecture of gene expression in human liver. *PLoS. Biol* 6, e107.
- Schaid, D. J., Rowland, C. M., Tines, D. E., Jacobson, R. M., and Poland, G. A. (2002). Score tests for association between traits and haplotypes when linkage phase is ambiguous. *Am J Hum. Genet* 70, 425-434.

- Sedykh, A., Zhu, H., Tang, H., Zhang, L., Richard, A., Rusyn, I., and Tropsha, A. (2011). Use of in Vitro HTS-Derived Concentration-Response Data as Biological Descriptors Improves the Accuracy of QSAR Models of in Vivo Toxicity. *Environ Health Perspect.* 119, 364-370.
- Shi, J., Springer, S., and Escobar, P. (2010). Coupling cytotoxicity biomarkers with DNA damage assessment in TK6 human lymphoblast cells. *Mutat. Res* 696, 167-178.
- Shishodia, S. and Aggarwal, B. B. (2004). Guggulsterone inhibits NF-kappaB and IkappaBalpha kinase activation, suppresses expression of anti-apoptotic gene products, and enhances apoptosis. *J Biol Chem* 279, 47148-47158.
- Siest, G., Jeannesson, E., Marteau, J. B., Samara, A., Marie, B., Pfister, M., and Visvikis-Siest, S. (2008). Transcription factor and drug-metabolizing enzyme gene expression in lymphocytes from healthy human subjects. *Drug Metab Dispos* 36, 182-189.
- 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., and Dermitzakis, E. T. (2007). Population genomics of human gene expression. *Nat Genet* 39, 1217-1224.
- Xia, M., Huang, R., Sun, Y., Semenza, G. L., Aldred, S. F., Witt, K. L., Inglese, J., Tice, R. R., and Austin, C. P. (2009). Identification of chemical compounds that induce HIF-1alpha activity. *Toxicol Sci* 112, 153-163.
- Xia, M., Huang, R., Witt, K. L., Southall, N., Fostel, J., Cho, M. H., Jadhav, A., Smith, C. S., Inglese, J., Portier, C. J., Tice, R. R., and Austin, C. P. (2008). Compound cytotoxicity profiling using quantitative high-throughput screening. *Environ Health Perspect.* 116, 284-291.
- Yu, B., Mitchell, G. A., and Richter, A. (2009). Cirhin up-regulates a canonical NF-kappaB element through strong interaction with Cirip/HIVEP1. *Exp Cell Res* 315, 3086-3098.
- Zhou, Y. H., Xia, K., and Wright, F. A. (2011). A powerful and flexible approach to the analysis of RNA sequence count data. *Bioinformatics* 27, 2672-2678.
- Zhu, H., Rusyn, I., Richard, A., and Tropsha, A. (2008). Use of cell viability assay data improves the prediction accuracy of conventional quantitative structure-activity relationship models of animal carcinogenicity. *Environ. Health Perspect.* 116, 506-513.

**CHAPTER 3: POPULATION-BASED IN VITRO HAZARD AND CONCENTRATION-
RESPONSE ASSESSMENT OF CHEMICALS:
THE 1000 GENOMES HIGH THROUGHPUT SCREENING STUDY**

Nour Abdo¹, Menghang Xia², Chad C. Brown³, Oksana Kosyk¹, Ruili Huang², Srilatha Sakamuru², Yihui Zhou^{3,4}, John Jack³, Paul Gallins⁶, Kai Xia⁷, Yun Li^{6,8}, Alison Motsinger-Reif^{3,4}, Christopher P. Austin², Raymond R. Tice⁹, Ivan Rusyn¹, Fred A. Wright^{3,4,5}

¹Department of Environmental Sciences and Engineering, University of North Carolina at Chapel Hill, North Carolina, USA.

²National Center for Advancing Translational Sciences, Bethesda, Maryland, USA.

³Bioinformatics Research Center and Departments of ⁴Statistics and ⁵Biological Sciences, North Carolina State University, Raleigh, North Carolina, USA.

⁶Departments of Genetics, ⁷Psychiatry, and ⁸Biostatistics, University of North Carolina at Chapel Hill, North Carolina, USA.

⁹National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina, USA.

Running title: The 1000 Genomes High Throughput Screening Study.

Disclaimer: The authors declare they have no actual or potential competing financial interests. The views expressed are those of the authors and do not necessarily reflect the statements, opinions, views, conclusions, or policies of the US EPA, NIEHS, NCATS, NIH, or the United States government.

Acknowledgements: This work was made possible by US EPA grants STAR RD83516601 and NIH grants RD83382501, R01CA161608, R01HG006292, and through an interagency agreement (IAG #Y2-ES-7020-01) from NIEHS to NCATS.

A. ABSTRACT

Background: Important gaps exist in our understanding of human variation in response to toxic environmental chemicals. **Objectives:** To address this critical need in next generation risk assessment, we tested a hypothesis that population-wide *in vitro* cytotoxicity screening has the potential to assess both the magnitude of and molecular causes for inter-individual genetic variability in toxicity of chemicals. **Methods:** We used 1086 lymphoblastoid cell lines, representing 9 populations from 5 continents, drawn from the 1000 Genomes Project to assess variation in cytotoxic responses to 179 chemicals used as phenotypes. The analysis included ranking of chemicals by average response and assessments of population variation and heritability, information that is immediately applicable to human health assessment of chemical toxicity. Genome-wide association mapping was also performed, with attention to phenotypic relevance to human exposures. **Results:** The extent of inter-individual variability in cytotoxicity was <10-fold for about 2/3 of the compounds; however, some compounds exhibited >100-fold range in variability. Genetic mapping suggested important roles for variation in membrane and trans-membrane genes with a number of chemicals showing association with rs13120371 in the solute carrier *SLC7A11*, which has been implicated in chemoresistance. Analysis of public RNA-sequencing profiles on the same cell lines provided evidence of association between basal transcription and cytotoxic response, with enrichment for genes with membrane localization. **Conclusions:** This experimental approach fills critical gaps of most recent large-scale toxicity testing programs by providing quantitative experimentally-based confidence intervals for estimating chemical hazard and variability, as well as generating testable hypotheses about potential mechanisms of toxicity.

B. INTRODUCTION

During the last decade, considerable progress has been made in high-throughput approaches for toxicity testing to address the challenges posed by (i) the expense and ethical constraints in animal testing; (ii) uncertainties in applicability of animal models to human susceptibility, and (iii) a large and increasing number of chemicals, many of which have never been subjected to adequate toxicity testing, and to which humans are exposed. Reports by the National Research Council (National Research Council 2007) and other prospective statements (Collins et al. 2008) have articulated a vision for the use of biochemical- and cell-based assays in a high throughput screening format to screen chemicals, providing an improved understanding of toxicity response and modes of action. The *in vitro* testing of human cell lines meets human relevance standards (Collins et al. 2008) while serving as a bridge to targeted *in vivo* assessment. Beyond characterizing an “average” response to chemicals, another goal of next-generation toxicity testing is to provide an improved understanding of population variability in susceptibility, to identify vulnerable subpopulations, and to refine uncertainty factors used in regulatory risk assessment (Zeise et al. 2013).

The Tox21 initiative (Tice et al. 2013) is conducting automated systematic screening of thousands of chemicals against hundreds of molecular and cellular toxicity phenotypes. Cell-based assays for cell viability are an established approach to prioritize chemicals for further evaluation or to classify into hypothesized modes of action (Huang et al. 2008). However, for environmental chemicals the number of cell lines used in these assays has typically been limited to several dozen (Lock et al. 2012; O'Shea et al. 2011), sometimes representing multiple species (Xia et al. 2008). Thus, an understanding of human population variability and, in particular, the role of constitutional genetic variation remains elusive. Epidemiological approaches to these

questions have also been limited to a relatively few chemicals with high occupational or other exposure (Zeise et al. 2013), and have mainly quantified the effects of drug metabolizing enzymes (Ginsberg et al. 2009). Furthermore, an epidemiological approach provides little basis to compare directly across chemicals, including new chemicals with little to no exposure or potential toxicity information.

Screening of lymphoblastoid cell lines (LCLs) is an established approach to identify genetic variants influencing cytotoxic response to pharmaceuticals, especially chemotherapeutic agents (Wheeler and Dolan 2012). Choy et al. (Choy et al. 2008) had challenged the value of these approaches, primarily due to potential confounding influences, including growth rates and batch effects. However, enrichment of human blood eQTLs has been established among weakly significant chemotherapeutic drug susceptibility loci (Gamazon et al. 2010). With the advent of statistical methods purpose-built for cytotoxicity profiling (Brown et al. 2012a), several robust associations were identified (Brown et al. 2014).

For environmental chemicals, quantifying the extent of population variation in cytotoxicity is of great interest, potentially providing data to inform uncertainty factors in risk assessment (Zeise et al. 2013). Direct connections to human risk assessment must consider genetic variation in cytotoxic response at relatively low concentrations relevant to human exposure (Baynes 2012). This goal may conflict somewhat with maximization of power to identify specific genotype-susceptibility associations, as the effects of genetic variation may be apparent only at concentrations higher than human-relevant exposure. Furthermore, for both these goals, the number of cell lines used in past studies of environmental chemical cytotoxicity has often been inadequate to establish population variation, or to assess genetic association for these complex traits with small effect.

Here, we describe profiling 1086 LCLs for cytotoxic response to 179 chemicals, each assayed over a range of 8 concentrations spanning six orders of magnitude. The compounds were primarily chemicals of environmental concern, cover a wide range of *in vivo* toxicity hazards, and were drawn from a larger set of 1408 compounds used for high-throughput screening (Lock et al. 2012; O'Shea et al. 2011; Xia et al. 2008). The LCLs were selected from the 1000 Genomes Project (1000 Genomes Project Consortium et al. 2012) spanning a variety of ancestral populations. Cytotoxic response was assessed using an effective concentration 10% (EC_{10}) and genome-wide association mapping was performed using both EC_{10} and with an omnibus test using the entire 8-concentration profile as a multivariate vector.

C. MATERIALS AND METHODS

Chemicals and cytotoxicity profiling. Chemicals were chosen as a subset of the National Toxicology Program's 1408 chemical library. Chemicals were dissolved with dimethyl sulfoxide (DMSO) into 8 stock concentrations and were transferred into 1536-well plate format via a pin tool station (Kalypsys, San Diego, CA). The final concentrations ranged from 0.33 nM to 92 μ M. The negative control was DMSO at 0.46% vol/vol; the positive control was tetra-octyl-ammonium bromide (46 μ M). The CellTiter-Glo Luminescent Cell Viability (Promega, Madison, WI) assay was used to assess intracellular ATP concentration, a marker for viability/cytotoxicity, 40 h post treatment. A ViewLux plate reader (PerkinElmer, Shelton, CT) was used to detect luminescent intensity in each well of the assay plates.

Cell lines. A set of 1104 immortalized lymphoblastoid cell lines established by the HapMap Consortium and the 1000 Genomes Project (1000 Genomes Project Consortium et al. 2012) was acquired from Coriell (Camden, NJ). Out of the 1104, 401 cell lines were related individuals of trios (both parents and child). Cell lines were randomly divided into screening batches with equal distribution of populations and gender in each batch and without regard to family structure. Cells were cultured at 37⁰C with 5% CO₂ in RPMI 1640 media (Invitrogen, Carlsbad, CA) supplemented with 15% fetal bovine serum (HyClone, South Logan, UT) and 100U/ml penicillin/100mg/ml streptomycin (Invitrogen). Media was replaced every 3 days. Cells were grown to a concentration of up to 10⁶ cells/ml and viability of >85% before treatment. Cells were plated into a tissue culture-treated 1536-well white/solid bottom assay plates (Greiner Bio-One North America, Monroe, NC) at 2000 cells/5 μ l/well using a flying reagent dispenser (BioRAPTR FRD dispenser, Beckman Coulter, Carlsbad, CA). Each cell line was seeded on

multiple plates (1-2 plates per batch and/or between batches) so that each compound was screened in each cell line on 1-2 plates (all chemicals were fit to a single plate).

Genotypes. The primary source of genotypes was the Illumina HumanOmni2.5 platform for 1000 Genomes (<http://www.1000genomes.org> and ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/technical/working/20120131_omni_genotypes_and_intensities), available for >90% of the samples. SNPs with a call rate below 95%, minor allele frequency (MAF)<0.01, or HWE p-value<1x10⁻⁶ were excluded. From the original 1086 samples, a maximal subset of 884 was chosen so as to remove first-degree relatives ('unrelated' set) using a combination of genotypes and sample annotation. Of the 884 samples, genotyped SNPs from the Illumina HumanOmni2.5 platform were available for 761 samples. The remaining 123 samples had been genotyped as part of the HapMap project (http://hapmap.ncbi.nlm.nih.gov/downloads/genotypes/hapmap3_r3/plink_format), and were imputed for the set of filtered Illumina SNPs using the MACH software. A subset from a larger 1000 Genomes set (totaling 875 samples, not restricted to these cell lines) were used as a reference for this genotype imputation. The final set of 1.3m SNPs were used for primary analysis. A further subset of 690 unrelated individuals represented Phase I individuals from 1000 Genomes (1000 Genomes Project Consortium et al. 2012), with more complete sequencing data available. After applying the quality filters, a total of 12m SNPs remained.

Cytotoxicity EC₁₀ estimation and outlier detection. Cytotoxicity data were normalized relative to positive/negative controls as described. Although the primary method for association mapping was based on a multivariate treatment of the cytotoxicity response values across the entire set of dose concentrations for each chemical, it is convenient for other analyses to provide a single

cytotoxicity dose summary per chemical and cell line. We devised an effective concentration 10th percentile (EC₁₀), using the logistic model

$$\log\left(\frac{y - \theta_{min}}{\theta_{max} - \theta_{min}}\right) = \beta_0 + \beta_1 d + \varepsilon,$$

with $\varepsilon \sim N(0, \sigma^2)$, where y is the observed normalized signal representing proportion of surviving cells (which we term the “cytotoxicity value”), d is the log(concentration) for each chemical, and θ_{max} is the mean cytotoxicity value on the logit scale for the zero concentration. θ_{min} was set to zero, to avoid difficulties in estimating the minimum cytotoxicity value for chemicals with low cytotoxicity. An exception was made for chemicals in which the cytotoxicity value at the highest concentration was higher than 0.4, as a very few number of plates/chemicals did not reliably reach maximum cytotoxicity. In those instances the cytotoxicity value was set at the observed cytotoxicity at the maximum concentration. Inspection of these data revealed good fits in such instances according the maximum likelihood estimation. Although in principle θ_{max} should have been 1.0, a number of plates exhibited a drift from this value, and thus the parameter was estimated from the data.

Fitting for the parameters $[\beta_0, \beta_1, \sigma^2, \theta_{max}]$ proceeded by maximum likelihood using numerical optimization in *R* v2.15. An automatic outlier detection algorithm was devised by considering the impact of dropping each concentration value in succession, and removing those values for which the maximum likelihood improved by a factor of 10 or more (example in **Figure 3.1** and refitting the model using the non-outlying observations.

Multivariate Association Analysis. The MAGWAS multivariate analysis of covariance model (Brown et al. 2012a) was used for primary association mapping. The approach allows for use of

the full concentration-response profile, as opposed to a univariate summary (such as EC₁₀) as a single response, with the advantage of robustness and power under a wide variety of association patterns. The model used for association for the j th individual and genotype i for the chemical/SNP is

$$Y_{ij} = \mathbf{X}_{ij}\beta + \mu_i + \mathbf{e}_{ij}$$

$$\mathbf{e}_{ij} \sim N(\mathbf{0}, \Sigma),$$

where Y_{ij} is the vector of responses (across the eight concentrations) for the j^{th} individual having genotype i , \mathbf{X}_{ij} is the design matrix of covariates, including sex, indicator variables for laboratory batch, and the first ten genotype principal components, and μ_i is the eight concentration-vector of parameters modeling the effects of genotype i on the response. The model assumes that the error terms are multivariate normally distributed, with mean vector $\mathbf{0}$ and variance-covariance matrix Σ , allowing for dependencies in the observations. P -values were obtained using Pillai's trace (Pillai 1955). Because this method makes use of asymptotic theory, markers with fewer than 20 individuals representing any genotype were removed, leaving 692,013 SNPs for analysis.

Heritability. Estimation of the proportion of chemical response variation due to genetic variation (heritability) was first calculated for each compound using the mean of the batch adjusted EC₁₀ value across the 401 related individuals belonging to the related individuals (trios). The Multipoint Engine for Rapid Likelihood Inference (MERLIN) (Abecasis et al. 2002) software package was used to estimate heritability with and without additional covariates including subpopulation by ethnicity (CEU, MXL and YRI) and population stratification (top three principal components). The covariates did not have a substantial effect on the heritability

estimates (results not shown). Additionally, variance component analysis and hypothesis testing were performed with Sequential Oligogenic Linkage Analysis Routines (Almasy and Blangero 1998), in order to evaluate the significance and standard error for each heritability estimates per compound. A false discovery threshold of 0.05 was used to ascribe significance and control for multiple testing.

Using the set of 884 unrelated individuals, we also ran the GCTA package (Yang et al. 2011) to estimate heritability, with default settings and using the 1.3m SNPs. To assess whether the degree of concordance between MERLIN and GCTA was at the level expected, we used the 179-vector of MERLIN heritability estimates as a hypothetical true set of heritabilities. These “true” values and associated standard errors from both MERLIN and GCTA were used to simulate independent normal errors to create 10,000 paired vectors of MERLIN and GCTA estimates, which were then compared for concordance.

RNA-Seq data. For dataset E-GEUV-1, mapped reads were downloaded (BAM format) from ArrayExpress (<http://www.ebi.ac.uk/arrayexpress>). IsoDOT (<http://www.bios.unc.edu/~weisun/software/isoform.htm>) was used to count reads for each non-overlapping exon, which have been preprocessed in IsoDOT library files (http://www.bios.unc.edu/~weisun/software/isoform_files/). Read counts for each gene were obtained by summing the read count of the corresponding exons.

Prediction of EC₁₀ using RNA-Seq. The RNA-Seq read count data were normalized by the library size for each sample, and principal components were computed from these data to use as predictors in LASSO-based regression using the R package *lars* v1.2, with each chemical’s EC₁₀ values used in succession as a response. Default cross-validation prediction error estimation was

performed within the package for each chemical. Following theoretical work on avoiding bias in principal component estimation (Lee et al. 2010), the entire set of principal components was computed once and used in the cross-validations.

Attenuated variability estimates to account for sampling variation. To account for the inflationary effect of sampling variance on expected ranges of EC_{10} values, we considered the simple model: $EC_{10} = \mu + \varepsilon$, where μ is the underlying true (unknown) EC_{10} and ε reflects sampling variation in the estimate. We suppose that each chemical has an underlying true sampling variability σ_s^2 per observation, while observed EC_{10} values were, in many instances, averaged across multiple observations. If an individual is measured n' times, $var(\varepsilon) = \sigma_s^2/n'$. To conservatively estimate σ_s^2 , we identified all paired replicate instances for the chemical across different batches, computed the sample variance within each pair, and average across these pairs to obtain $\widehat{\sigma_s^2}$. Then we computed a variance inflation factor

$$VIF = \frac{var(EC_{10})}{\widehat{\sigma_s^2}/mean(n')}$$

where $mean(n')$ is the average number of replicates per individual. Finally, we consider individual measurements to have been inflated by \sqrt{VIF} so that, for example, the deflated inter-susceptibility range is $(\log(q_{95}) - \log(q_{05}))/\sqrt{VIF}$.

D. RESULTS

Cell lines and genotypes

An initial set of 1104 LCLs (1000 Genomes Project Consortium et al. 2012) was representative of 9 geographically- and ancestry-diverse populations: Utah residents with Northern & Western European ancestry (CEU); Han Chinese in Beijing, China (CHB); Japanese in Tokyo, Japan (JPT); Luhya in Webuye, Kenya (LWK); Mexican ancestry in Los Angeles, CA (MXL); Tuscans in Italy (TSI); Yoruban in Ibadan, Nigeria (YRI); British from England and Scotland (GBR); and Colombian in Medellin, Colombia (CLM). Genotypes were obtained from the 1000 Genomes web site. A few cell lines (18; 1.6%) were not viable or grew very slowly, or had insufficiently available genotypes, and the final data analysis set consisted of 1086 cell lines.

Due to sample size limitations in comparison with modern disease genome-wide association studies, and to reduce multiple comparisons, we initially focused on ~2.5 million markers further filtered by minor allele frequency (MAF). 172 of the individuals had not been genotyped on the platform, and so MaCH (Li et al. 2010) dosage imputation to these markers was performed using the appropriate 1000G reference population. Analyses were performed separately on 400 individuals belonging to parent-child trios (not all complete) in the CEU (164), MXL (83), and YRI (153) populations, and on a maximal set of 884 individuals in the remaining populations with no first-degree relationships (unrelateds). Association analyses were performed using both the Omni 2.5 set with $MAF \geq 0.05$ (~1.3m) and a larger set (~12m) of typed SNPs available from the sequencing data, as described below.

Figure 3.2a shows the distribution of populations and continental ancestry. LCLs were randomly divided into screening batches with equal distribution of populations and sex in each

batch, without regard to family structure. The major HapMap/1000G continental ancestry populations were represented, as well as admixed populations from the Americas (**Figure 3.2b**).

Cytotoxicity profiling

Figure 3.3 shows a flow chart of cytotoxicity profiling across 8 concentrations ranging from 0.33 nM to 92 μ M. Logistic curve fitting with outlier detection (**Figure 3.1**) was used to obtain EC₁₀ values, which were batch-corrected and averaged across replicates for each cell line.

To place the current study in context with other cell line cytotoxicity genetic mapping studies, we reviewed comparable studies, identifying 18 reports (Lock et al. 2012; O'Shea et al. 2011; Gamazon et al. 2010; Wheeler et al. 2013; Innocenti et al. 2009; Stark et al. 2010; Aksoy et al. 2009; O'Donnell et al. 2010; Li et al. 2009; Peters et al. 2011; Huang et al. 2011; Wheeler et al. 2011; Kulkarni et al. 2012; Brown et al. 2012b). These studies had (i) more than one chemical (except (Brown et al. 2012b)), and (ii) at least 50 cell lines. **Figure 3.4a** depicts a heatmap of our cytotoxicity measurements across cell lines and chemicals. The figure also depicts, to scale, the size of the other 18 studies in terms of cell lines X chemicals/drugs. In these terms, the size of the study reported here is an order of magnitude greater than any single previous study, and several times larger than the other reports combined.

For the ~700 cell lines for which there was at least one replicate plate, **Figure 3.4b** depicts the EC₁₀ values for replicates ($r=0.90$). Nine of the chemicals were assayed in duplicate on the each plate, and duplicate chemicals showed similar median EC₁₀ and range of variability across cell lines (**Figure 3.4c**). The entire range of EC₁₀ values across all chemicals exhibited remarkable variation in cytotoxicity (**Figure 3.4d**). Only one other report has been of similar scale in chemicals represented (240 chemicals investigated in (Lock et al. 2012)). However, our

comparisons across chemicals are much more definitive in the ability to rank and prioritize compounds by cytotoxic activity, due to the large number of cell lines profiled ($n=1086$ in the current report, vs. $n=81$ in (Lock et al. 2012)).

Figure 3.5a shows the results of EC_{10} estimation for all cell lines for an illustrative chemical, as well as the results from the logistic fit applied to the pooled data (similar to (Lock et al. 2012)). The histogram depicts the individual EC_{10} estimates, showing variation of more than an order of magnitude. To quantify “susceptible” subgroups in the population, we recorded the 5th and 95th percentiles of EC_{10} values for each chemical, and the quantile difference $q_{95}-q_{05}$ on the log scale represents a *fold-range* of population variability. **Figure 3.5b** shows the fold-range across the 179 chemicals, as a histogram and as a function of mean EC_{10} values (inset). The fold-range is low for the chemicals at the high and low extremes due to measurement constraints at the extremes. Overall, the measured fold-ranges are expected to be inflated due to technical sampling variation, and the figure also shows an estimate of the true underlying distribution after removing this source of variation. The majority (166; 92.7%) of the chemicals show a shrunken fold-range less than 10. For a few chemicals, the estimated fold-range in cytotoxicity is much greater than 10, and may be as great as 100 or more (**Table 3.1**).

For each population, we produced a profile of EC_{10} values across the 179 chemicals by averaging for each chemical across the individuals within the population. Hierarchical clustering of the averaged profiles (**Figure 3.5c**) showed general assortment by continental ancestry, although variation within populations was generally greater than across populations. While a large number of chemicals showed significant EC_{10} variation across populations or by sex (false discovery rate $q<0.05$, **Table 3.2**), this variation tended to be modest (two examples in **Figure 3.5d**).

Heritability and mapping

Trio-based analysis provided significant evidence of additive heritability for 22 chemicals ($q < 0.05$), with significant h^2 ranging from ~ 0.25 to ~ 0.5 (**Figure 3.5e**). This analysis was augmented by essentially independent heritability estimation using GCTA (Yang et al. 2011) performed using the maximal set of 884 unrelated individuals. GCTA-based h^2 ranged from ~ 0.4 to 0.8 for 34 significant chemicals (**Figures 3.5a-b**).

Estimation of the EC_{10} as a phenotype was motivated by relevance to human health assessment practices (Baynes 2012); however, elucidation of the underlying genetic mechanisms may be more powerful without assumptions about the point-of-departure. Moreover, the EC_{10} is not sensitive to genetic influences that are apparent only at high concentrations. We thus adopted a three-stage approach to mapping, using ten genotype principal components and sex as covariates. For the primary analysis, using the set of unrelated individuals, we applied the multivariate MAGWAS approach (Brown et al. 2012a), which is sensitive to any pattern of variation of cytotoxicity measurements due to genotype. Second, for the same individuals, we used EC_{10} values as a quantitative phenotype in regression analysis for an additive model of SNP allele effects, using the larger set of 1.3m SNPs (chr1-X). At the level of individual SNPs, this analysis was used to identify significant associations that might have been missed by MAGWAS and to investigate pathway-based associations (Schaid et al. 2012). Finally, in order to capture a larger number of SNPs and variants with lower MAF (Gamazon et al. 2012), we applied the EC_{10} regression approach to 690 of the unrelated individuals who were part of 1000 Genomes Phase I (1000 Genomes Project Consortium et al. 2012), and used ~ 12.4 m variants with $MAF \geq 0.01$.

Preliminary examination of the last analysis indicated phenotype outlier effects causing spurious significant findings due to the lower MAF threshold, and so after applying an initial filter of association $P < 5 \times 10^{-8}$, the (chemical)×(SNP) analyses were recomputed after applying an inverse quantile normalization to the EC_{10} values.

For the primary analyses, each chemical is deemed worthy of separate investigation. Thus, we reasoned that a balanced approach to multiple comparisons was to apply per-chemical false discovery control, following proposals that SNPs with $FDR\ q < 0.10$ be declared significant (van den Oord and Sullivan 2003). **Table 3.3** shows these 48 chemical-SNP associations, after removing redundant regional findings within ± 1 Mb. The nearest gene is reported, along with partial R^2 , the portion of variance explained in the MAGWAS model across the entire 8 concentrations after considering covariates. The most significant MAGWAS findings tend to have larger partial R^2 , but the relationship to P -values is complex (**Figure 3.6**).

Table 3.3 is presented for each chemical, but a re-ranking by P -values reveals that the top 10 significant associations includes three solute carriers (*SLC7A11* for 2-amino-4-methylphenol, *SLC39A14* for 1,3-dicyclohexylcarbodiimide, and *SLCO3A1* for titanocene dichloride), the transmembrane protein *TMEM196* for N-isopropyl-N'-phenyl-p-phenylen, and *NFAT5* for o-aminophenol, a gene which activates several solute carriers in response to osmotic stress (Haltermann et al. 2012). These findings suggest a major role for membrane proteins and solute carrier transporters in mediating cytotoxicity, as has been reported for the chemotherapeutic agent paclitaxel (Njiaju et al. 2012).

The most significant MAGWAS association ($P = 8.4 \times 10^{-10}$) was for 2-amino-4-methylphenol at rs13120371 in the 3' UTR of *SLC7A11*, a cystine and glutamate transporter also

known as xCT . The result was highly significant on a per-chemical basis ($q=0.0006$), and at the significance threshold for multiple comparison correction for the entire set of SNPs X chemicals combined ($q=0.10$). **Figure 3.7a** shows the corresponding MAGWAS Manhattan plot, with a regional plot (**Figure 3.7b**). The same SNP also appeared with $q<0.10$ for methyl mercuric chloride and N-methyl-p-aminophenol sulfate (**Table 3.3**). Comparative curves for 2-amino-4-methylphenol and the three genotypes show that the difference in cytotoxic response appears mainly at the highest concentration (**Figure 3.7c**). The plot illustrates the contrast between EC_{10} , which does not differ significantly by genotype, and the multivariate MAGWAS finding, which is sensitive to concentration-response variation.

For the secondary analysis using regression of EC_{10} on SNP genotype, we computed the genomic control value λ (Devlin and Roeder 1999). The mean λ (\pm s.d.) across chemicals was 0.988 ± 0.017 suggesting that population stratification was well-controlled. **Table 3.4** shows results from the EC_{10} regression analyses, with all significant findings (per-chemical $q<0.10$) shown after removing redundant regional findings (63 unique chemicals, 260 unique nearest gene assignments). For many chemicals, the effects of genotype are observable both for EC_{10} and across the multivariate response, and the two approaches provide similar evidence (**Figure 3.8**). At the false discovery rate (Storey and Tibshirani 2003) of <0.1 only ~ 18 unique chemicals would be expected to appear in the table. SNPs in four genes appear in the table for three or more chemicals: *GRIPI* (Glutamate receptor interacting protein 1), which directs localization of transmembrane proteins (Setou et al. 2002); *FMN2*, a component of p21-based cell cycle arrest (Yamada et al. 2013); *DNER*, a transmembrane protein associated with glioblastoma propagation (Sun et al. 2009); and the cell membrane cadherin *CDH13*, which acts as an epithelial tumor suppressor (Chan et al. 2008). As with the MAGWAS results, membrane localized proteins

appear to play an important role. Because EC_{10} were available for 179 chemicals, we could also illustrate that GCTA-based heritability estimates are largely reflected in a tendency toward small association P -values, a phenomenon that is difficult to discern for single-trait GWAS studies (**Figure 3.9c**). **Table 3.5** shows the significant associations for the analysis of the larger number (12.4m) of sequenced SNPs.

With regards to rs13120371 in *SLC7A11*, we hypothesized that it may modify resistance to a larger number of chemicals. We examined the EC_{10} P -values for rs13120371 across all 179 chemicals, observing a clear excess of small P -values (**Figure 3.7d**). Using a standard false-discovery computation (Storey and Tibshirani 2003), we estimated the proportion of true discoveries for the SNP across the chemicals as 0.25, a significant trend that remained even after removal of the three top MAGWAS-identified chemicals. The estimated number of true discoveries, corresponding to an estimated 44 chemicals showing true cytotoxicity association with rs13120371, is subject to considerable sampling variation. Nonetheless, the data indicate that *SLC7A11* may be a cytotoxicity mediator, and a role for *SLC7A11* has been proposed in glutathione-mediated chemotherapeutic resistance (Huang et al. 2005).

“Pathway” association analysis of gene sets/ontologies was performed for EC_{10} phenotypes and the 1.3m Omni 2.5 SNPs using *gene set scan* (Schaid et al. 2012) which computes significance of SNPs, genes, and ontologies (KEGG and Gene Ontologies). Eleven chemicals had significant pathways, and several chemicals showed significant associations with immune-response pathways and ontologies (**Table 3.6**) at family-wise error rate (FWER) <0.05 .

RNA-Seq analysis

For 344 cell lines in the study, RNA-Seq expression data on 36,142 genes was available (Lappalainen et al. 2013). We performed association analyses of normalized expression vs. EC₁₀ values for each of the 179 chemicals, with per-chemical false discovery $q < 0.1$. A total of 260 genes met the threshold for at least one of 52 chemicals (**Figure 3.10a**).

Analysis of these genes by DAVID/EASE (Dennis et al. 2003) suggested enrichment for ribosomal proteins and those involved with actin binding and nucleotide binding. However, the use of such “gene-list” methods has been shown to greatly increase false-positive rates due to gene-gene correlations (Gatti et al. 2010). Accordingly, we performed analysis of Gene Ontology Cellular Components terms using the *safeExpress*, which accounts for correlation structures in gene expression (Zhou et al. 2013). **Figure 3.10b** shows the results for 22 chemicals and 33 terms, with at least one significant per-chemical finding at $q < 0.1$. Several chemicals exhibit enrichment of genes with protein localization to the cell membrane, in addition to other localizations.

The RNA-Seq findings support the underlying theme that genes with cell membrane localization are important to cytotoxic susceptibility, with transcription acting as a mediator to genetic variation. However, the ability to jointly model the effects of transcriptomic and genetic variation is hampered by sample size limitations, and the overall association of EC₁₀ with expression is relatively modest. Using the entire expression dataset to predict EC₁₀ for each chemical using a penalized prediction procedure, only three chemicals exhibited more than 5% explained variation (**Table 3.7**). The explained variation among these few chemicals may be

exaggerated due to winner's curse inflation, and the average explained variation across the chemicals was less than 1%.

E. DISCUSSION

Our study, with a large number of environmental chemicals interrogated on a much larger sample than previous datasets, demonstrates the feasibility of addressing the challenge of assessing the individual variability in next-generation risk assessments (Zeise et al. 2013). Despite concerns over the ability to map meaningful drug response traits in cell lines (Choy et al. 2008), our results suggest that large sample sizes, on the order necessary for mapping human complex traits (Goldstein 2009), can overcome the challenges. Importantly, these data are also a useful resource for future investigation. Our results generated dozens of associations about the mechanism of toxicity for hundreds of chemicals that can be used to generate testable hypotheses.

Although we present results as a survey across the 179 chemicals, the results for population variation and association for each chemical will be useful for future targeted investigations. Moreover, the use of a common protocol and automated system enabled comparisons across chemicals in a manner that is difficult to perform across separate studies of individual chemicals. Cytotoxicity in LCLs is just one among multiple criteria that may be used in prioritization and to refine uncertainty factors (Zeise et al. 2013), but fills a great need for additional data on population variation (Collins et al. 2008).

Beyond the immediate utility of such data for context-specific decisions in human health assessment of chemicals, we were able to discover important biological associations. We have observed that genes with protein localization to cell membranes, including solute carriers, are enriched in a subset of chemicals, both in genetic association and RNA-Seq association analyses. Solute carrier transporters have been investigated as potential mediators of cytotoxicity for chemotherapeutic agents (DeGorter et al. 2012), with specific genes investigated for paclitaxel

(Njiaju et al. 2012), oxaliplatin (Zhang et al. 2006), gemcitabine (Marce et al. 2006). These genes control cellular influx and efflux of drugs and toxicants, and thus are a plausible source of genetically-induced variation in susceptibility. Moreover, several families of solute carriers are important mediators of toxicity in liver and kidney (DeGorter et al. 2012), which are the primary relevant organs for a large number of the chemical studies. To our knowledge, our study is the first to highlight membrane transporters as potentially important in a wider range of chemical compounds, beyond the cancer chemotherapeutic agents that have been the subject of most LCL mapping studies.

The significant association results for rs13120371 in *SLC7A11* were interesting, and supported by a growing literature on the direct and indirect importance of *SLC7A11* in chemoresistance (Lo et al. 2008). Small interfering *SLC7A11* RNA, including x(c)(-) inhibitors, have been shown to increase sensitivity to various agents in various cancer cell lines (Pham et al. 2010). Expression of *SLC7A11* is altered in drug-resistant ovarian cancer cell lines (Januchowski et al. 2013), is downregulated in response to thymoquinone in breast cancer cells (Motaghed et al. 2014), and is clinically predictive of poor survival in hepatocellular carcinoma and glioblastoma (Kinoshita et al. 2013). Also, *SLC7A11* was shown to be inversely correlated with clinical outcome in bladder cancer, and to be negatively regulated by the micro RNA *miRNA-27a* for cisplatin-resistant cell lines (Drayton et al. 2014).

Conclusions: This study provides an example of how a large-scale systems biology experiment (toxicity phenotyping, genetic mapping, and correlation with basal gene expression data) can aid in translation to public health protection. Testing human cell lines that represent various

populations and identify sensitive subpopulations can substantially improve any risk assessment paradigm where decision making will be relied on actual estimated variability rather than mere statistical assumptions that may or may not be satisfied. Traditional GWAS studies are concerned with uncovering genetic modifiers that may underpin a person's susceptibility to a particular disease. At the same time, little information on how much inter-individual variability may exist in the population for a particular chemical-associated adverse health outcome, even though current risk assessment practices favor data-driven estimation. The use of genetically-defined/-diverse models in chemical safety/toxicity testing is uncommon primarily because of the complexity of such studies. Although the risk assessment process is shifting toward greater reliance on *in vitro* data, none of the *in vitro* assays in Tox21, ToxCast, or other large-scale screening programs is designed to address individual variability (Rusyn and Daston 2010). The availability of the genetically-diverse, genetically-defined renewable source of human cells, such as LCLs from the HapMap and 1000 Genomes, opens an opportunity for *in vitro* toxicity testing at the population scale. Our heritability estimates show that genetic variation may have a profound effect on differences between individual cell lines, and that such variability can be quantified and used to generate testable hypotheses about the mechanisms of toxicity.

FIGURES

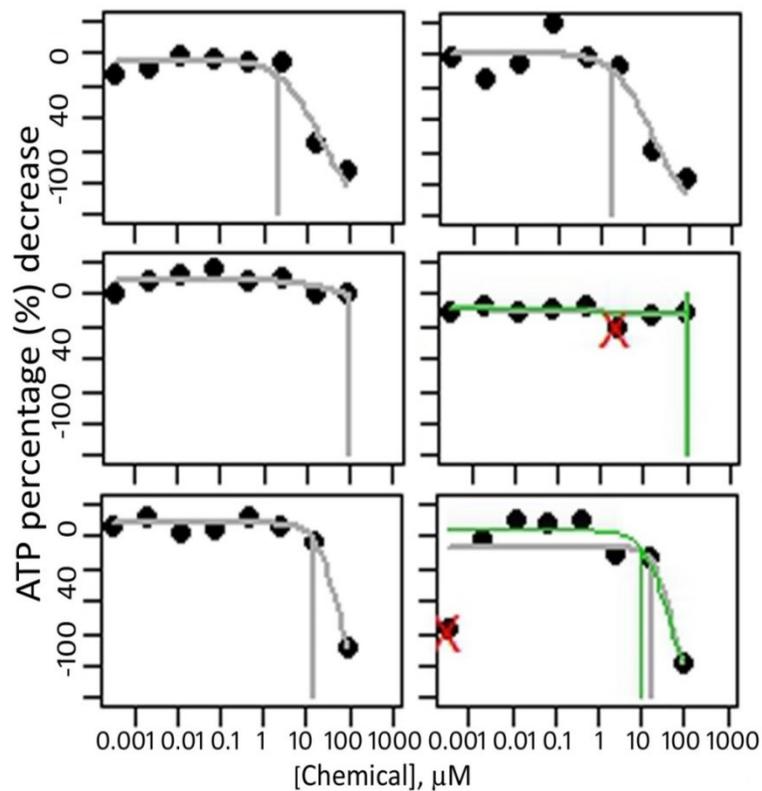


Figure 3.1. Illustrative Fits for Cytotoxicity Estimation. The fits and EC10 point estimate (vertical lines) shown in grey. The top and bottom rows show instances of cytotoxic compounds. The middle panels show compounds that are non-cytotoxic for the range of concentrations, and EC10 was fixed at the maximum concentration. Points marked in red were excluded on the basis of the likelihood ratio criterion described in Online Methods, providing new fits and EC10 values shown in green.

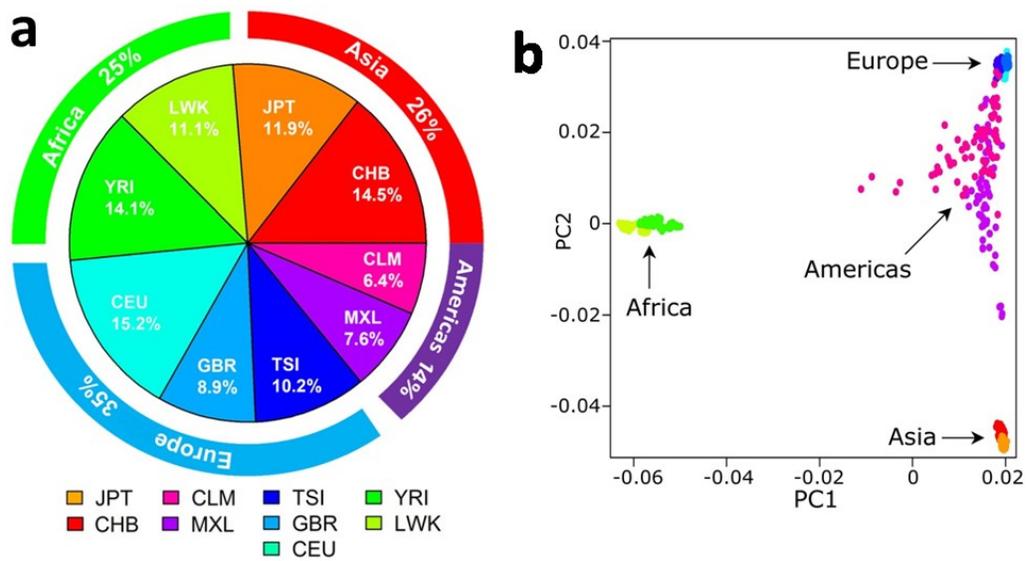
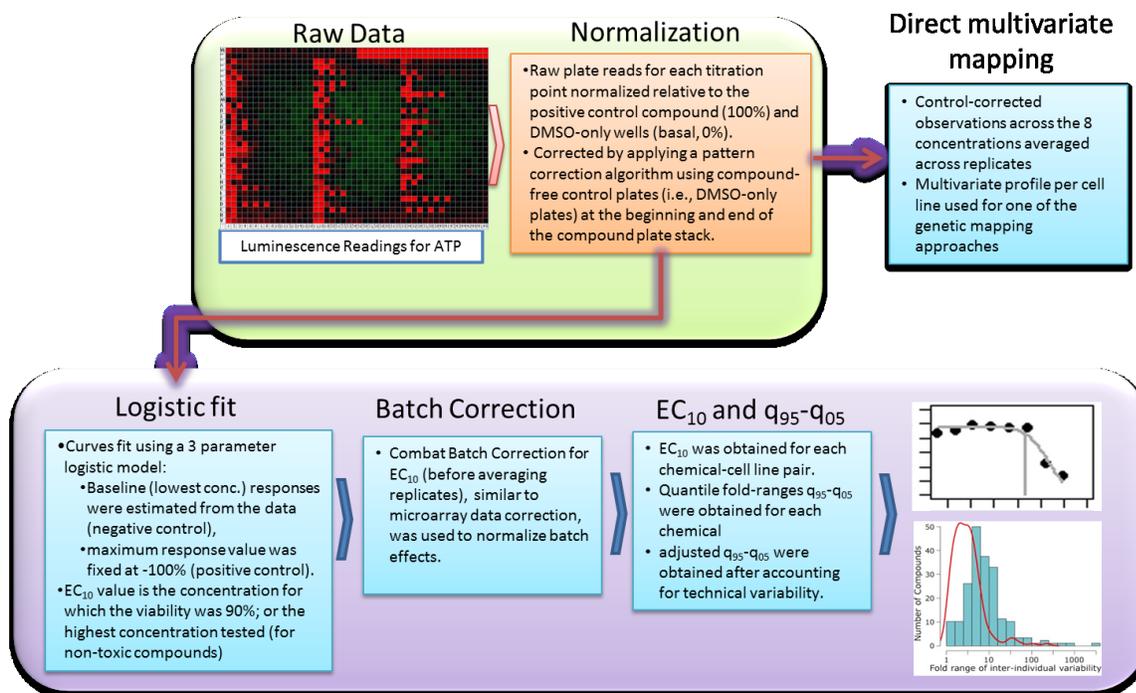


Figure 3.2. Distributions of the LCLS among the 9 Populations (a) Distribution of the LCLS among the 9 populations used in this study. Abbreviations follow the 1000 Genomes nomenclature (see text). Outer boundaries show continental/ancestral origin. (b) Scatter plot for the 1st and 2nd principal components for genotypes across all cell lines, colored by population.



Supplementary Figure 1. Flow chart of data processing to obtain cytotoxicity response values and EC_{10} .

Figure 3.3 Flow Chart of Data Processing to Obtain Cytotoxicity Response Values and EC_{10} .

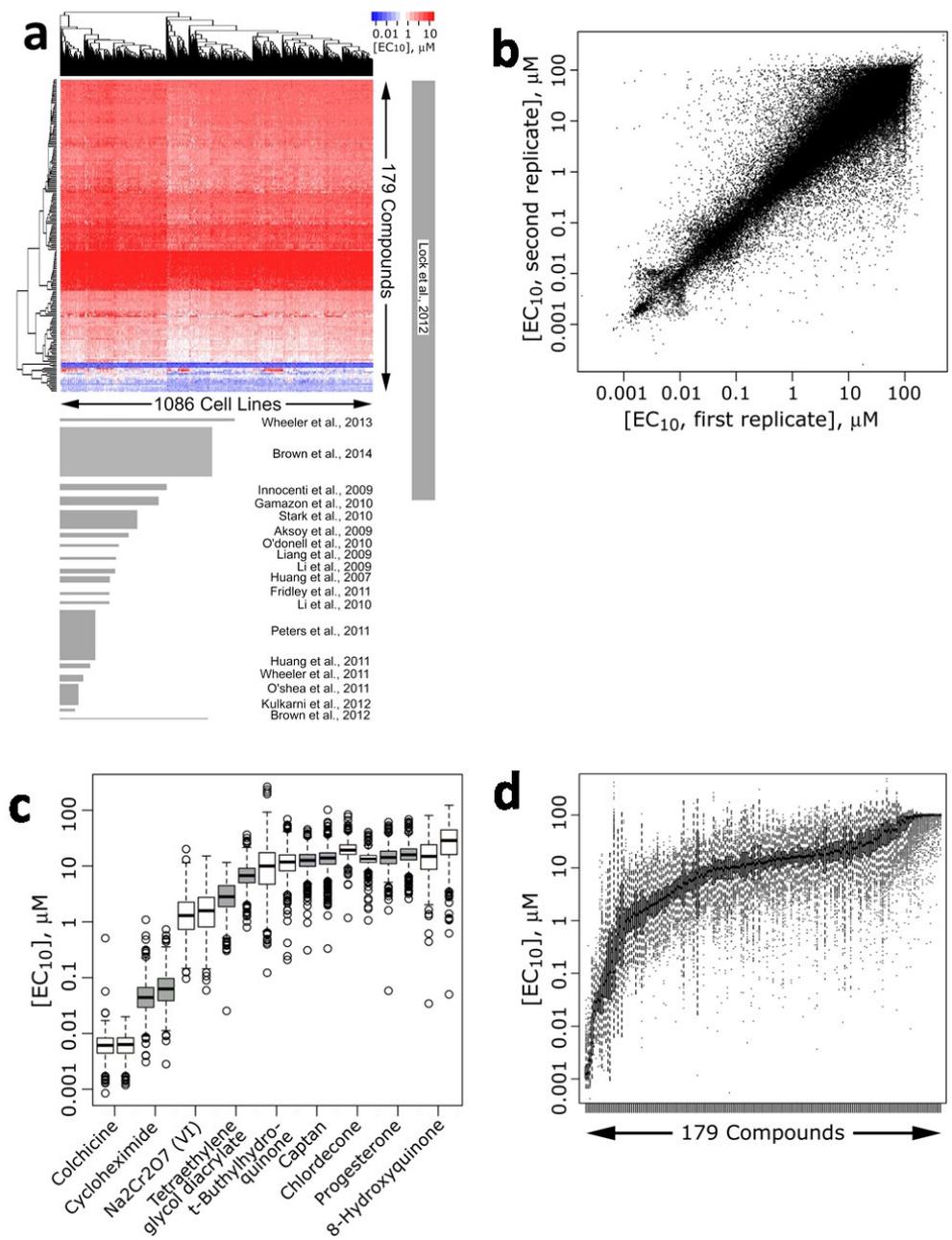


Figure 3.4 Comparison of the Current Study to Other Comparable LCL cell line/screening studies and Reproducibility (a) Comparison of the current study to other comparable LCL cell line/screening studies, in terms of the number of cell lines and chemicals screened. EC₁₀ values are shown in the heatmap, while the area of each depicted report is in proportion to the current

study. Published studies with at least 2 compounds (except for the largest single drug study, Ref. 28), and with at least 50 cell lines, are used for comparison. (b) Intra-experimental reproducibility of EC_{10} values for randomly selected pairs of within-batch replicate plates for all chemicals and cell lines. (c) Side-by-side boxplots show 9 compounds that were assayed in two independent sets of wells on each plate. (d) Boxplots of cytotoxicity EC_{10} values for the 179 chemicals (arranged by mean activity) across the 1086 cell lines.

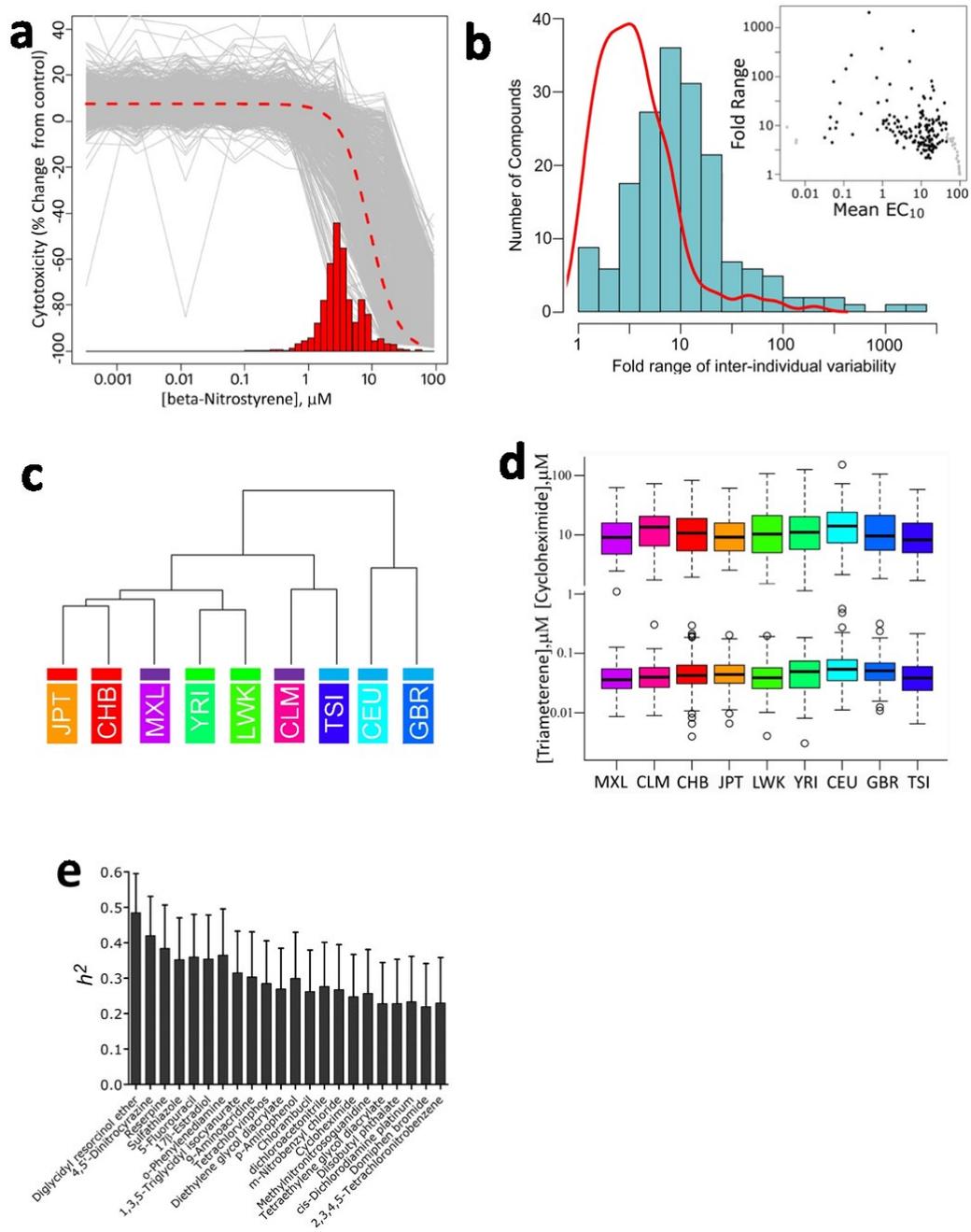


Figure 3.5. Magnitude of Inter-Individual Variability Across Chemicals and Populations

(a) Modeling *in vitro* quantitative high-throughput screening data, using β -nitrostyrene as an

example chemical. Logistic dose-response modeling was performed for each individual (plate) to the values shown in gray, providing individual 10% effect concentration estimates (EC_{10} , histogram). The fit of the logistic model to the pooled data is also shown as a dashed curve, and EC_{10} estimation based on this curve is similar to the average of the individual EC_{10} values). (b) A histogram (blue bars) of EC_{10} fold-range (q95-q05) for 179 compounds across 1086 cell lines. The red curve shows the same distribution when values are shrunken to account for technical variability. The inset shows the relationship between fold-range and mean estimated EC_{10} for each chemical. (c) Hierarchical clustering for the 179-length profiles of mean EC_{10} , computed within each population. The upper bar's color depicts continental ancestral origin of each population. (d) Boxplot of EC_{10} values by population for 2 example chemicals with different potency levels, which showed significant population differences by ANOVA (cycloheximide, $P=6.0 \times 10^{-6}$ and triamterene, $P=3.6 \times 10^{-4}$). (e) Trio-based heritability estimates (h^2) for compounds with significant additive heritability ($q < 0.05$, 22 out of 179 compounds tested).

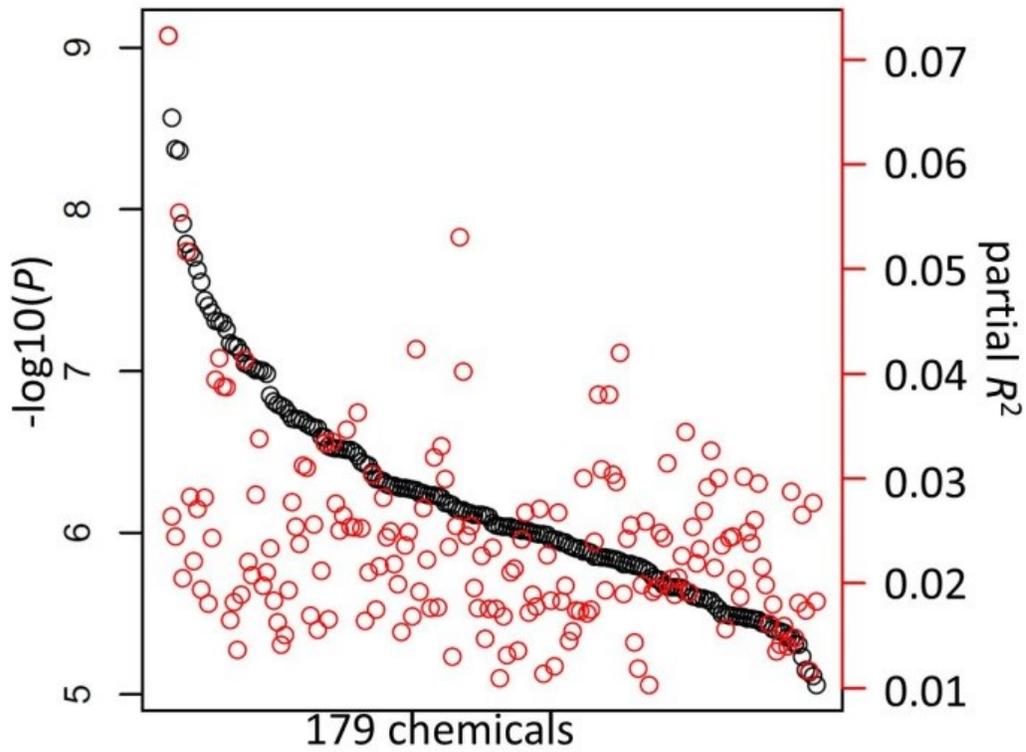


Figure 3.6 *P*-values vs. partial R². A plot of the most significant genetic variant for each chemical, with black dots depicting the $-\log_{10}$ p-values for the association, and red dots showing the maximum partial R-squared across all 8 concentrations.

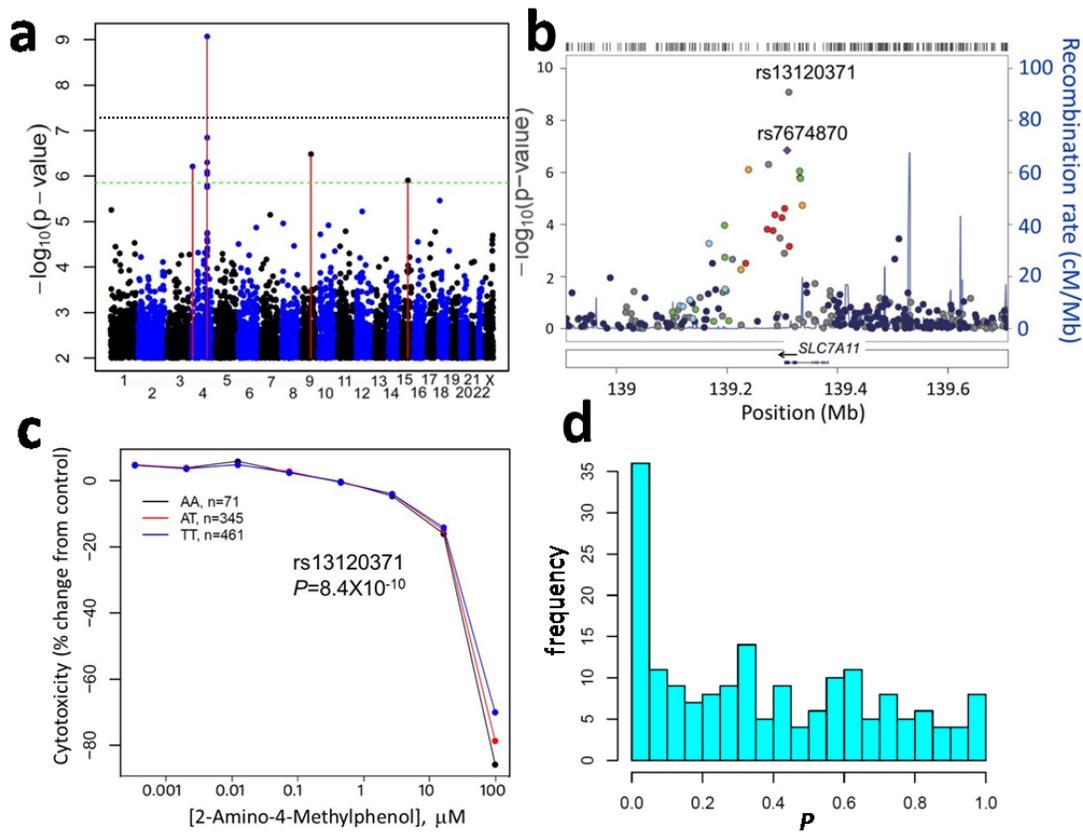


Figure 3.7. Genotype-Cytotoxicity Associations (a) Manhattan plot of MAGWAS $-\log_{10}(P)$ vs. genomic position, for association of genotype and cytotoxicity to 2-amino-4-methylphenol. The line of suggestive association (expected once per genome scan) is in green, and Bonferroni-corrected significance for a single chemical in black. (b) LocusZoom plot of the most significant region. SNP rs13120371 was the most significant ($P=8.4 \times 10^{-10}$), while the nearby rs7674870 was used for comparison of linkage disequilibrium patterns in the region. (c) Average concentration-response profiles of cytotoxicity of 2-amino-4-methylphenol plotted separately for each rs13120371 genotype. Genotype effects appear only for the highest concentrations. (d) Histogram of EC10-based P-values for all 179 chemicals for rs13120371 shows an excess of small P-values.

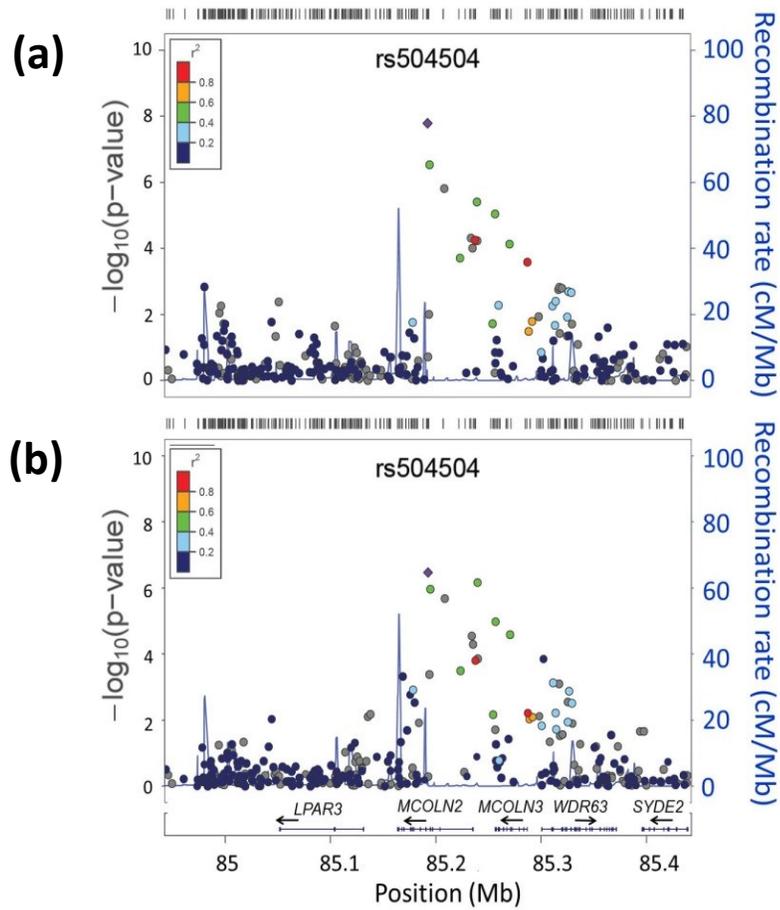


Figure 3.8 MAGWAS vs. PLINK for an example chemical/region. LocusZoom plots of the most significant SNP (rs504504, chr1p22) associated with cytotoxicity of dieldrin from (a) MAGWAS (a) and (b) PLINK regression analyses, using EC10 as a quantitative phenotype.

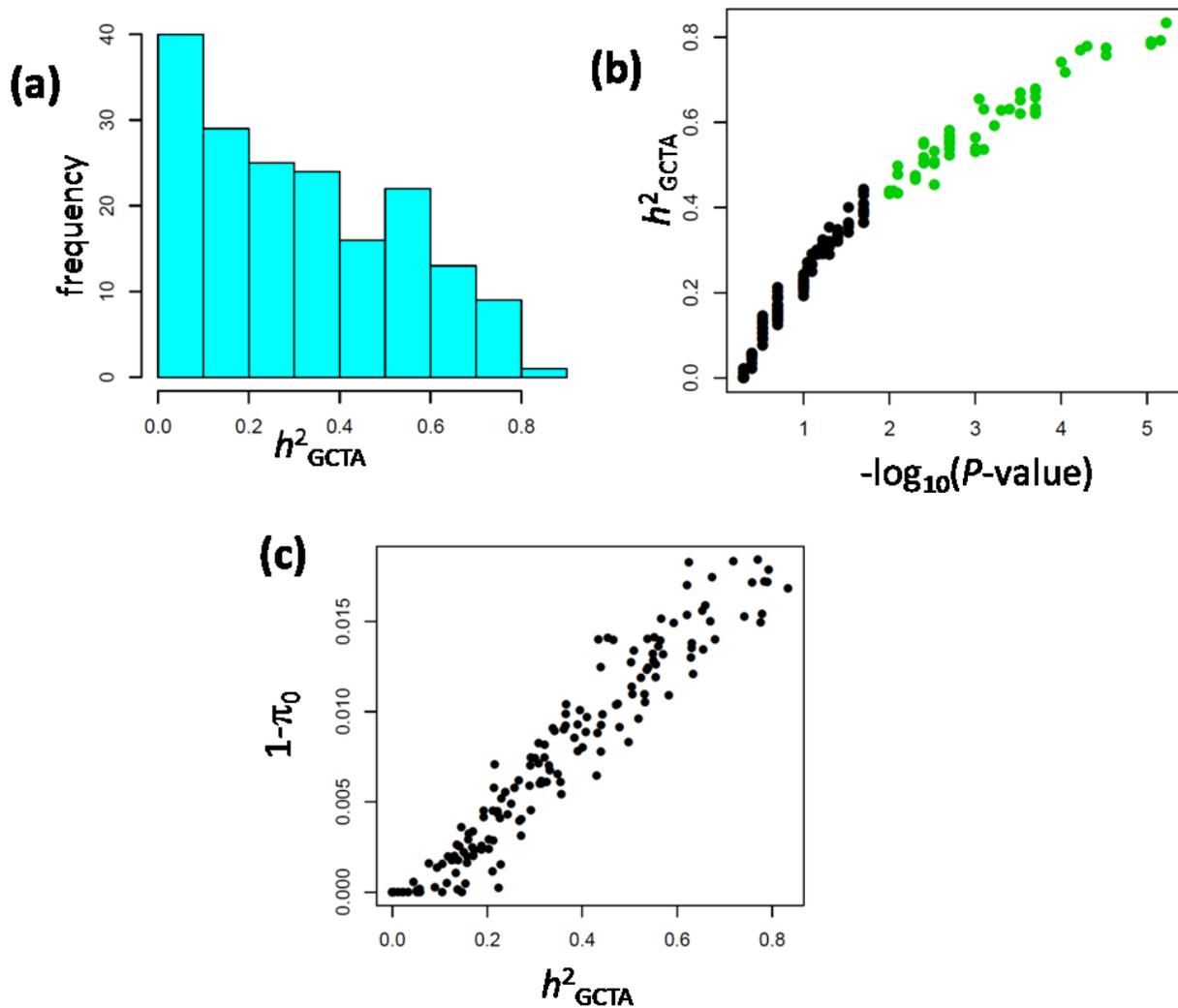


Figure 3.9. Heritability of EC₁₀ (a) Histogram of GCTA heritability (h^2) estimates for EC₁₀. (b) Heritability estimates for each chemical vs. GCTA $-\log_{10}(P\text{-values})$. For GCTA heritability, 34 chemicals with false discovery rate $q < 0.05$ are shown in green. (c) The estimated proportion of true discoveries (among 1.3M SNPs) vs. GCTA heritability estimates for EC₁₀ ($r=0.96$).

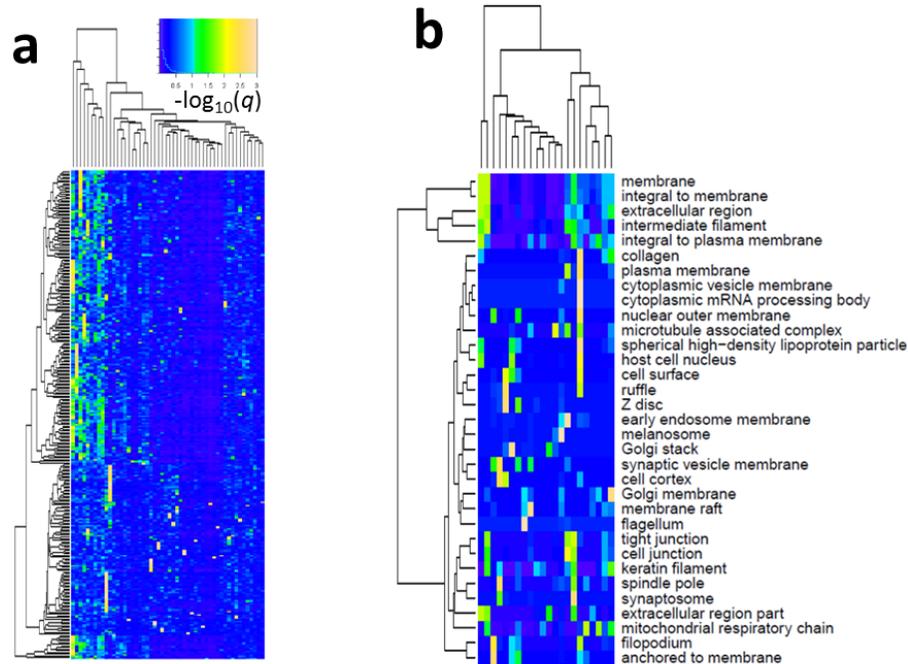


Figure 3.10 Heatmaps of Gene- and Pathway-Chemical Associations (a) Heatmap of $-\log_{10}(q)$ for 260 genes and 52 chemicals, in which $q < 0.1$ for at least one gene-chemical combination. (b) Heatmap of $-\log_{10}(q)$ for 33 Gene Ontology Cellular Component pathways and 22 chemicals, in which $q < 0.1$ for at least one pathway-chemical combinations.

Table 3.1. Average EC10 Cytotoxicity Values and Fold-Ranges

Drug Name	CAS#	Mean ^a	Median ^b	SD	q ₀₅ ^c	q ₉₅	Fold Range q ₉₅ - q ₀₅	Shrunk en Fold Range Estimate ^d
Daunomycin HCL	23541-50-6	-2.467	-2.562	0.317	-3.116	-1.344	59.198	7.422
Colchicine	64-86-8	-2.232	-2.215	0.214	-2.699	-1.531	14.726	1.710
Colchicine	64-86-8	-2.228	-2.199	0.212	-2.651	-1.668	9.603	2.679
Malachite green oxalate	2437-29-8	-1.492	-1.473	0.242	-2.114	-0.795	20.842	3.718
Tetramethylthiouoram disulfide	137-26-8	-1.347	-1.362	0.403	-1.991	-0.635	22.655	7.521
Cycloheximide	66-81-9	-1.357	-1.357	0.290	-1.909	-0.870	10.944	2.565
Digoxin	20830-75-5	-1.303	-1.302	0.222	-1.839	-0.691	14.052	3.229
Ziram	137-30-4	-1.264	-1.296	0.557	-2.217	0.081	199.014	50.632
Cycloheximide	66-81-9	-1.208	-1.204	0.294	-1.853	-0.606	17.683	5.106
Phenylmercuric acetate	62-38-4	-1.183	-1.195	0.345	-1.846	-0.517	21.363	8.569
Zinc pyrithione	13463-41-7	-1.100	-1.140	0.463	-1.815	0.041	71.802	22.069
Mercuric chloride	7487-94-7	-0.946	-1.024	0.678	-2.053	0.181	171.242	42.958
6-Thioguanine (6-TG)	154-42-7	-0.803	-0.863	0.786	-2.233	0.542	595.493	91.906
Hexamethyl-p-rosaniline chloride	548-62-9	-0.774	-0.746	0.262	-1.232	-0.287	8.809	3.668
Methyl mercuric (II) chloride	115-09-3	-0.554	-0.544	0.373	-1.398	0.140	34.552	13.299
6-Mercaptopurine monohydrate	6112-76-1	-0.348	-0.524	0.973	-1.674	1.701	2368.405	201.289
5-Fluorouracil	51-21-8	-0.152	-0.188	0.613	-1.141	0.949	123.005	25.675
Azathioprine	446-86-6	-0.021	-0.187	0.835	-1.177	1.333	323.460	75.384
Chloranil	118-75-2	-0.116	-0.145	0.442	-0.908	0.844	56.582	8.202
tetra-N-Octylammonium bromide	14866-33-2	-0.135	-0.109	0.291	-0.636	0.537	14.911	6.379
Ethidium bromide	1239-45-8	0.019	-0.006	0.322	-0.657	0.747	25.357	10.620
p-Nitrosodiphenylamine	156-10-5	0.024	0.052	0.341	-0.500	0.597	12.493	6.943
N,N-Dimethyl-p-nitrosoaniline	138-89-6	0.023	0.062	0.330	-0.634	0.511	13.984	5.040
Cetylpyridinium bromide	140-72-7	0.080	0.072	0.332	-0.470	0.708	15.086	7.573
Sodium dichromate dihydrate (VI)	7789-12-0	0.109	0.112	0.348	-0.467	0.729	15.685	5.277
Nitrogen mustard hydrochloride	55-86-7	0.114	0.113	0.507	-0.643	0.990	42.993	16.172
Pentaerythritol triacrylate	3524-68-3	0.101	0.152	0.297	-0.461	0.538	9.977	4.039
Potassium dichromate	7778-50-9	0.157	0.168	0.315	-0.470	0.810	19.050	6.706
Pyrimethamine	58-14-0	0.189	0.173	0.571	-0.765	1.249	103.396	11.847
Sodium dichromate dihydrate (VI)	7789-12-0	0.181	0.197	0.369	-0.590	0.851	27.615	9.685
2-Octyl-3-isothiazolone	26530-20-1	0.200	0.231	0.291	-0.418	0.673	12.335	4.571
Chlorambucil	305-03-3	0.267	0.277	0.276	-0.213	0.763	9.465	3.480
Diglycidyl resorcinol ether (DGRE)	101-90-6	0.304	0.308	0.362	-0.381	0.966	22.269	7.881

9-Aminoacridine, monohydrochloride, monohydrate	52417-22-8	0.326	0.330	0.249	-0.166	0.735	7.972	4.079
1,3,5-Triglycidyl isocyanurate	2451-62-9	0.319	0.337	0.329	-0.360	1.044	25.305	8.179
Fumaronitrile	764-42-1	0.413	0.371	0.330	-0.063	0.972	10.847	3.021
o-Phenanthroline	66-71-7	0.369	0.372	0.198	0.026	0.752	5.328	2.200
N,N'-Di-sec-butyl-p-phenyldiamine	101-96-2	0.384	0.394	0.207	0.003	0.734	5.384	3.496
Hexachlorophene	70-30-4	0.387	0.399	0.295	-0.275	1.046	20.979	6.739
Chlorhexidine	55-56-1	0.389	0.425	0.353	-0.549	1.085	43.071	9.064
Tetraethylene glycol diacrylate	17831-71-9	0.438	0.447	0.275	-0.039	0.878	8.259	2.915
Methylene bis(thiocyanate)	6317-18-6	0.487	0.472	0.256	0.115	0.942	6.714	3.793
Catechol	120-80-9	0.477	0.486	0.304	-0.101	0.978	11.997	5.334
beta-Nitrostyrene	5153-67-3	0.539	0.509	0.270	0.137	1.037	7.954	3.195
Benzethonium chloride	121-54-0	0.554	0.542	0.168	0.276	0.877	3.990	2.138
N-Methyl-p-aminophenol sulfate	55-55-0	0.584	0.593	0.258	0.152	1.032	7.600	4.078
Ergotamine tartrate	379-79-3	0.621	0.623	0.361	-0.097	1.214	20.464	3.631
Domiphen bromide	538-71-6	0.614	0.633	0.239	0.229	0.967	5.466	2.947
1,1,1,2-Tetrabromoethane	630-16-0	0.612	0.647	0.266	-0.022	1.027	11.189	5.288
Nitazoxanide	55981-09-4	0.700	0.705	0.276	0.152	1.209	11.412	4.428
2,2',4'-Trichloroacetophenone	4252-78-2	0.705	0.706	0.236	0.245	1.157	8.167	1.456
2,4-Difluoronitrobenzene	446-35-5	0.700	0.715	0.680	-0.514	1.915	268.741	20.770
Iodochlorohydroxyquinoline	130-26-7	0.739	0.755	0.253	0.134	1.178	11.082	3.679
Diethylene glycol diacrylate	4074-88-8	0.735	0.770	0.278	0.091	1.195	12.700	6.164
2-Amino-4-chlorophenol	95-85-2	0.749	0.776	0.526	-0.357	1.579	86.417	9.579
Turmeric (>98% curcumin)	458-37-7	0.790	0.805	0.208	0.370	1.163	6.198	2.313
1-Methyl-3-nitro-1-nitroso-guanidine	70-25-7	0.857	0.826	0.320	0.174	1.396	16.668	7.219
Tetraethylene glycol diacrylate	17831-71-9	0.825	0.829	0.209	0.345	1.292	8.861	3.413
Dibromonitromethane (water disinfection byproducts)	598-91-4	0.865	0.869	0.235	0.319	1.290	9.351	4.609
1,8-Dihydroxy-4,5-dinitroanthraquinone	81-55-0	0.834	0.871	0.221	0.261	1.209	8.863	4.152
Dexamethazone	50-02-2	0.792	0.898	0.966	-1.049	2.000	1119.682	43.483
1,6-Hexamethylene diacrylate	13048-33-4	0.906	0.928	0.184	0.615	1.108	3.111	1.382
1,3-Diiminobenz (f)-isoindoline	65558-69-2	0.931	0.939	0.180	0.318	1.266	8.873	2.115
4-Amino-4'-hydroxy-3-methyl-diphenylamine	6219-89-2	0.927	0.942	0.207	0.449	1.393	8.779	3.345
1,3-Dicyclohexylcarbodiimide	538-75-0	0.962	0.947	0.257	0.517	1.453	8.632	4.137
Guggulsterones E	39025-24-6	0.942	0.966	0.301	0.386	1.383	9.913	3.755
2,3,4,5-Tetrachloronitrobenzene	879-39-0	0.956	0.980	0.322	0.355	1.432	11.957	2.319
Dieldrin	60-57-1	0.978	0.982	0.325	0.276	1.476	15.869	6.769
3-Chloro-4-(dichloromethyl)-5-hydroxy-2(5H)-furanone(MX)	77439-76-0	0.959	0.990	0.212	0.443	1.258	6.542	2.034
17beta-Estradiol	50-28-2	0.947	0.994	0.326	0.384	1.370	9.672	1.422
Aldrin	309-00-2	0.977	0.997	0.402	0.262	1.573	20.474	4.223

t-Butylhydroquinone	1948-33-0	0.943	0.998	0.435	-0.057	1.479	34.406	4.193
Acetochlor	34256-82-1	1.002	1.007	0.206	0.702	1.309	4.048	2.302
4,4-Thiobis(6-tert-butyl-m-cresol)	96-69-5	1.002	1.012	0.143	0.681	1.251	3.712	2.085
Hydroquinone	123-31-9	0.988	1.013	0.362	0.369	1.535	14.663	7.213
2-Amino-4-methylphenol	95-84-1	0.973	1.016	0.291	0.463	1.325	7.283	3.983
N-Isopropyl-N'-phenyl-p-phenylenediamine	101-72-4	0.994	1.018	0.196	0.514	1.268	5.677	1.661
p-Quinone	106-51-4	0.996	1.018	0.231	0.527	1.348	6.627	1.580
Triamterene	396-01-0	1.035	1.020	0.370	0.367	1.727	22.859	10.261
Amiloride hydrochloride	2016-88-8	1.008	1.043	0.392	0.256	1.591	21.661	1.968
Tamoxifen citrate	54965-24-1	1.060	1.059	0.159	0.647	1.327	4.787	1.243
N,N,N',N'-Tetramethyl-p-phenylenediamine	100-22-1	1.048	1.066	0.387	0.418	1.635	16.503	2.355
Aflatoxin B1 from Aspergillus flavus	1162-65-8	1.033	1.068	0.415	0.357	1.649	19.589	5.202
t-Butylhydroquinone	1948-33-0	1.053	1.069	0.248	0.646	1.428	6.049	3.112
Hematoxylin	517-28-2	1.055	1.073	0.249	0.514	1.412	7.909	1.575
Retinal	116-31-4	1.073	1.086	0.189	0.665	1.353	4.874	2.904
p-tert-Butylcatechol	98-29-3	1.086	1.089	0.198	0.753	1.481	5.353	3.450
Reserpine	50-55-5	1.085	1.090	0.163	0.783	1.319	3.438	1.375
o-Nitrobenzyl chloride	612-23-7	1.056	1.093	0.512	0.035	1.924	77.577	6.760
Captan	133-06-2	1.087	1.100	0.194	0.633	1.399	5.838	2.406
Melatonin	73-31-4	1.108	1.116	0.330	0.551	1.600	11.199	6.378
Chlordecone (kepone)	143-50-0	1.119	1.129	0.124	0.916	1.297	2.408	1.179
4-Chloro-o-phenylenediamine	95-83-0	1.084	1.130	0.496	0.238	1.875	43.388	6.949
Saquinavir mesylate (AIDS Initiative)	149845-06-7	1.128	1.138	0.167	0.804	1.379	3.764	1.304
Captan	133-06-2	1.114	1.141	0.218	0.640	1.389	5.613	1.892
Progesterone	57-83-0	1.145	1.151	0.197	0.866	1.431	3.678	1.838
7,12-Dimethylbenzanthracene	57-97-6	1.123	1.157	0.381	0.514	1.669	14.291	1.420
8-Hydroxyquinoline	148-24-3	1.142	1.172	0.325	0.465	1.684	16.551	7.506
N-(1,3-Dimethylbutyl)-N'-phenyl-p-phenylenediamine	793-24-8	1.169	1.173	0.129	0.989	1.350	2.300	1.481
HC blue 2	33229-34-4	1.137	1.174	0.292	0.584	1.552	9.285	1.833
Toxaphene	8001-35-2	1.187	1.176	0.129	0.980	1.412	2.704	1.798
Titanocene dichloride	1271-19-8	1.157	1.176	0.262	0.723	1.569	7.019	2.756
Chlordane (technical grade)	12789-03-6	1.149	1.188	0.353	0.566	1.547	9.576	2.009
Progesterone	57-83-0	1.200	1.200	0.171	0.900	1.466	3.679	1.807
p-Aminophenol	123-30-8	1.208	1.201	0.192	0.976	1.503	3.364	1.332
Dazomet	533-74-4	1.200	1.202	0.301	0.650	1.789	13.792	7.957
dichloroacetonitrile	3018-12-0	1.191	1.203	0.291	0.750	1.598	7.034	2.769
Amitriptyline HCl	549-18-8	1.200	1.205	0.189	0.883	1.492	4.058	1.545
Ethacrynic acid	58-54-8	1.209	1.206	0.216	0.843	1.554	5.140	1.973

Dibromoacetonitrile	3252-43-5	1.219	1.212	0.221	0.849	1.639	6.173	2.679
2,2'-Thiobis(4,6-dichlorophenol)	97-18-7	1.207	1.213	0.243	0.487	1.670	15.248	3.791
N-(1-Naphthyl)ethylenediamine dihydrochloride	1465-25-4	1.231	1.222	0.110	1.066	1.420	2.260	1.595
Endosulfan	115-29-7	1.212	1.224	0.237	0.840	1.548	5.113	2.521
N,N'-Diphenyl-p-phenylenediamine	74-31-7	1.205	1.224	0.354	0.609	1.754	13.991	1.556
Tetrachlorvinphos	961-11-5	1.219	1.234	0.285	0.620	1.649	10.689	1.723
2,4-Decadienal	25152-84-5	1.269	1.242	0.229	0.576	1.860	19.229	7.996
13-cis-Retinal	472-86-6	1.262	1.247	0.168	1.062	1.583	3.316	1.625
Bisphenol A diglycidyl ether	1675-54-3	1.231	1.250	0.291	0.831	1.678	7.024	2.548
Cadmium chloride	10108-64-2	1.273	1.259	0.225	0.908	1.656	5.591	3.115
N,N-Diethyl-p-phenylenediamine	93-05-0	1.278	1.260	0.201	1.019	1.720	5.022	3.012
Vitamin D3	67-97-0	1.336	1.263	0.254	1.036	1.893	7.202	4.846
2,3,4,5-Tetrachlorophenol	4901-51-3	1.254	1.265	0.186	0.977	1.528	3.552	1.361
Chlordecone (kepone)	143-50-0	1.289	1.284	0.143	0.990	1.649	4.563	2.838
Ethyl linolenate	1191-41-9	1.307	1.288	0.164	1.086	1.630	3.504	1.634
Bis(cyclopentadienyl)vanadium chloride	12083-48-6	1.282	1.296	0.288	0.647	1.770	13.261	4.582
Retinol acetate	127-47-9	1.321	1.299	0.194	1.054	1.712	4.547	1.323
p-Benzoquinone dioxime	105-11-3	1.305	1.319	0.297	0.831	1.765	8.589	2.301
p-n-Nonylphenol	104-40-5	1.291	1.321	0.346	0.712	1.830	13.136	3.046
Chlorpheniramine maleate	113-92-8	1.241	1.322	0.505	0.368	1.893	33.512	3.510
4-(Chloroacetyl)acetanilide	140-49-8	1.295	1.323	0.340	0.739	1.785	11.119	3.698
Flutamide (pubertal study)	13311-84-7	1.319	1.357	0.369	0.733	1.789	11.391	4.921
p-Nitrophenethyl alcohol	100-27-6	1.267	1.359	0.606	0.097	2.000	80.068	2.231
Glutaraldehyde	111-30-8	1.382	1.365	0.171	1.123	1.705	3.820	1.912
1-Naphthylamine	134-32-7	1.290	1.366	0.554	0.186	2.000	65.228	13.509
Oxymetholone	434-07-1	1.356	1.369	0.197	0.901	1.743	6.951	2.337
Cadmium acetate dihydrate	4/4/5743	1.391	1.381	0.245	0.832	1.827	9.868	3.891
2-Biphenylamine	90-41-5	1.321	1.409	0.589	0.234	2.000	58.342	4.916
2-Chloroacetophenone (CN)	532-27-4	1.426	1.422	0.218	1.062	1.761	5.004	2.494
2-Pivalyl-1,3-indandione	83-26-1	1.430	1.433	0.259	1.019	1.845	6.693	4.470
8-Hydroxyquinoline	148-24-3	1.393	1.457	0.350	0.613	1.824	16.237	6.894
o-Phenylenediamine	95-54-5	1.451	1.468	0.369	0.779	1.994	16.406	4.867
2,4-Hexadienal	142-83-6	1.481	1.471	0.209	1.166	1.850	4.833	2.308
m-Nitrobenzyl chloride	619-23-8	1.416	1.478	0.427	0.641	2.000	22.878	3.445
3,4-Diaminotoluene	496-72-0	1.457	1.480	0.370	0.768	2.000	17.068	5.441
3,4-Dinitrotoluene	610-39-9	1.466	1.487	0.344	0.922	2.000	11.962	3.184
2,3,5-Trichlorophenol	933-78-8	1.534	1.538	0.323	1.024	2.000	9.472	1.817
2',4',5'-Trihydroxybutyrophenone	1421-63-2	1.502	1.552	0.337	0.916	1.961	11.117	4.527
Ninhydrin	485-47-2	1.549	1.570	0.234	1.122	2.000	7.544	4.182
cis-Dichlorodiamine platinum	15663-27-1	1.561	1.620	0.351	0.887	2.000	12.978	3.574

Propiconazole	60207-90-1	1.648	1.674	0.215	1.254	1.923	4.662	1.245
Rhein (1,8-dihydroxy-3-carboxyl anthraquinone)	478-43-3	1.637	1.690	0.318	1.098	2.000	7.980	4.135
o-Aminophenol	95-55-6	1.633	1.703	0.346	0.978	2.000	10.513	2.559
Nifedipine	21829-25-4	1.679	1.719	0.218	1.153	2.000	7.026	2.466
Di(2-ethylhexyl) phthalate	117-81-7	1.684	1.723	0.286	0.735	2.000	18.405	6.967
Alizarin Yellow R, free acid	2243-76-7	1.642	1.723	0.346	0.969	2.000	10.730	1.874
Verapamil HCl	152-11-4	1.662	1.727	0.283	1.142	2.000	7.219	1.468
Sulfathiazole	72-14-0	1.601	1.813	0.489	0.564	2.000	27.262	5.784
Phenformin hydrochloride	834-28-6	1.765	1.820	0.224	1.241	2.000	5.745	3.414
Dichlorvos (Vapona)	62-73-7	1.810	1.862	0.203	1.422	2.000	3.785	2.203
Danthron	117-10-2	1.761	1.865	0.259	1.308	2.000	4.916	2.561
dimethyldipropylene-triamine	10563-29-8	1.810	1.895	0.244	1.323	2.000	4.752	1.392
Systhane	88671-89-0	1.870	1.942	0.221	1.506	2.000	3.118	1.907
Azobenzene	103-33-3	1.827	1.943	0.263	1.351	2.000	4.459	1.554
1-(2,6,6-Trimethyl-2-cyclohexene-1-yl)-1-penten-3-one	7779-30-8	1.913	1.954	0.141	1.585	2.000	2.603	1.669
Diisobutyl phthalate	84-69-5	1.937	1.965	0.167	1.573	2.000	2.671	2.303
4-Methoxy-3-nitro-N-phenylbenzamide	97-32-5	1.957	1.972	0.124	1.737	2.000	1.832	2.041
Ethoxyquin	91-53-2	1.871	1.972	0.281	1.456	2.000	3.497	1.823
t-Butyl formate	762-75-4	1.930	1.989	0.217	1.576	2.000	2.652	3.165
Aldicarb	116-06-3	1.976	1.993	0.123	1.894	2.000	1.277	1.126
permethrin	52645-53-1	1.974	1.995	0.088	1.864	2.000	1.367	1.176
4-Chloro-3,5-dinitro-a,a,a-trifluorotoluene	393-75-9	1.957	1.998	0.196	1.744	2.000	1.803	1.259
3,4-Dichlorophenyl isocyanate	102-36-3	1.978	1.998	0.108	1.855	2.000	1.395	1.172
Styrene	100-42-5	1.993	1.999	0.053	2.000	2.000	1.000	1.000
5-(Hydroxymethyl)-2-furoic acid	6338-41-6	1.996	2.000	0.044	2.000	2.000	1.000	1.000
trans-1,4-dichloro-2-butene	110-57-6	1.965	2.001	0.143	1.795	2.000	1.602	1.168
Hexachloro-1,3-butadiene	87-68-3	1.990	2.001	0.079	1.986	2.000	1.034	1.016
Mono(2-ethylhexyl)phthalate	4376-20-9	1.986	2.001	0.079	1.947	2.000	1.131	1.163
Methacrylonitrile	126-98-7	1.981	2.002	0.146	1.990	2.000	1.024	1.009
1,2-Epoxy-3-chloropropane (Epichlorohydrin)	106-89-8	1.985	2.003	0.095	1.995	2.000	1.012	1.018

^aValues are \log_{10} (molar concentration). ^bTable entries are sorted by the median. ^c5th percentile. ^dSee Online Methods for the shrinkage procedure, which estimates the fold-range after removing the effect of technical sampling variation.

Table 3.2. Chemicals showing significant EC10 variation across populations or by sex

Drug Name	CAS#	P population differences	q population differences	P sex differences	q sex differences
Azathioprine	446-86-6	3.8E-13	6.8E-11	0.228	0.506
5-Fluorouracil	51-21-8	9.3E-13	8.3E-11	0.003	0.078
1,3,5-Triglycidyl isocyanurate	2451-62-9	7.4E-10	4.4E-08	0.223	0.504
Diglycidyl resorcinol ether (DGRE)	101-90-6	1.6E-09	7.0E-08	0.042	0.217
6-Thioguanine (6-TG)	154-42-7	2.4E-09	8.4E-08	0.804	0.882
6-Mercaptopurine monohydrate	6112-76-1	8.1E-09	2.4E-07	0.898	0.918
Turmeric (>98% curcumin)	458-37-7	2.1E-08	5.5E-07	0.031	0.196
Phenformin hydrochloride	834-28-6	8.1E-08	1.8E-06	0.289	0.556
Tetrachlorvinphos	961-11-5	3.4E-06	6.7E-05	0.517	0.735
Cycloheximide	66-81-9	6.0E-06	1.1E-04	0.778	0.882
N,N-Dimethyl-p-nitrosoaniline	138-89-6	8.9E-06	1.4E-04	0.018	0.162
Hematoxylin	517-28-2	1.1E-05	1.6E-04	0.643	0.822
Diethylene glycol diacrylate	4074-88-8	1.2E-05	1.6E-04	0.012	0.131
Cadmium acetate dihydrate	4-4-5743	1.4E-05	1.8E-04	0.757	0.882
Sodium dichromate dihydrate (VI)	7789-12-0	1.6E-05	1.9E-04	0.050	0.218
beta-Nitrostyrene	5153-67-3	1.8E-05	2.0E-04	0.229	0.506
N-Methyl-p-aminophenol sulfate	55-55-0	2.1E-05	2.2E-04	0.051	0.219
Tetraethylene glycol diacrylate	17831-71-9	3.8E-05	3.8E-04	0.023	0.180
Tetraethylene glycol diacrylate	17831-71-9	4.6E-05	4.4E-04	0.050	0.218
1,1,1,2-Tetrabromoethane	630-16-0	5.8E-05	0.001	0.368	0.609
Pyrimethamine	58-14-0	7.1E-05	0.001	0.650	0.826
2,2',4'-Trichloroacetophenone	4252-78-2	6.9E-05	0.001	0.746	0.879
Malachite green oxalate	2437-29-8	6.7E-05	0.001	0.567	0.774

Retinal	116-31-4	7.6E-05	0.001	0.880	0.906
t-Butylhydroquinone	1948-33-0	1.2E-04	0.001	0.070	0.249
Methyl mercuric (II) chloride	115-09-3	1.2E-04	0.001	0.067	0.247
Dieldrin	60-57-1	1.2E-04	0.001	0.007	0.121
Sodium dichromate dihydrate (VI)	7789-12-0	1.1E-04	0.001	0.027	0.196
2-Amino-4-chlorophenol	95-85-2	1.7E-04	0.001	0.808	0.882
Nitrogen mustard hydrochloride	55-86-7	2.0E-04	0.001	0.077	0.251
Verapamil HCl	152-11-4	2.4E-04	0.001	0.425	0.661
Triamterene	396-01-0	3.6E-04	0.002	0.486	0.722
N-Isopropyl-N'-phenyl-p-phenylenediamine	101-72-4	3.6E-04	0.002	0.359	0.601
Cycloheximide	66-81-9	3.8E-04	0.002	0.880	0.906
Alizarin Yellow R, free acid	2243-76-7	4.0E-04	0.002	0.061	0.239
Benzethonium chloride	121-54-0	0.001	0.003	0.076	0.251
9-Aminoacridine, monohydrochloride, monohydrate	52417-22-8	0.001	0.003	0.604	0.795
Dexamethazone	50-02-2	0.001	0.003	0.571	0.774
3,4-Dinitrotoluene	610-39-9	0.001	0.003	0.008	0.121
Retinol acetate	127-47-9	0.001	0.004	0.094	0.295
p-Nitrosodiphenylamine	156-10-5	0.001	0.004	0.022	0.180
Glutaraldehyde	111-30-8	0.001	0.004	0.131	0.392
trans-1,4-dichloro-2-butene	110-57-6	0.001	0.004	0.177	0.441
Hexachlorophene	70-30-4	0.001	0.005	0.856	0.902
Acetochlor	34256-82-1	0.001	0.005	0.257	0.548
Titanocene dichloride	1271-19-8	0.001	0.005	0.849	0.902
Chlordecone (kepone)	143-50-0	0.001	0.005	0.046	0.218
Nifedipine	21829-25-4	0.002	0.006	0.000	0.022
Saquinavir mesylate (AIDS)	149845-	0.002	0.006	0.141	0.395

Initiative)	06-7				
Chlordecone (kepone)	143-50-0	0.002	0.006	0.010	0.128
dimethyldipropylene-triamine	10563-29-8	0.002	0.006	0.308	0.562
Zinc pyrrithione	13463-41-7	0.002	0.007	0.275	0.556
1-Methyl-3-nitro-1-nitroso-guanidine	70-25-7	0.003	0.009	0.015	0.150
Cetylpyridinium bromide	140-72-7	0.003	0.011	0.010	0.128
8-Hydroxyquinoline	148-24-3	0.004	0.012	0.193	0.474
Potassium dichromate	7778-50-9	0.004	0.013	0.067	0.247
2,4-Difluoronitrobenzene	446-35-5	0.004	0.013	0.163	0.418
p-tert-Butylcatechol	98-29-3	0.004	0.013	0.618	0.802
t-Butylhydroquinone	1948-33-0	0.004	0.013	0.702	0.861
p-n-Nonylphenol	104-40-5	0.005	0.015	0.143	0.395
Amiloride hydrochloride	2016-88-8	0.006	0.016	0.011	0.128
17beta-Estradiol	50-28-2	0.006	0.016	0.154	0.399
Ethyl linolenate	1191-41-9	0.006	0.016	0.045	0.218
Pentaerythritol triacrylate	3524-68-3	0.008	0.023	0.001	0.050
Di(2-ethylhexyl) phthalate	117-81-7	0.008	0.023	0.216	0.499
Bis(cyclopentadienyl)vanadium chloride	12083-48-6	0.008	0.023	0.720	0.871
Dichlorvos (Vapona)	62-73-7	0.009	0.025	0.282	0.556
1,6-Hexamethylene diacrylate	13048-33-4	0.010	0.025	0.002	0.073
N,N-Diethyl-p-phenylenediamine	93-05-0	0.010	0.027	0.303	0.562
8-Hydroxyquinoline	148-24-3	0.011	0.027	0.280	0.556
Ethidium bromide	1239-45-8	0.011	0.027	0.787	0.882
4-Methoxy-3-nitro-N-phenylbenzamide	97-32-5	0.013	0.033	0.558	0.768
Dibromoacetone nitrile	3252-43-5	0.015	0.038	0.248	0.535
4-Amino-4'-hydroxy-3-methyl-	6219-	0.017	0.042	0.786	0.882

diphenylamine	89-2				
Hexamethyl-p-rosaniline chloride	548-62-9	0.018	0.043	0.864	0.902
4,4-Thiobis(6-tert-butyl-m-cresol)	96-69-5	0.019	0.044	0.345	0.600
2-Chloroacetophenone (CN)	532-27-4	0.020	0.046	0.071	0.249
2-Amino-4-methylphenol	95-84-1	0.021	0.047	0.594	0.794
Ninhydrin	485-47-2	0.022	0.050	0.389	0.627
o-Phenanthroline	66-71-7	0.022	0.050	0.075	0.251
Phenylmercuric acetate	62-38-4	0.169	0.237	0.000	0.037

P-values were obtained by running analyses of variance on $\log_{10}(EC_{10})$ with subpopulation or sex as a categorical variable. *q*-values were obtained after Benjamini-Hochberg false discovery rate correction per chemical.

Table 3.3. MAGWAS Multivariate Association Results.

Chemical ^a	CAS #	SNP	bp ^b	Chrom	Gene	P ^c	qvalue ^d	Explained R ² _e
2-Amino-4-methylphenol	95-84-1	rs13120371	139092719	4	<i>SLC7A11</i>	8.42E-10	0.0006	0.0723
Methyl mercuric (II) chloride	115-09-3	rs13120371	139092719	4	<i>SLC7A11</i>	8.89E-08	0.0632	0.0414
N-Methyl-p-aminophenol sulfate	55-55-0	rs13120371	139092719	4	<i>SLC7A11</i>	4.88E-08	0.0347	0.0395
N-Isopropyl-N'-phenyl-p-phenylenediamine	101-72-4	rs1159874	19916619	7	<i>TMEM196</i>	2.71E-09	0.0019	0.0264
		rs6430301	148953669	2	<i>MBD5</i>	2.84E-07	0.0674	0.0262
		rs3935192	75878841	17	<i>FLJ45079</i>	5.44E-07	0.0968	0.0281
2-Amino-4-methylphenol	95-84-1	rs57046479	99635548	9	<i>ZNF782</i>	3.25E-07	0.0769	0.0181
		rs6446632	4355380	4	<i>ZBTB49</i>	6.15E-07	0.0875	0.0340
o-Aminophenol	95-55-6	rs1800566	69745145	16	<i>NFAT5</i>	4.32E-09	0.0031	0.0554
		rs4244032	142794725	5	<i>NR3C1</i>	3.79E-07	0.0430	0.0210
		rs8073076	63454129	17	<i>AXIN2</i>	1.10E-06	0.0784	0.0193
		rs11062381	2954423	12	<i>FKBP4</i>	1.46E-06	0.0945	0.0337
Titanocene dichloride	1271-19-8	rs62009303	92805261	15	<i>SLCO3A1</i>	1.97E-08	0.0140	0.0222
		rs62189869	162922728	2	<i>LOC151171</i>	1.82E-07	0.0431	0.0197
		rs12902246	49274274	15	<i>SECISBP2L</i>	4.62E-07	0.0657	0.0311
		rs1906308	104333651	11	<i>PDGFD</i>	7.90E-07	0.0703	0.0261
13-cis-Retinal	472-86-6	rs541217	106564400	6	<i>PRDM1</i>	1.23E-08	0.0087	0.0205
		rs4532252	12397379	4	<i>RAB28</i>	3.72E-07	0.0715	0.0329
N,N-Diethyl-p-phenylenediamine	93-05-0	rs6691053	173868955	1	<i>DARS2</i>	2.82E-08	0.0200	0.0194
		rs61879371	19852683	11	<i>NAV2</i>	1.39E-07	0.0494	0.0181

	CAS #	SNP	bp ^b	Chrom	Gene	P ^c	qvalue ^d	Explained R ² _e
2,4-Decadienal	25152-84-5	rs1194596	154238383	1	<i>C1orf43</i>	3.60E-08	0.0207	0.0282
		rs4689451	6458552	4	<i>PPP2R2C</i>	9.97E-08	0.0236	0.0211
Malachite green oxalate	2437-29-8	rs3742522	24906534	14	<i>KHNYN</i>	5.53E-08	0.0062	0.0388
		rs10772306	10677140	12	<i>KLRAP1</i>	3.59E-07	0.0283	0.0169
		rs717818	141830833	4	<i>RNF150</i>	1.40E-06	0.0908	0.0180
Fumaronitrile	764-42-1	rs11048994	27530778	12	<i>ARNTL2</i>	7.08E-08	0.0504	0.0136
		rs12962668	444687	18	<i>COLEC12</i>	2.64E-07	0.0940	0.0262
Retinal	116-31-4	rs11590090	113313563	1	<i>FAM19A3</i>	9.91E-08	0.0508	0.0198
		rs34835780	3842112	1	<i>LOC100133612</i>	2.14E-07	0.0508	0.0143
Permethrin	52645-53-1	rs2408151	5912100	8	<i>MCPH1</i>	1.04E-07	0.0740	0.0211
		rs2598	47241618	20	<i>PREX1</i>	2.26E-07	0.0805	0.0197
1,3-Dicyclohexylcarbodiimide	538-75-0	rs28437300	22224506	8	<i>SLC39A14</i>	4.25E-09	0.0030	0.0245
Dieldrin	60-57-1	rs504504	85420044	1	<i>MCOLN2</i>	1.64E-08	0.0116	0.0517
Flutamide (pubertal study)	13311-84-7	rs17186961	103630028	8	<i>KLF10</i>	1.83E-08	0.0130	0.0283
Aldrin	309-00-2	rs340251	158599864	3	<i>MFSD1</i>	2.37E-08	0.0118	0.0271
2,3,4,5-Tetrachlorophenol	4901-51-3	rs7879360	88236251	X	<i>CPXCR1</i>	3.95E-08	0.0281	0.0181
Colchicine	64-86-8	rs7777880	48275852	7	<i>ABCA13</i>	4.34E-08	0.0308	0.0244
Sulfathiazole	72-14-0	rs1796415	121543011	12	<i>P2RX7</i>	4.94E-08	0.0351	0.0416
Reserpine	50-55-5	rs13143102	131264117	4	<i>C4orf33</i>	5.05E-08	0.0359	0.0388
Dichlorvos (Vapona)	62-73-7	rs1037353	83525588	11	<i>DLG2</i>	6.73E-08	0.0479	0.0165
1,2-Epoxy-3-chloropropane	106-89-8	rs3130884	72228285	X	<i>PABPC1L2B</i>	6.97E-08	0.0496	0.0182
Cycloheximide	66-81-9	rs8053118	79168698	16	<i>WWOX</i>	7.66E-08	0.0545	0.0189
Benzethonium chloride	121-54-0	rs62496173	9309398	8	<i>TNKS</i>	9.07E-08	0.0645	0.0220

Chemical ^a	CAS #	SNP	bp ^b	Chrom	Gene	P ^c	qvalue ^d	Explained R ² _e
Tetrachlorvinphos	961-11-5	rs7642013	32638632	3	<i>DYNC1LI1</i>	9.38E-08	0.0667	0.0208
Mono(2-ethylhexyl)phthalate	4376-20-9	rs1204399	99886830	X	<i>TNMD</i>	9.79E-08	0.0577	0.0285
7,12-Dimethylbenzanthracene	57-97-6	rs9932935	16247471	16	<i>ABCC1</i>	9.84E-08	0.0700	0.0338
Phenylmercuric acetate	62-38-4	rs12899102	40495067	15	<i>BUB1B</i>	1.41E-07	0.0522	0.0233
o-Phenanthroline	66-71-7	rs11716740	182831688	3	<i>MCCC1</i>	1.98E-07	0.0740	0.0238

^aThe first three entries highlight that rs13120371 in SLC7A11 was observed with FDR $q < 0.10$ for three chemicals. Remaining entries are sorted first by chemical, and then P -value. ^bNCBI build 37. ^cMAGWAS P -value. ^dFDR q -value obtained per chemical, using ~700K SNPs analyzed by MAGWAS. ^ePartial R^2 attributable to variation in genotype.

Table 3.4 Significant EC10 –SNP associations among set of 1.3m SNPs

Chemical	CAS #	SNP	bp ^a	Chrom	gene	P ^b	q-value ^c
Chlordecone (kepone)	143-50-0	rs28502033	714403 92	7	CALN1	1.29 E-10	2E-04
Tetramethylthiouoram disulfide	137-26-8	rs74487456	877081 46	15	AGBL1	3.13 E-10	4E-04
2,2'-Thiobis(4,6-dichlorophenol)	97-18-7	rs78022668	220093 749	1	SLC30A1 0	3.28 E-10	4E-04
3-Chloro-4-(dichloromethyl)-5-hydroxy-2(5H)-furanone(MX)droxy-2(5H)-furanone(MX)	77439-76-0	rs11414767 1	806998 99	16	CDYL2	1.23 E-09	0.002
2,4-Hexadienal	142-83-6	rs499384	335820 90	6	BAK1	1.93 E-09	0.003
1,3-Diiminobenz (f)-isoindoline	65558-69-2	rs11813930	214294 32	10	C10orf1 13	2.13 E-09	0.003
2,3,4,5-Tetrachlorophenol	4901-51-3	rs61732507	381289 32	21	HLCS	2.35 E-09	0.003
N-Methyl-p-aminophenol sulfate	55-55-0	rs13120371	139092 719	4	SLC7A11	3.97 E-09	0.005
4-Amino-4'-hydroxy-3-methyl-diphenylamineenylamine	6219-89-2	rs75523194	863169 69	7	GRM3	6.78 E-09	0.009
1,8-Dihydroxy-4,5-dinitroanthraquinoneuionone	81-55-0	rs75198473	241641 98	10	KIAA121 7	7.92 E-09	0.011
cis-Dichlorodiamine platinum	15663-27-1	rs6461533	210829 86	7	SP8	8.25 E-09	0.011
dichloroacetonitrile	3018-12-0	rs75207133	123195 84	9	TYRP1	8.74 E-09	0.012
Amiloride hydrochloride	2016-88-8	rs74631305	667728 43	12	GRIP1	9.18 E-09	0.012
Domiphen bromide	538-71-6	rs78468118	827006 51	16	CDH13	9.24 E-09	0.009
o-Phenylenediamine	95-54-5	rs28473317	100407 305	15	ADAMTS 17	9.34 E-09	0.009
t-Butyl formate	762-75-4	rs17044941	234331 91	2	KLHL29	9.43 E-09	0.013
N,N,N',N'-Tetramethyl-p-phenylenediamineediamine	100-22-1	rs9825285	171722 475	3	TMEM2 12	9.50 E-09	0.013
Cycloheximide	66-81-9	rs75192401	107761 225	11	RAB39	1.02 E-08	0.014
Progesterone	57-83-0	rs60732724	153798 056	5	GALNT1 0	1.14 E-08	0.015
Domiphen bromide	538-71-6	rs75611658	103934 813	13	SLC10A2	1.29 E-08	0.009
o-Phenylenediamine	95-54-5	rs11301912 0	291965 46	13	SLC46A3	1.39 E-08	0.009
1,2-Epoxy-3-chloropropane (Epichlorohydrin)	106-89-8	rs74422342	128543 398	7	KCP	1.53 E-08	0.015
Alizarin Yellow R, free acid	2243-76-7	rs78618741	168250 523	2	XIRP2	1.59 E-08	0.021
17beta-Estradiol	50-28-2	rs76341199	739328 70	6	KHDC1L	1.90 E-08	0.025
Aflatoxin B1 from Aspergillus flavusavus	1162-65-8	rs73277691	354953 26	20	C20orf1 18	1.96 E-08	0.017
1,1,1,2-Tetrabromoethane	630-16-0	rs73737933	801557 5	5	MTRR	2.07 E-08	0.015
2',4',5'-Trihydroxybutyrophenone	1421-63-2	rs10213832	172302 409	5	ERGIC1	2.10 E-08	0.028
Diethylene glycol diacrylate	4074-88-8	rs75191956	516398 20	15	GLDN	2.11 E-08	0.028

8-Hydroxyquinoline	148-24-3	rs9477330	171350 67	6	RBM24	2.20 E-08	0.021
1,2-Epoxy-3-chloropropane (Epichlorohydrin)	106-89-8	rs73109015	637907 27	5	RGS7BP	2.28 E-08	0.015
Mercuric chloride	7487-94-7	rs5963392	381443 23	X	RPGR	2.30 E-08	0.031
1,3-Diiminobenz (f)-isoindoline	65558-69-2	rs11444823 5	101178 035	12	ANO4	2.43 E-08	0.016
Melatonin	73-31-4	rs6473423	841340 51	8	SNX16	2.55 E-08	0.034
2-Octyl-3-isothiazolone	26530-20-1	rs57831318	154355 658	3	GPR149	2.58 E-08	0.020
Chlordane (technical grade)	12789-03-6	rs4531541	756681 92	12	CAPS2	2.67 E-08	0.035
N-(1,3-Dimethylbutyl)-N'-phenyl-p-phenylenediamine	793-24-8	rs17173040	144146 779	7	NOBOX	2.67 E-08	0.031
2-Chloroacetophenone (CN)	532-27-4	rs7262054	566213 12	20	ANKRD6 0	2.86 E-08	0.038
Iodochlorohydroxyquinoline	130-26-7	rs2122382	623264 84	2	COMMD 1	2.88 E-08	0.019
1-Naphthylamine	134-32-7	rs11418175 7	341515 15	14	NPAS3	3.01 E-08	0.04
dichloroacetonitrile	3018-12-0	rs78571529	543504 74	2	ACYP2	3.11 E-08	0.018
Guggulsterones E	39025-24-6	rs11249990 8	127825 897	12	TMEM1 32C	3.13 E-08	0.042
Zinc pyrithione	13463-41-7	rs531928	165901 231	1	UCK2	3.14 E-08	0.042
2,2'-Thiobis(4,6-dichlorophenol)	97-18-7	rs11658045 1	231391 030	1	GNPAT	3.18 E-08	0.021
8-Hydroxyquinoline	148-24-3	rs4721944	207907 67	7	ABCB5	3.23 E-08	0.021
1,1,1,2-Tetrabromoethane	630-16-0	rs4566162	513227 20	16	SALL1	3.37 E-08	0.015
N-(1-Naphthyl)ethylenediamine dihydrochloride	1465-25-4	rs74706294	918492 17	6	BACH2	3.39 E-08	0.025
Acetochlor	34256-82-1	rs11523761 8	800406 82	2	CTNNA2	3.40 E-08	0.045
Nitazoxanide	55981-09-4	rs182482	160710 20	16	ABCC1	3.44 E-08	0.026
Colchicine	64-86-8	rs9419276	134798 249	10	C10orf9 3	3.53 E-08	0.047
Cycloheximide	66-81-9	rs78272463	955130 38	8	KIAA142 9	3.74 E-08	0.023
dichloroacetonitrile	3018-12-0	rs72894316	743786 60	3	CNTN3	3.96 E-08	0.018
Bisphenol A diglycidyl ether	1675-54-3	rs7708761	118856 707	5	HSD17B 4	4.01 E-08	0.032
N,N-Dimethyl-p-nitrosoaniline	138-89-6	rs11523761 8	800406 82	2	CTNNA2	4.02 E-08	0.041
Aflatoxin B1 from Aspergillus flavus	1162-65-8	rs78373020	986730 48	1	DPYD	4.09 E-08	0.017
Captan	133-06-2	rs9533891	450992 24	13	TSC22D1	4.11 E-08	0.055
7,12-Dimethylbenzanthracene	57-97-6	rs7051976	786578 92	X	ITM2A	4.26 E-08	0.046
Aflatoxin B1 from Aspergillus flavus	1162-65-8	rs73892641	267573 2	20	EBF4	4.31 E-08	0.017
m-Nitrobenzyl chloride	619-23-8	rs11656302 7	107358 795	4	AIMP1	4.37 E-08	0.037
1,1,1,2-Tetrabromoethane	630-16-0	rs80050627	165228 913	1	LMX1A	4.48 E-08	0.015
1,2-Epoxy-3-chloropropane (Epichlorohydrin)	106-89-8	rs10175753	230503 892	2	DNER	4.53 E-08	0.020
t-Butylhydroquinone	1948-33-0	rs76069668	101975 867	2	CREG2	4.78 E-08	0.063

17beta-Estradiol	50-28-2	rs79788122	907598 26	13	GPC5	4.78 E-08	0.028
N-(1-Naphthyl)ethylenediamine dihydrochloride	1465-25-4	rs3736146	334576 97	17	NLE1	4.81 E-08	0.025
Bisphenol A diglycidyl ether	1675-54-3	rs7699080	728858 99	4	NPFFR2	4.83 E-08	0.032
Toxaphene	8001-35-2	rs60756373	249105 34	22	UPB1	4.84 E-08	0.036
Aflatoxin B1 from Aspergillus flavus	1162-65-8	rs19089932	244197 75	21	NCAM2	5.04 E-08	0.017
4-Amino-4'-hydroxy-3-methyl-diphenylamine	6219-89-2	rs78933262	867603 40	16	FOXL1	5.06 E-08	0.034
Cycloheximide	66-81-9	rs11221963	894582 22	14	EML5	5.06 E-08	0.023
Hexachloro-1,3-butadiene	87-68-3	rs72774515	294911	2	ACP1	5.22 E-08	0.069
N,N,N',N'-Tetramethyl-p-phenylenediamine	100-22-1	rs74706317	150228 076	2	LYPD6	5.32 E-08	0.035
Systhane	88671-89-0	rs74798209	770383 27	15	SCAPER	5.36 E-08	0.036
2,3,4,5-Tetrachlorophenol	4901-51-3	rs1904540	670454 72	12	GRIP1	5.39 E-08	0.036
1,8-Dihydroxy-4,5-dinitroanthraquinone	81-55-0	rs58733988	374005 97	4	KIAA123 9	5.60 E-08	0.025
Cadmium chloride	10108-64-2	rs4751095	131322 781	10	MGMT	5.61 E-08	0.074
m-Nitrobenzyl chloride	619-23-8	rs2570981	121765 728	5	SNCAIP	5.63 E-08	0.037
N-(1-Naphthyl)ethylenediamine dihydrochloride	1465-25-4	rs55681361	472988 54	21	PCBP3	5.71 E-08	0.025
Cycloheximide	66-81-9	rs5974707	136862 247	X	GPR101	5.72 E-08	0.038
Chlordecone (kepone)	143-50-0	rs34098910	497209 32	12	TROAP	5.97 E-08	0.040
p-Nitrosodiphenylamine	156-10-5	rs236356	368011 88	6	CPNE5	6.00 E-08	0.059
Domiphen bromide	538-71-6	rs8057753	561752 42	16	LOC2838 56	6.14 E-08	0.027
Titanocene dichloride	1271-19-8	rs10521479	555833 83	X	FOXR2	6.16 E-08	0.082
N,N-Dimethyl-p-nitrosoaniline	138-89-6	rs11305997	194701 443	3	C3orf21	6.23 E-08	0.041
Saquinavir mesylate (AIDS Initiative)	149845-06-7	rs80233769	917429 37	15	SV2B	6.24 E-08	0.083
Nitazoxanide	55981-09-4	rs7522764	147001 029	1	BCL9	6.41 E-08	0.028
4-Chloro-o-phenylenediamine	95-83-0	rs11943865	148650 93	4	CC2D2A	6.41 E-08	0.059
3-Chloro-4-(dichloromethyl)-5-hydroxy-2(5H)-furanone(MX)droxy-2(5H)-furanone(MX)	77439-76-0	rs11338951	143441 17	16	MKL2	6.43 E-08	0.020
17beta-Estradiol	50-28-2	rs78468118	827006 51	16	CDH13	6.54 E-08	0.028
1-Naphthylamine	134-32-7	rs11618649	668004 64	3	LRIG1	6.55 E-08	0.044
Acetochlor	34256-82-1	rs16839696	138609 221	2	HNMT	6.72 E-08	0.045
Dazomet	533-74-4	rs11481148	140314 947	2	LRP1B	6.79 E-08	0.085
N,N'-Diphenyl-p-phenylenediamine	74-31-7	rs9523298	919535 23	13	MIR17H G	6.79 E-08	0.045
dichloroacetonitrile	3018-12-0	rs74598282	871158 33	9	KLF4	6.97 E-08	0.021
7,12-Dimethylbenzanthracene	57-97-6	rs6681688	818812 12	1	LPHN2	7.00 E-08	0.046
Ziram	137-30-4	rs6696727	214422 475	1	SMYD2	7.16 E-08	0.095

Styrene	100-42-5	rs73154933	694486 2	2	CMPK2	7.33 E-08	0.097
Aflatoxin B1 from Aspergillus flavus	1162-65-8	rs57916829	672786 17	18	DOK6	7.44 E-08	0.020
permethrin	52645-53-1	rs12547834	103661 811	8	KLF10	7.66 E-08	0.067
N-(1,3-Dimethylbutyl)-N'-phenyl-p-phenylenediamine	793-24-8	rs11891305	379473 8	2	ALLC	7.75 E-08	0.031
dichloroacetonitrile	3018-12-0	rs11620259	118722 702	8	EXT1	8.06 E-08	0.021
1,3-Diiminobenz (f)-isoindoline	65558-69-2	rs11579730	715790 45	5	MRPS27	8.89 E-08	0.034
3-Chloro-4-(dichloromethyl)-5-hydroxy-2(5H)-furanone(MX)	77439-76-0	rs11307345	883482 09	3	C3orf38	9.27 E-08	0.021
Amiloride hydrochloride	2016-88-8	rs9805629	110688 370	13	COL4A1	9.30 E-08	0.062
Captan	133-06-2	rs7144942	306995 18	14	PRKD1	9.43 E-08	0.042
2,3,4,5-Tetrachlorophenol	4901-51-3	rs28483005	888327 41	7	ZNF804B	9.81 E-08	0.043
2-Amino-4-chlorophenol	95-85-2	rs7871969	667346 5	9	GLDC	9.95 E-08	0.057
Systhane	88671-89-0	rs75774128	535059 43	13	PCDH8	9.99 E-08	0.044
17beta-Estradiol	50-28-2	rs36022476	766737 85	15	SCAPER	1.03 E-07	0.028
1,3-Diiminobenz (f)-isoindoline	65558-69-2	rs75108342	186491 772	4	SORBS2	1.03 E-07	0.034
Triamterene	396-01-0	rs75524003	811778	9	DMRT1	1.04 E-07	0.075
p-Nitrosodiphenylamine	156-10-5	rs11278084	991209 81	7	ZKSCAN 5	1.06 E-07	0.059
2,4-Hexadienal	142-83-6	rs76209455	838786	17	NXN	1.07 E-07	0.071
Toxaphene	8001-35-2	rs11216844	233419 69	6	HDGFL1	1.07 E-07	0.036
Colchicine	64-86-8	rs60118770	914432 59	8	TMEM6 4	1.11 E-07	0.074
Nitazoxanide	55981-09-4	rs10187279	228612 202	2	SLC19A3	1.12 E-07	0.037
3-Chloro-4-(dichloromethyl)-5-hydroxy-2(5H)-furanone(MX)	77439-76-0	rs11216844	233419 69	6	HDGFL1	1.15 E-07	0.022
Captan	133-06-2	rs1793170	118150 205	11	MPZL3	1.17 E-07	0.042
2-Amino-4-chlorophenol	95-85-2	rs75408210	676088 38	2	ETAA1	1.18 E-07	0.057
Chlordecone (kepone)	143-50-0	rs9933397	553966 53	16	IRX6	1.20 E-07	0.047
4-Amino-4'-hydroxy-3-methyl-diphenylamine	6219-89-2	rs76662197	142805 555	4	IL15	1.20 E-07	0.053
Captan	133-06-2	rs35333194	123412 290	6	CLVS2	1.27 E-07	0.042
2-Amino-4-chlorophenol	95-85-2	rs77766744	400234 74	12	C12orf4 0	1.31 E-07	0.057
17beta-Estradiol	50-28-2	rs14698611	968950 44	X	DIAPH2	1.33 E-07	0.028
p-Nitrosodiphenylamine	156-10-5	rs11462131	216531 001	2	LOC6463 24	1.33 E-07	0.059
Triamterene	396-01-0	rs74000350	239858 614	2	HDAC4	1.41 E-07	0.075
17beta-Estradiol	50-28-2	rs77203508	174412 976	2	ZAK	1.42 E-07	0.028
1,2-Epoxy-3-chloropropane (Epichlorohydrin)	106-89-8	rs77143142	104926 726	12	CHST11	1.43 E-07	0.047
N-(1,3-Dimethylbutyl)-N'-phenyl-p-phenylenediamine	793-24-8	rs11487255	716542 41	17	SDK2	1.45 E-07	0.031

2,3,4,5-Tetrachlorophenol	4901-51-3	rs74444408	27796616	14	NOVA1	1.51 E-07	0.050
o-Aminophenol	95-55-6	rs114765730	156894511	6	ARID1B	1.53 E-07	0.088
N-(1,3-Dimethylbutyl)-N'-phenyl-p-phenylenediamine	793-24-8	rs74831443	54356852	2	ACYP2	1.53 E-07	0.031
Hexachloro-1,3-butadiene	87-68-3	rs72661424	103823726	13	SLC10A2	1.61 E-07	0.071
dichloroacetonitrile	3018-12-0	rs11215989	116407490	11	BUD13	1.62 E-07	0.036
8-Hydroxyquinoline	148-24-3	rs2705858	14897262	2	NBAS	1.63 E-07	0.066
1,1,1,2-Tetrabromoethane	630-16-0	rs80184146	53967078	7	POM121L12	1.64 E-07	0.043
17beta-Estradiol	50-28-2	rs77023555	41518715	8	MIR486	1.67 E-07	0.028
1,3-Diiminobenz (f)-isoindoline	65558-69-2	rs79099303	178107512	5	ZNF354A	1.68 E-07	0.045
4-Chloro-o-phenylenediamine	95-83-0	rs1863773	171861497	2	TLK1	1.68 E-07	0.059
Triamterene	396-01-0	rs80124975	55171150	7	EGFR	1.69 E-07	0.075
N-(1,3-Dimethylbutyl)-N'-phenyl-p-phenylenediamine	793-24-8	rs73077728	39666223	5	DAB2	1.70 E-07	0.031
Cycloheximide	66-81-9	rs938729	127996472	9	RABEPK	1.70 E-07	0.045
2-Amino-4-chlorophenol	95-85-2	rs74445911	99188202	3	COL8A1	1.71 E-07	0.057
Acetochlor	34256-82-1	rs75608771	181458496	4	ODZ3	1.73 E-07	0.051
N,N'-Diphenyl-p-phenylenediamine	74-31-7	rs852350	46828160	20	PREX1	1.73 E-07	0.057
Systhane	88671-89-0	rs76341199	73932870	6	KHDC1L	1.73 E-07	0.057
Aflatoxin B1 from Aspergillus flavus	1162-65-8	rs73946114	115126665	2	ACTR3	1.73 E-07	0.038
permethrin	52645-53-1	SNP2-19172082	19308601	2	OSR1	1.73 E-07	0.067
4-Chloro-o-phenylenediamine	95-83-0	rs35972823	127572550	9	OLFML2A	1.76 E-07	0.059
o-Aminophenol	95-55-6	rs73367814	144496350	8	MAFA	1.86 E-07	0.088
Acetochlor	34256-82-1	rs115850579	108846714	12	FICD	1.91 E-07	0.051
dichloroacetonitrile	3018-12-0	rs35516880	49704023	19	TRPM4	1.97 E-07	0.037
o-Aminophenol	95-55-6	rs116214814	145659037	1	RNF115	1.98 E-07	0.088
Aflatoxin B1 from Aspergillus flavus	1162-65-8	rs16863051	222617754	2	EPHA4	1.99 E-07	0.038
permethrin	52645-53-1	rs11879047	48535012	19	CABP5	2.00 E-07	0.067
4-Chloro-o-phenylenediamine	95-83-0	rs915258	99654001	X	PCDH19	2.01 E-07	0.059
1,3-Diiminobenz (f)-isoindoline	65558-69-2	rs78119276	40879472	13	FOXO1	2.05 E-07	0.045
Chlordecone (kepone)	143-50-0	rs73490735	7839616	16	A2BP1	2.12 E-07	0.047
p-Nitrosodiphenylamine	156-10-5	rs2697819	14073705	11	SPON1	2.12 E-07	0.07
7,12-Dimethylbenzanthracene	57-97-6	rs77953084	130557966	3	ATP2C1	2.22 E-07	0.074
Domiphen bromide	538-71-6	rs11110500	101123449	12	GAS2L3	2.23 E-07	0.054
4-Chloro-o-phenylenediamine	95-83-0	rs16886866	12267803	4	HS3ST1	2.23 E-07	0.059

Toxaphene	8001-35-2	rs74822199	92015924	1	CDC7	2.25 E-07	0.047
dichloroacetonitrile	3018-12-0	rs78443892	101205305	12	ANO4	2.34 E-07	0.039
17beta-Estradiol	50-28-2	rs76272767	31321204	3	STT3B	2.39 E-07	0.035
Domiphen bromide	538-71-6	rs77606067	12357201	2	LPIN1	2.41 E-07	0.054
8-Hydroxyquinoline	148-24-3	rs78767283	135001510	5	CXCL14	2.42 E-07	0.066
Domiphen bromide	538-71-6	rs77635386	97941484	14	VRK1	2.43 E-07	0.054
1,8-Dihydroxy-4,5-dinitroanthraquinone	81-55-0	rs61749065	37921424	1	LOC728431	2.48 E-07	0.064
2-Amino-4-chlorophenol	95-85-2	rs10980588	113609826	9	MUSK	2.49 E-07	0.066
Chlordecone (kepone)	143-50-0	rs13229	41165878	17	IFI35	2.52 E-07	0.047
N-Methyl-p-aminophenol sulfate	55-55-0	rs523638	23465886	1	TARBP1	2.54 E-07	0.067
Cycloheximide	66-81-9	rs114561024	8886997	6	SLC35B3	2.61 E-07	0.058
2,3,4,5-Tetrachlorophenol	4901-51-3	rs16930149	29601508	10	LYZL1	2.67 E-07	0.057
p-Nitrosodiphenylamine	156-10-5	rs72884793	59742833	3	FHIT	2.76 E-07	0.073
Toxaphene	8001-35-2	rs116409713	167134776	5	ODZ2	2.76 E-07	0.047
Toxaphene	8001-35-2	rs9660010	240588502	1	FMN2	2.82 E-07	0.047
Acetochlor	34256-82-1	rs114752023	49897238	2	FSHR	2.86 E-07	0.063
2,3,4,5-Tetrachlorophenol	4901-51-3	SNP5-131756739	131728840	5	SLC22A5	2.95 E-07	0.057
17beta-Estradiol	50-28-2	rs58049330	90611155	9	CDK20	2.96 E-07	0.039
2,3,4,5-Tetrachlorophenol	4901-51-3	rs1052179	71267010	10	TSPAN15	3.01 E-07	0.057
3-Chloro-4-(dichloromethyl)-5-hydroxy-2(5H)-furanone(MX)	77439-76-0	rs114377156	230451029	2	DNER	3.05 E-07	0.038
3-Chloro-4-(dichloromethyl)-5-hydroxy-2(5H)-furanone(MX)	77439-76-0	rs72984722	140332059	2	LRP1B	3.10 E-07	0.038
1-Naphthylamine	134-32-7	rs73742862	54763462	6	FAM83B	3.14 E-07	0.095
Nitazoxanide	55981-09-4	rs74060294	70178204	14	KIAA0247	3.15 E-07	0.084
Chlordecone (kepone)	143-50-0	rs1154879	42814417	14	LRFN5	3.16 E-07	0.047
Chlordecone (kepone)	143-50-0	rs74424738	194716032	3	C3orf21	3.17 E-07	0.047
8-Hydroxyquinoline	148-24-3	rs45507700	30782357	13	KATNAL1	3.18 E-07	0.066
Domiphen bromide	538-71-6	rs11568403	86955535	9	SLC28A3	3.20 E-07	0.061
dichloroacetonitrile	3018-12-0	rs115567386	84537793	1	TTL7	3.25 E-07	0.043
o-Phenylenediamine	95-54-5	rs114557691	16158386	6	MYLIP	3.34 E-07	0.087
1,3-Diiminobenz (f)-isoindoline	65558-69-2	rs79150690	23147311	9	ELAVL2	3.39 E-07	0.064
Cycloheximide	66-81-9	rs28679586	31188401	17	MYO1D	3.39 E-07	0.064
8-Hydroxyquinoline	148-24-3	rs12311965	131268403	12	STX2	3.49 E-07	0.066
Progesterone	57-83-0	rs77458047	98005227	8	PGCP	3.54 E-07	0.060

Progesterone	57-83-0	rs11214105	112037 653	11	TEX12	3.64 E-07	0.060
3-Chloro-4-(dichloromethyl)-5-hydroxy-2(5H)-furanone(MX)	77439-76-0	rs73460084	801736 64	9	GNA14	3.67 E-07	0.038
Chlordecone (kepone)	143-50-0	rs15087363	145799 3 41	20	MACRO D2	3.70 E-07	0.049
2,3,4,5-Tetrachlorophenol	4901-51-3	rs7734259	120522 6	5	SLC6A19	3.78 E-07	0.063
Cycloheximide	66-81-9	rs7263958	124846 78	20	SPTLC3	3.86 E-07	0.064
Domiphen bromide	538-71-6	rs11437715	230451 6 029	2	DNER	3.87 E-07	0.062
17beta-Estradiol	50-28-2	rs16893110	240929 18	5	PRDM9	3.89 E-07	0.047
4-Chloro-o-phenylenediamine	95-83-0	rs75858454	782004 96	9	MIR548 H3	4.42 E-07	0.093
1,1,1,2-Tetrabromoethane	630-16-0	rs74053705	917148 8	1	GPR157	4.48 E-07	0.085
p-Nitrosodiphenylamine	156-10-5	rs13390320	631649 2	2	SOX11	4.49 E-07	0.099
Domiphen bromide	538-71-6	rs1587323	251591 33	15	SNRPN	4.54 E-07	0.062
Progesterone	57-83-0	rs11702590	425246 28	21	BACE2	4.55 E-07	0.06
o-Phenylenediamine	95-54-5	rs76475950	540566 15	2	GPR75- ASB3	4.56 E-07	0.087
o-Phenylenediamine	95-54-5	rs16925298	708167 4	9	KDM4C	4.58 E-07	0.087
Cycloheximide	66-81-9	rs78767283	135001 510	5	CXCL14	4.59 E-07	0.068
Nitazoxanide	55981-09-4	rs12346608	280620	9	DOCK8	4.63 E-07	0.088
Domiphen bromide	538-71-6	rs56708270	279254 85	X	DCAF8L1	4.66 E-07	0.062
1,8-Dihydroxy-4,5-dinitroanthraquinone	81-55-0	rs11860817	553886 44	16	IRX6	4.76 E-07	0.079
17beta-Estradiol	50-28-2	rs10076309	927706 38	5	FLJ4270 9	4.80 E-07	0.052
4-Chloro-o-phenylenediamine	95-83-0	rs3211770	113793 849	13	F10	4.93 E-07	0.093
1-Naphthylamine	134-32-7	rs11481811	791433 8 99	7	MAGI2	5.02 E-07	0.095
1-Naphthylamine	134-32-7	rs59993949	129924 957	12	TMEM1 32D	5.07 E-07	0.095
Acetochlor	34256-82-1	rs73235161	207897 38	2	HS1BP3	5.16 E-07	0.078
1-Naphthylamine	134-32-7	rs6468813	103139 683	8	NCALD	5.29 E-07	0.095
3-Chloro-4-(dichloromethyl)-5-hydroxy-2(5H)-furanone(MX)	77439-76-0	rs78748755	665034 51	11	C11orf8 0	5.32 E-07	0.047
dichloroacetonitrile	3018-12-0	rs77088325	898513 2	20	PLCB1	5.33 E-07	0.059
Progesterone	57-83-0	rs7059570	264216 97	X	MAGEB6	5.43 E-07	0.060
1-Naphthylamine	134-32-7	SNP14- 50902158	518324 08	14	TMX1	5.49 E-07	0.095
Acetochlor	34256-82-1	rs74088312	965979 17	14	BDKRB2	5.51 E-07	0.078
3-Chloro-4-(dichloromethyl)-5-hydroxy-2(5H)-furanone(MX)	77439-76-0	rs6540459	208369 736	1	PLXNA2	5.64 E-07	0.047
Domiphen bromide	538-71-6	rs4237673	393670 04	11	LRR4C	5.66 E-07	0.068
Chlordecone (kepone)	143-50-0	rs73833731	568396 49	3	ARHGEF 3	5.79 E-07	0.064
1,8-Dihydroxy-4,5-dinitroanthraquinone	81-55-0	rs9567646	467630 28	13	LCP1	5.93 E-07	0.087

Cycloheximide	66-81-9	rs111782938	26528764	1	CATSPER4	6.04 E-07	0.076
Aflatoxin B1 from Aspergillus flavus	1162-65-8	rs76487712	35245296	2	CRIM1	6.11 E-07	0.075
Acetochlor	34256-82-1	rs11110500	101123449	12	GAS2L3	6.47 E-07	0.078
Acetochlor	34256-82-1	rs74059988	8880606	12	RIMKLB	6.48 E-07	0.078
Nitazoxanide	55981-09-4	rs9313437	168395192	5	SLIT3	6.55 E-07	0.097
Domiphen bromide	538-71-6	rs6028580	38282421	20	DHX35	6.56 E-07	0.073
17beta-Estradiol	50-28-2	rs75555706	96015792	8	C8orf38	6.57 E-07	0.058
Chlordecone (kepone)	143-50-0	rs114816993	63423482	8	NKAIN3	6.63 E-07	0.064
Chlordecone (kepone)	143-50-0	rs73126926	53696607	5	HSPB3	6.73 E-07	0.064
1,3-Diiminobenz (f)-isoindoline	65558-69-2	rs12813380	118296931	12	KSR2	6.95 E-07	0.097
1-Naphthylamine	134-32-7	rs73200328	14878222	2	NBAS	7.05 E-07	0.095
N-(1,3-Dimethylbutyl)-N'-phenyl-p-phenylenediamine	793-24-8	rs724997	173708117	2	RAPGEF4	7.08 E-07	0.083
Cycloheximide	66-81-9	rs11130253	50681158	3	MAPKAPK3	7.08 E-07	0.076
1-Naphthylamine	134-32-7	rs76046261	16577365	11	C11orf58	7.10 E-07	0.095
Aflatoxin B1 from Aspergillus flavus	1162-65-8	rs115521847	149101485	3	TM4SF1	7.12 E-07	0.075
1-Naphthylamine	134-32-7	rs55711976	42300980	19	CEACAM3	7.16 E-07	0.095
N-(1,3-Dimethylbutyl)-N'-phenyl-p-phenylenediamine	793-24-8	rs4427625	40182606	12	SLC2A13	7.17 E-07	0.083
1,3-Diiminobenz (f)-isoindoline	65558-69-2	rs4077144	127622908	2	LOC339760	7.22 E-07	0.097
Aflatoxin B1 from Aspergillus flavus	1162-65-8	rs62016193	6339309	16	RBFOX1	7.31 E-07	0.075
N-(1,3-Dimethylbutyl)-N'-phenyl-p-phenylenediamine	793-24-8	SNP16-62118696	63561195	16	CDH11	7.38 E-07	0.083
N-(1,3-Dimethylbutyl)-N'-phenyl-p-phenylenediamine	793-24-8	rs17015753	25111128	3	TOP2B	7.46 E-07	0.083
Chlordecone (kepone)	143-50-0	rs35972823	127572550	9	OLFML2A	7.90 E-07	0.07
Cycloheximide	66-81-9	rs77828608	29542970	22	KREMEN1	8.32 E-07	0.076
17beta-Estradiol	50-28-2	rs74631305	66772843	12	GRIP1	8.51 E-07	0.071
Progesterone	57-83-0	rs11009601	34237166	10	PARD3	8.57 E-07	0.071
Aflatoxin B1 from Aspergillus flavus	1162-65-8	rs281886	131785132	7	PLXNA4	8.65 E-07	0.082
Cycloheximide	66-81-9	rs114442034	16159615	7	ISPD	8.90 E-07	0.076
Cycloheximide	66-81-9	rs115317550	2476936	17	KIAA0664	9.15 E-07	0.076
Chlordecone (kepone)	143-50-0	rs9540605	66486429	13	PCDH9	9.26 E-07	0.070
N-(1,3-Dimethylbutyl)-N'-phenyl-p-phenylenediamine	793-24-8	rs599023	60172344	13	DIAPH3	9.86 E-07	0.094
3-Chloro-4-(dichloromethyl)-5-hydroxy-2(5H)-furanone(MX)	77439-76-0	rs74444408	27796616	14	NOVA1	9.96 E-07	0.075
Acetochlor	34256-82-1	rs947213	105682496	6	PREP	1.01 E-06	0.100
Chlordecone (kepone)	143-50-0	rs9660010	240588502	1	FMN2	1.10 E-06	0.070

Chlordecone (kepone)	143-50-0	rs11836196	163919 99	12	SLC15A5	1.10 E-06	0.070
Chlordecone (kepone)	143-50-0	rs6076084	235285 36	20	CST9L	1.10 E-06	0.070
Chlordecone (kepone)	143-50-0	rs78571529	543504 74	2	ACYP2	1.11 E-06	0.07
N-(1,3-Dimethylbutyl)-N'-phenyl-p-phenylenediamine	793-24-8	rs77953084	130557 966	3	ATP2C1	1.13 E-06	0.094
Acetochlor	34256- 82-1	rs76289940	949003 17	12	TMCC3	1.15 E-06	0.100
Progesterone	57-83-0	rs77697536	199224 99	2	TTC32	1.16 E-06	0.071
3-Chloro-4-(dichloromethyl)-5-hydroxy-2(5H)- furanone(MX)	77439- 76-0	rs77746073	117912 00	12	ETV6	1.17 E-06	0.080
Chlordecone (kepone)	143-50-0	rs2697819	140737 05	11	SPON1	1.18 E-06	0.071
Acetochlor	34256- 82-1	rs56922398	552184 60	6	GFRAL	1.19 E-06	0.100
Chlordecone (kepone)	143-50-0	rs79304881	164845 251	4	1-Mar	1.23 E-06	0.071
Progesterone	57-83-0	rs292582	134882 680	7	WDR91	1.25 E-06	0.071
Acetochlor	34256- 82-1	rs79450652	180881 757	2	CWC22	1.25 E-06	0.100
3-Chloro-4-(dichloromethyl)-5-hydroxy-2(5H)- furanone(MX)	77439- 76-0	rs74631305	667728 43	12	GRIP1	1.31 E-06	0.083
Acetochlor	34256- 82-1	rs58612427	687863 70	2	APLF	1.35 E-06	0.100
Acetochlor	34256- 82-1	rs73432502	224269 49	13	FGF9	1.35 E-06	0.100
3-Chloro-4-(dichloromethyl)-5-hydroxy-2(5H)- furanone(MX)	77439- 76-0	rs1545116	589932 26	16	GOT2	1.44 E-06	0.083
Chlordecone (kepone)	143-50-0	rs35315427	87179	17	RPH3AL	1.56 E-06	0.084
3-Chloro-4-(dichloromethyl)-5-hydroxy-2(5H)- furanone(MX)	77439- 76-0	rs77635386	979414 84	14	VRK1	1.69 E-06	0.090
Progesterone	57-83-0	rs4758204	761904 1	11	PPFIBP2	1.73 E-06	0.084
Chlordecone (kepone)	143-50-0	rs28539080	599446 06	20	CDH4	1.75 E-06	0.087
Progesterone	57-83-0	rs8044956	704970 86	16	FUK	1.84 E-06	0.084
3-Chloro-4-(dichloromethyl)-5-hydroxy-2(5H)- furanone(MX)	77439- 76-0	rs78468118	827006 51	16	CDH13	1.90 E-06	0.093
3-Chloro-4-(dichloromethyl)-5-hydroxy-2(5H)- furanone(MX)	77439- 76-0	rs78297143	436043 13	14	LRFN5	2.14 E-06	0.098
Chlordecone (kepone)	143-50-0	rs9931421	779260 21	16	VAT1L	2.19 E-06	0.098
Chlordecone (kepone)	143-50-0	rs7868175	665808 0	9	GLDC	2.56 E-06	0.098
Chlordecone (kepone)	143-50-0	rs2012691	105453 745	7	ATXN7L1	2.62 E-06	0.098
Chlordecone (kepone)	143-50-0	rs74059988	888060 6	12	RIMKLB	2.71 E-06	0.098
Chlordecone (kepone)	143-50-0	rs73808311	334980 9	3	CRBN	2.75 E-06	0.098
Chlordecone (kepone)	143-50-0	rs61214761	114039 909	7	FOXP2	2.75 E-06	0.098
Chlordecone (kepone)	143-50-0	rs56381502	258459 22	5	CDH9	2.97 E-06	0.098
Chlordecone (kepone)	143-50-0	rs79494514	341223 46	13	STARD1 3	2.97 E-06	0.098
Chlordecone (kepone)	143-50-0	rs79843432	239514 826	1	FMN2	2.99 E-06	0.098
Chlordecone (kepone)	143-50-0	rs73145960	519209 9	2	SOX11	3.02 E-06	0.098

^aNCBI build 37 . ^b *P*-value. ^c FDR *q*-value obtained per chemical.

Table 3.5. Significant EC10 –SNP associations among larger set of 12m SNPs

Chemical	CAS #	SNP	bp	Chrom	gene	P	q_value
Vitamin D3	67-97-0	rs75591162	179520663	3	PEX5L	2.79E-14	3.47E-07
Amiloride hydrochloride	2016-88-8	rs149885464	114388490	X	LRCH2	9.75E-14	1.21E-06
Systhane	88671-89-0	rs114097262	57530631	20	TH1L	2.22E-12	2.76E-05
Cycloheximide	66-81-9	rs112895158	143565653	3	C3orf58	3.19E-12	3.96E-05
Hexachlorophene	70-30-4	rs149041648	4081594	7	SDK1	3.88E-12	4.82E-05
17beta-Estradiol	50-28-2	rs148557261	85025463	X	CHM	4.44E-12	2.76E-05
1,8-Dihydroxy-4,5-dinitroanthraquinone	81-55-0	rs115918434	70712674	6	COL19A1	5.56E-12	6.90E-05
o-Nitrobenzyl chloride	612-23-7	rs9966852	20953234	18	C18orf45	9.72E-12	1.00E-04
Cycloheximide	66-81-9	rs116560380	53916956	16	FTO	1.70E-11	1.00E-04
1,3-Diiminobenz (f)-isoindoline	65558-69-2	rs150407211	204994131	1	NFASC	1.82E-11	2.00E-04
N-Methyl-p-aminophenol sulfate	55-55-0	rs139684082	55659431	20	BMP7	1.88E-11	2.00E-04
2,2'-Thiobis(4,6-dichlorophenol)	97-18-7	rs78022668	220093749	1	EPRS	2.19E-11	9.06E-05
17beta-Estradiol	50-28-2	rs114097262	57530631	20	TH1L	2.90E-11	6.86E-05
1,1,1,2-Tetrabromoethane	630-16-0	rs116253534	7734317	3	GRM7	3.24E-11	4.00E-04
Saquinavir mesylate (AIDS Initiative)	149845-06-7	rs77326389	27959745	15	OCA2	3.77E-11	5.00E-04
Nitazoxanide	55981-09-4	rs116345917	38690860	5	LIFR	4.41E-11	3.91E-05
Dichlorvos (Vapona)	62-73-7	rs5915692	4004014	X	PRKX	5.19E-11	6.00E-04
17beta-Estradiol	50-28-2	rs77919593	178511146	2	PDE11A	5.52E-11	6.86E-05
Ziram	137-30-4	rs114456684	89021170	13	SLITRK5	6.92E-11	9.00E-04
cis-Dichlorodiamine platinum	15663-27-1	rs13236745	21072550	7	SP8	7.19E-11	3.00E-04
Nitazoxanide	55981-09-4	rs114256919	106718263	14	LOC100288568	7.81E-11	6.06E-05
1,6-Hexamethylene diacrylate	13048-33-4	rs62231930	24939620	22	C22orf13	7.91E-11	0.001
N,N,N',N'-Tetramethyl-p-phenylenediamine	100-22-1	rs111412077	92398645	11	FAT3	9.36E-11	5.00E-04
p-Nitrosodiphenylamine	156-10-5	rs78789459	188016223	4	FAT1	1.04E-10	7.00E-04
Hexachlorophene	70-30-4	rs185960291	107817291	11	RAB39	1.11E-10	7.00E-04
p-Nitrosodiphenylamine	156-10-5	rs185480677	214584272	1	PTPN14	1.14E-10	7.00E-04
17beta-Estradiol	50-28-2	rs77023555	41518715	8	ANK1	1.20E-10	1.00E-04

1-Methyl-3-nitro-1-nitroso-guanidine	81-55-0	rs139598440	89777354	X	TGIF2LX	1.22E-10	5.00E-04
1,8-Dihydroxy-4,5-dinitroanthraquinone	70-25-7	rs115531561	69465188	2	ANTXR1	1.22E-10	0.002
o-Aminophenol	95-55-6	rs182028010	230236320	1	GALNT2	1.26E-10	0.002
1,3-Diiminobenz (f)-isoindoline	65558-69-2	rs146270840	194601175	3	FAM43A	1.56E-10	6.00E-04
Glutaraldehyde	111-30-8	rs7992282	31601586	13	C13orf26	1.58E-10	0.002
Nitazoxanide	55981-09-4	rs182761810	47129104	X	USP11	1.59E-10	1.00E-04
4-Chloro-o-phenylenediamine	95-83-0	rs146160738	103390555	X	MCART6	1.68E-10	0.002
Systhane	88671-89-0	rs111763750	53503234	13	PCDH8	1.72E-10	3.00E-04
6-Thioguanine (6-TG)	154-42-7	rs72484656	73259654	9	TRPM3	1.99E-10	0.003
1,3-Diiminobenz (f)-isoindoline	65558-69-2	rs148557261	85025463	X	CHM	2.02E-10	6.00E-04
Vitamin D3	67-97-0	rs79427953	155449132	5	SGCD	2.31E-10	9.00E-04
p-Quinone	106-51-4	rs140356758	96472953	13	UGGT2	2.35E-10	0.003
Vitamin D3	67-97-0	rs73001915	180641510	4	LOC285501	2.47E-10	9.00E-04
4-Chloro-o-phenylenediamine	95-83-0	rs61790870	34693156	4	ARAP2	2.53E-10	0.002
17beta-Estradiol	50-28-2	rs138873795	93861771	15	RGMA	2.59E-10	3.00E-04
Retinal	116-31-4	rs61741388	114156579	9	KIAA0368	3.11E-10	0.002
17beta-Estradiol	50-28-2	rs9957628	57478543	18	PMAIP1	3.18E-10	3.00E-04
1,2-Epoxy-3-chloropropane (Epichlorohydrin)	106-89-8	rs114097262	57530631	20	TH1L	3.35E-10	0.004
Vitamin D3	67-97-0	rs77571603	29323478	10	LYZL1	3.37E-10	9.00E-04
Chlordecone (kepone)	143-50-0	rs6834412	4731136	4	MSX1	3.58E-10	0.003
Cycloheximide	66-81-9	rs4075856	127996778	9	RABEPK	3.66E-10	8.00E-04
N,N,N',N'-Tetramethyl-p-phenylenediamine	100-22-1	rs139345780	35242366	2	CRIM1	3.73E-10	0.001
Vitamin D3	67-97-0	rs140124941	58119464	12	AGAP2	3.75E-10	9.00E-04
Sodium dichromate dihydrate (VI)	7789-12-0	rs144044660	175439270	4	HPGD	3.76E-10	0.005
Systhane	88671-89-0	rs182761810	47129104	X	USP11	3.79E-10	5.00E-04
Azathioprine	446-86-6	rs72625563	30966462	X	TAB3	3.90E-10	0.004
2',4',5'-Trihydroxybutyrophenone	1421-63-2	rs77671255	18038964	7	PRPS1L1	4.15E-10	0.005
Retinol acetate	127-47-9	rs115240733	83143062	8	SNX16	4.28E-10	0.005
N,N'-Diphenyl-p-phenylenediamine	74-31-7	rs150553218	180241673	4	LOC285501	4.57E-10	5.00E-04

1,3-Diiminobenz (f)-isoindoline	65558-69-2	rs79373644	68284871	17	KCNJ2	4.60E-10	0.001
N,N-Diethyl-p-phenylenediamine	93-05-0	rs184971028	23937090	9	ELAVL2	4.63E-10	0.006
Nifedipine	21829-25-4	rs74383932	211608006	2	CPS1	4.76E-10	0.002
o-Nitrobenzyl chloride	612-23-7	rs75820586	88320786	14	GALC	4.89E-10	0.003
17beta-Estradiol	50-28-2	rs114750717	143943362	X	SPANXN1	4.99E-10	4.00E-04
Guggulsterones E	39025-24-6	rs74929760	146457160	4	MMAA	5.00E-10	0.006
Hydroquinone	123-31-9	rs78268501	7128396	4	SORCS2	5.03E-10	0.006
Endosulfan	115-29-7	rs75512317	132143171	X	USP26	5.16E-10	0.006
2,2'-Thiobis(4,6-dichlorophenol)	97-18-7	rs76656117	24217993	10	KIAA1217	5.29E-10	0.002
Nifedipine	21829-25-4	rs187575156	33840771	16	ZNF267	5.51E-10	0.002
Hexachlorophene	70-30-4	rs188272213	38447642	5	EGFLAM	5.59E-10	0.002
Nitazoxanide	55981-09-4	rs115715858	88034425	15	AGBL1	6.03E-10	3.00E-04
Rhein (1,8-dihydroxy-3-carboxyl anthraquinone)	478-43-3	rs181686881	28901928	15	GOLGA8F	6.04E-10	0.008
Mercuric chloride	7487-94-7	rs115872305	65584758	5	SFRS12	6.05E-10	0.008
Saquinavir mesylate (AIDS Initiative)	50-28-2	rs115129751	102066268	14	DIO3	6.16E-10	5.00E-04
17beta-Estradiol	149845-06-7	rs28712763	38664057	5	LIFR	6.16E-10	0.002
Chlordecone (kepone)	143-50-0	rs28502033	71440392	7	WBSCR17	6.28E-10	0.003
1,8-Dihydroxy-4,5-dinitroanthraquinone	81-55-0	rs186751172	137938171	X	FGF13	7.36E-10	0.001
1-Naphthylamine	134-32-7	rs56166709	26718141	16	JMJD5	7.81E-10	0.002
Oxymetholone	434-07-1	rs118085300	72050372	10	NPFFR1	7.90E-10	0.010
1-Naphthylamine	134-32-7	rs148976999	89292543	2	EIF2AK3	7.96E-10	0.002
Fumaronitrile	764-42-1	rs112841261	95693291	8	ESRP1	8.09E-10	0.010
Nitazoxanide	55981-09-4	rs139729800	176909149	1	ASTN1	8.33E-10	3.00E-04
Pentaerythritol triacrylate	3524-68-3	rs12263759	127748420	10	ADAM12	8.37E-10	0.007
Cycloheximide	66-81-9	rs2774024	21573881	1	ECE1	8.42E-10	9.00E-04
Saquinavir mesylate (AIDS Initiative)	149845-06-7	rs115362446	59318384	10	IPMK	8.50E-10	0.002
9-Aminoacridine, monohydrochloride, monohydrate	52417-22-8	rs148846460	115474911	X	SLC6A14	8.52E-10	0.011
Saquinavir mesylate (AIDS Initiative)	149845-06-7	rs140911187	32806643	11	CCDC73	8.61E-10	0.002
p-Nitrosodiphenylamine	156-10-5	rs115484855	116269294	10	ABLIM1	8.63E-10	0.002

						10	
2,2',4'-Trichloroacetophenone	4252-78-2	rs73371185	3098435	18	MYOM1	8.84E-10	0.011
Sodium dichromate dihydrate (VI)	7789-12-0	rs117147263	3894712	12	PARP11	9.07E-10	0.006
1,1,1,2-Tetrabromoethane	630-16-0	rs73561088	87780267	13	SLITRK5	9.11E-10	0.002
Azathioprine	446-86-6	rs191755607	40475255	X	ATP6AP2	9.13E-10	0.004
Retinal	116-31-4	rs113516752	106955309	14	LOC100288568	9.20E-10	0.002
Retinal	116-31-4	rs115605961	36196063	3	STAC	9.46E-10	0.002
o-Phenylenediamine	95-54-5	rs78961197	55949853	19	SHISA7	9.46E-10	0.006
p-Aminophenol	123-30-8	rs113103160	107838962	11	RAB39	9.50E-10	0.007
2,4-Hexadienal	142-83-6	rs79177535	114025195	11	ZBTB16	9.74E-10	0.012
Ziram	137-30-4	rs116855907	21587330	16	METTL9	1.00E-09	0.006
Diethylene glycol diacrylate	4074-88-8	rs78140462	110470359	12	ANKRD13A	1.02E-09	0.013
Cetylpyridinium bromide	140-72-7	rs80257852	23561114	9	ELAVL2	1.03E-09	0.013
Azathioprine	3018-12-0	rs28365025	183643255	3	ABCC5	1.06E-09	0.003
dichloroacetonitrile	446-86-6	rs117910051	164141498	6	PACRG	1.06E-09	0.004
1,1,1,2-Tetrabromoethane	630-16-0	rs142817312	110092992	10	SORCS1	1.09E-09	0.002
Saquinavir mesylate (AIDS Initiative)	149845-06-7	rs145544371	6472066	X	VCX3A	1.10E-09	0.002
Colchicine	64-86-8	rs80328096	19803987	3	EFHB	1.11E-09	0.002
Pentaerythritol triacrylate	3524-68-3	rs192972674	72721640	X	PABPC1L2A	1.11E-09	0.007
2,3,4,5-Tetrachloronitrobenzene	879-39-0	rs115782131	85339219	2	TCF7L1	1.13E-09	0.014
p-Nitrosodiphenylamine	50-28-2	rs12087685	23727348	1	TCEA3	1.15E-09	8.00E-04
17beta-Estradiol	156-10-5	rs193048198	243945733	1	ZNF238	1.15E-09	0.002
Phenformin hydrochloride	834-28-6	rs140883532	53503005	20	CYP24A1	1.16E-09	0.014
Azathioprine	446-86-6	rs146213067	136068920	X	GPR101	1.17E-09	0.004
p-Aminophenol	123-30-8	rs57325800	10214905	1	UBE4B	1.19E-09	0.007
1,3,5-Triglycidyl isocyanurate	2451-62-9	rs150927141	11429085	1	UBIAD1	1.19E-09	0.007
Tetramethylthiouram disulfide	100-27-6	rs140556470	48843069	X	GRIPAP1	1.20E-09	0.004
p-Nitrophenethyl alcohol	137-26-8	rs142624309	58608899	1	OMA1	1.20E-09	0.012
Progesterone	143-50-0	rs185665429	129727410	8	MYC	1.22E-09	0.003
Chlordecone (kepone)	57-83-0	rs73285294	153826706	5	SAP30L	1.22E-09	0.015

						09	
17beta-Estradiol	50-28-2	rs76191860	116013224	12	MED13L	1.24E-09	8.00E-04
Retinol acetate	127-47-9	rs7838214	118761053	8	EXT1	1.29E-09	0.008
1,2-Epoxy-3-chloropropane (Epichlorohydrin)	106-89-8	rs77143142	104926726	12	TXNRD1	1.32E-09	0.005
o-Phenylenediamine	88671-89-0	rs11917950	8725408	3	C3orf32	1.37E-09	0.002
Systhane	95-54-5	rs138701165	33431943	X	DMD	1.37E-09	0.006
1,1,1,2-Tetrabromoethane	630-16-0	rs113917611	145680395	5	RBM27	1.41E-09	0.002
Ergotamine tartrate	379-79-3	rs78908092	91139027	8	CALB1	1.42E-09	0.002
Nitrogen mustard hydrochloride	55-86-7	rs185629858	72754141	9	MAMDC2	1.44E-09	0.009
Nitrogen mustard hydrochloride	55-86-7	rs150079115	229475297	1	C1orf96	1.46E-09	0.009
Triamterene	396-01-0	rs116253534	7734317	3	GRM7	1.47E-09	0.018
t-Butylhydroquinone	1948-33-0	rs78839292	4302598	6	PECI	1.48E-09	0.018
Pyrimethamine	58-14-0	rs150650924	19223048	16	SYT17	1.53E-09	0.019
1,2-Epoxy-3-chloropropane (Epichlorohydrin) dichloroacetonitrile	106-89-8	rs73380687	33444259	15	FMN1	1.56E-09	0.005
	3018-12-0	rs116561096	173810265	4	GALNT7	1.57E-09	0.004
Chlordecone (kepone)	143-50-0	rs189105020	53254392	X	IQSEC2	1.60E-09	0.003
1,1,1,2-Tetrabromoethane	630-16-0	rs113557926	37267705	2	HEATR5B	1.61E-09	0.002
2-Chloroacetophenone (CN)	532-27-4	rs138236656	17144946	17	FLCN	1.63E-09	0.020
p-Nitrophenethyl alcohol	100-27-6	rs74821785	180732042	4	LOC285501	1.65E-09	0.004
Diisobutyl phthalate	84-69-5	rs59163027	3618869	8	CSMD1	1.71E-09	0.021
Systhane	88671-89-0	rs114062278	67964362	5	PIK3R1	1.75E-09	0.002
Chlordecone (kepone)	143-50-0	rs147841840	121413130	12	HNF1A	1.76E-09	0.003
Toxaphene	8001-35-2	rs2232859	24890981	22	UPB1	1.77E-09	0.021
Dexamethazone	50-02-2	rs148291248	94417682	1	ABCA4	1.77E-09	0.022
2-Octyl-3-isothiazolone	26530-20-1	rs114113654	104765328	2	POU3F3	1.83E-09	0.023
Vitamin D3	67-97-0	rs115033986	35362684	2	CRIM1	1.87E-09	0.002
Vitamin D3	67-97-0	rs76371158	87850932	11	RAB38	1.90E-09	0.002
Amiloride hydrochloride	2016-88-8	rs138475390	144438110	X	SPANXN1	1.94E-09	0.008
3,4-Diaminotoluene	496-72-0	rs187884265	30037137	X	MAGEB2	1.94E-09	0.024
Endosulfan	115-29-7	rs141860023	147034359	X	FMR1	1.98E-09	0.012

						09	
Tetramethylthiouam disulfide	137-26-8	rs189210853	100196239	1	FRRS1	1.98E-09	0.012
2',4',5'-Trihydroxybutyrophenone	1421-63-2	rs10213832	172302409	5	ERGIC1	1.98E-09	0.012
Diethylene glycol diacrylate	149845-06-7	rs76663109	219066443	1	LYPLAL1	2.01E-09	0.004
Saquinavir mesylate (AIDS Initiative)	4074-88-8	rs34600565	51549528	15	CYP19A1	2.01E-09	0.013
1,3-Diiminobenz (f)-isoindoline	65558-69-2	rs184897828	23101302	7	KLHL7	2.02E-09	0.003
Hydroquinone	123-31-9	rs259613	238432583	1	ZP4	2.05E-09	0.013
Systhane	88671-89-0	rs115987646	17518759	21	USP25	2.08E-09	0.002
Vitamin D3	67-97-0	rs114742274	78274036	9	PCSK5	2.11E-09	0.002
Hexachlorophene	70-30-4	rs4599696	72213159	7	TYW1B	2.16E-09	0.007
p-Nitrosodiphenylamine	156-10-5	rs76022962	57713812	16	GPR97	2.19E-09	0.003
1-Naphthylamine	143-50-0	rs79365910	127437542	5	SLC12A2	2.28E-09	0.003
Chlordecone (kepone)	134-32-7	rs73982303	20719652	17	CCDC144NL	2.28E-09	0.003
Fumaronitrile	764-42-1	rs150469316	38260872	15	TMCO5A	2.33E-09	0.015
13-cis-Retinal	472-86-6	rs141310758	57507867	20	GNAS	2.35E-09	0.004
N,N-Diethyl-p-phenylenediamine	93-05-0	rs141821443	78500126	16	WVOX	2.36E-09	0.015
Sulfathiazole	72-14-0	rs192596112	18253562	17	SHMT1	2.37E-09	0.023
Malachite green oxalate	2437-29-8	rs77904586	3698213	3	LRRN1	2.38E-09	0.018
Dibromonitromethane (water disinfection byproducts)	598-91-4	rs139080195	33110833	6	COL11A2	2.51E-09	0.005
Pentaerythritol triacrylate	3524-68-3	rs114097262	57530631	20	TH1L	2.56E-09	0.011
Aldicarb	116-06-3	rs117120562	76067201	7	ZP3	2.59E-09	0.032
Melatonin	472-86-6	rs5030075	186456754	3	KNG1	2.63E-09	0.004
13-cis-Retinal	73-31-4	rs146412275	190261780	3	TMEM207	2.63E-09	0.033
Retinol acetate	127-47-9	rs112136721	76923060	1	ST6GALNAC3	2.71E-09	0.008
Captan	133-06-2	rs80260839	55452081	11	OR4C6	2.72E-09	0.018
Nitazoxanide	55981-09-4	rs112047343	25684936	6	SCGN	2.75E-09	9.00E-04
Vitamin D3	67-97-0	rs16908781	122453569	9	DBC1	2.76E-09	0.002
Chlorhexidine	55-56-1	rs2069493	99967186	14	CCNK	2.78E-09	0.035
2-Amino-4-chlorophenol	95-85-2	rs149669986	188717980	4	TRIML2	2.82E-09	0.032

Flutamide (pubertal study)	13311-84-7	rs147788105	36895798	14	SFTA3	2.92E-09	0.036
8-Hydroxyquinoline	148-24-3	rs190163116	14927274	2	NBAS	2.93E-09	0.036
p-Nitrosodiphenylamine	156-10-5	rs142249979	22127356	3	ZNF385D	2.96E-09	0.003
p-Nitrosodiphenylamine	156-10-5	rs139756912	53520212	19	ZNF160	3.02E-09	0.003
13-cis-Retinal	472-86-6	rs112109406	41635317	19	CYP2F1	3.06E-09	0.004
Mono(2-ethylhexyl)phthalate	4376-20-9	rs142849625	87323965	X	KLHL4	3.08E-09	0.038
o-Phenylenediamine	95-54-5	rs78673081	159546875	6	FNDC1	3.14E-09	0.006
2-Chloroacetophenone (CN)	532-27-4	rs7885074	79470548	X	TBX22	3.15E-09	0.020
Nitazoxanide	55981-09-4	rs147784707	39731980	19	IL28B	3.22E-09	0.001
1,2-Epoxy-3-chloropropane (Epichlorohydrin)	106-89-8	rs113130251	16684250	8	FGF20	3.24E-09	0.008
dichloroacetonitrile	3018-12-0	rs17517181	132048205	9	C9orf106	3.25E-09	0.004
dichloroacetonitrile	3018-12-0	rs116202592	118722702	8	EXT1	3.25E-09	0.004
Retinal	116-31-4	rs75551402	144924136	6	UTRN	3.25E-09	0.005
Hydroquinone	123-31-9	rs141081361	12301484	6	EDN1	3.36E-09	0.014
Mercuric chloride	7487-94-7	rs114980796	146855180	7	CNTNAP2	3.44E-09	0.012
1,1,1,2-Tetrabromoethane	630-16-0	rs76616047	165230707	1	LMX1A	3.49E-09	0.003
Chlordecone (kepone)	143-50-0	rs117340306	159702181	6	FNDC1	3.49E-09	0.041
Cycloheximide	74-31-7	rs140838976	42607192	2	EML4	3.50E-09	0.001
N,N'-Diphenyl-p-phenylenediamine	66-81-9	rs141347833	95509996	8	KIAA1429	3.50E-09	0.003
1,3-Diiminobenz (f)-isoindoline	65558-69-2	rs115797302	71579045	5	MRPS27	3.56E-09	0.003
Nitazoxanide	55981-09-4	rs9587366	108038145	13	FAM155A	3.58E-09	0.001
Nitazoxanide	55981-09-4	rs112507751	13672142	7	ETV1	3.59E-09	0.001
1,2-Epoxy-3-chloropropane (Epichlorohydrin)	106-89-8	rs143124581	127508299	2	GYPC	3.66E-09	0.008
Chlordecone (kepone)	143-50-0	rs9305010	23959373	19	ZNF681	3.69E-09	0.003
Chlorambucil	100-22-1	rs73477286	10512356	X	CLCN4	3.73E-09	0.008
N,N,N',N'-Tetramethyl-p-phenylenediamine	305-03-3	rs148476652	230219445	2	DNER	3.73E-09	0.024
HC blue 2	33229-34-4	rs150501661	80565923	2	CTNNA2	3.78E-09	0.047
Cycloheximide	66-81-9	rs77426543	23327356	14	MMP14	3.86E-09	0.003
p-Nitrophenethyl alcohol	100-27-6	rs7235815	42824095	18	SETBP1	3.91E-09	0.008

Methylene bis(thiocyanate)	6317-18-6	rs111371582	200201474	1	FAM58B	3.94E-09	0.049
Azathioprine	446-86-6	rs145504708	104046255	14	C14orf153	3.97E-09	0.008
1,8-Dihydroxy-4,5-dinitroanthraquinone	81-55-0	rs58965029	55366235	7	LANCL2	4.02E-09	0.002
2,2',4'-Trichloroacetophenone	4252-78-2	rs149411402	24732313	X	POLA1	4.05E-09	0.025
Retinol acetate	127-47-9	rs7220141	8596496	17	CCDC42	4.18E-09	0.008
Glutaraldehyde	111-30-8	rs17080050	179587743	5	RASGEF1C	4.22E-09	0.005
9-Aminoacridine, monohydrochloride, monohydrate	52417-22-8	rs143574721	71726068	4	GRSF1	4.23E-09	0.019
dichloroacetonitrile	3018-12-0	rs116306510	9045458	20	PLCB1	4.25E-09	0.004
Saquinavir mesylate (AIDS Initiative)	149845-06-7	rs148557261	85025463	X	CHM	4.30E-09	0.005
1,3-Diiminobenz (f)-isoindoline	65558-69-2	rs1937167	121239310	X	GLUD2	4.32E-09	0.003
4-Chloro-3,5-dinitro-a,a,-trifluorotoluene	393-75-9	rs6784365	16264220	3	GALNTL2	4.37E-09	0.009
Phenylmercuric acetate	62-38-4	rs183322569	144144565	X	SPANXN1	4.38E-09	0.054
1,3-Dicyclohexylcarbodiimide	538-75-0	rs185941966	48916163	X	CCDC120	4.42E-09	0.055
Ergotamine tartrate	379-79-3	rs7060812	143883587	X	SPANXN1	4.45E-09	0.002
1-Naphthylamine	134-32-7	rs76224072	66811251	3	KBTBD8	4.45E-09	0.005
Captan	133-06-2	rs117766494	134786971	10	C10orf93	4.55E-09	0.029
Vitamin D3	67-97-0	rs852471	5679853	7	FSCN1	4.61E-09	0.003
1,1,1,2-Tetrabromoethane	630-16-0	rs75201540	107697197	3	CD47	4.63E-09	0.003
9-Aminoacridine, monohydrochloride, monohydrate	52417-22-8	rs150993793	107634677	6	PDSS2	4.64E-09	0.019
Captan	133-06-2	rs143629216	108864949	X	KCNE1L	4.69E-09	0.029
Saquinavir mesylate (AIDS Initiative)	149845-06-7	rs77216777	7737628	16	A2BP1	4.74E-09	0.005
Sodium dichromate dihydrate (VI)	7789-12-0	rs76202242	77000874	15	SCAPER	4.80E-09	0.020
p-Nitrosodiphenylamine	156-10-5	rs13390320	6316492	2	SOX11	4.87E-09	0.004
2-Amino-4-chlorophenol	95-85-2	rs114488597	6692895	9	GLDC	5.09E-09	0.032
Tamoxifen citrate	54965-24-1	rs73418488	39876321	22	MGAT3	5.11E-09	0.017
Chlordecone (kepone)	143-50-0	rs7604210	220464122	2	STK11IP	5.14E-09	0.004
17beta-Estradiol	50-28-2	rs73515607	12081555	16	TNFRSF17	5.16E-09	0.002
Nitazoxanide	55981-09-4	rs142350989	11163875	18	FAM38B	5.18E-09	0.002

Glutaraldehyde	111-30-8	rs145788790	76706431	3	ZNF717	5.26E-09	0.005
o-Phenylenediamine	95-54-5	rs10487606	24377991	7	NPY	5.27E-09	0.008
Tetrachlorvinphos	961-11-5	rs17037561	108142257	4	DKK2	5.40E-09	0.065
Endosulfan	115-29-7	rs77954654	28003512	7	JAZF1	5.42E-09	0.022
Turmeric (>98% curcumin)	458-37-7	rs151050814	3717523	2	ALLC	5.47E-09	0.068
Chlordecone (kepone)	143-50-0	rs114366607	9548336	11	ZNF143	5.48E-09	0.004
17beta-Estradiol	50-28-2	rs78044298	13282331	17	HS3ST3A1	5.61E-09	0.002
17beta-Estradiol	50-28-2	rs73400780	100613871	7	MUC12	5.62E-09	0.002
p-Nitrosodiphenylamine	156-10-5	rs8056610	80414263	16	DYNLRB2	5.69E-09	0.004
1,1,1,2-Tetrabromoethane	630-16-0	rs114472896	51352245	16	SALL1	5.70E-09	0.004
Azobenzene	103-33-3	rs8034597	50564358	15	HDC	5.81E-09	0.022
Catechol	120-80-9	rs5970515	150169694	X	HMGB3	6.01E-09	0.044
Dichlorvos (Vapona)	62-73-7	rs147257714	119289133	4	PRSS12	6.07E-09	0.025
Alizarin Yellow R, free acid	2243-76-7	rs9484620	142617993	6	GPR126	6.08E-09	0.076
p-Benzoquinone dioxime	105-11-3	rs116199788	14629043	9	ZDHHC21	6.09E-09	0.025
Digoxin	20830-75-5	rs4668546	7802776	2	RNF144A	6.17E-09	0.033
Retinol acetate	127-47-9	rs77518266	16948889	9	CNTLN	6.18E-09	0.008
Vitamin D3	67-97-0	rs147712607	131033194	12	RIMBP2	6.21E-09	0.003
o-Phenylenediamine	95-54-5	rs10197500	6314210	2	SOX11	6.24E-09	0.009
Tamoxifen citrate	54965-24-1	rs140206324	162533710	3	OTOL1	6.24E-09	0.017
1,3-Diiminobenz (f)-isoindoline	65558-69-2	rs111425337	45991437	20	ZMYND8	6.25E-09	0.003
4-Chloro-3,5-dinitro-a,a,-trifluorotoluene	393-75-9	rs73317253	111956225	8	KCNV1	6.27E-09	0.009
Vitamin D3	67-97-0	rs114278053	41810548	14	LRFN5	6.28E-09	0.003
p-Nitrophenethyl alcohol	100-27-6	rs114585626	6905318	16	A2BP1	6.28E-09	0.009
1-Naphthylamine	134-32-7	rs116257974	83616924	7	SEMA3A	6.31E-09	0.005
Nitazoxanide	55981-09-4	rs16887170	55652860	6	BMP5	6.34E-09	0.002
Mono(2-ethylhexyl)phthalate	4376-20-9	rs140233380	150658923	X	PASD1	6.35E-09	0.039
1,1,1,2-Tetrabromoethane	630-16-0	rs112605190	74299348	17	QRICH2	6.48E-09	0.004
Bisphenol A diglycidyl ether	1675-54-3	rs189546210	89544224	2	EIF2AK3	6.52E-09	0.077

Cycloheximide	66-81-9	rs59128236	187624402	3	BCL6	6.60E-09	0.003
1,1,1,2-Tetrabromoethane	630-16-0	rs77326389	27959745	15	OCA2	6.70E-09	0.004
Rhein (1,8-dihydroxy-3-carboxyl anthraquinone)	478-43-3	rs80120215	119009775	1	SPAG17	6.70E-09	0.042
Nifedipine	21829-25-4	rs148079596	74646587	13	KLF12	6.73E-09	0.017
Cetylpyridinium bromide	140-72-7	rs192455865	77187063	5	AP3B1	6.73E-09	0.030
1,8-Dihydroxy-4,5-dinitroanthraquinone	81-55-0	rs190416563	40716108	20	PTPRT	6.74E-09	0.003
1,3-Diiminobenz (f)-isoindoline	65558-69-2	rs7214324	75241094	17	SEC14L1	6.74E-09	0.003
Methyl mercuric (II) chloride	115-09-3	rs140638640	3157831	X	MXRA5	6.96E-09	0.059
1,3-Diiminobenz (f)-isoindoline	65558-69-2	rs114723141	71915908	18	CYB5A	7.06E-09	0.003
1,2-Epoxy-3-chloropropane (Epichlorohydrin)	3018-12-0	rs16958628	11860133	16	ZC3H7A	7.08E-09	0.004
dichloroacetonitrile	106-89-8	rs148108647	388033	7	FAM20C	7.08E-09	0.011
Captan	133-06-2	rs143336822	79139911	15	MORF4L1	7.10E-09	0.018
Catechol	120-80-9	rs75909882	186181894	3	CRYGS	7.16E-09	0.044
Saquinavir mesylate (AIDS Initiative)	149845-06-7	rs139792645	69525002	14	DCAF5	7.20E-09	0.006
Flutamide (pubertal study)	13311-84-7	rs75591162	179520663	3	PEX5L	7.30E-09	0.039
4-Chloro-3,5-dinitro-a,a,a-trifluorotoluene	393-75-9	rs140966476	1185683	19	STK11	7.32E-09	0.009
4-Chloro-o-phenylenediamine	95-83-0	rs114170015	8036026	3	GRM7	7.42E-09	0.023
Aldicarb	116-06-3	rs77630313	67951530	2	C1D	7.43E-09	0.046
2,3,5-Trichlorophenol	933-78-8	rs72743494	210497480	1	HHAT	7.50E-09	0.093
Retinol acetate	127-47-9	rs75201540	107697197	3	CD47	7.51E-09	0.008
Triamterene	396-01-0	rs61880875	7924360	11	OR10A6	7.54E-09	0.047
4-Chloro-3,5-dinitro-a,a,a-trifluorotoluene	393-75-9	rs146080764	49534816	11	FOLH1	7.58E-09	0.009
o-Nitrobenzyl chloride	612-23-7	rs113423597	53897175	4	SCFD2	7.59E-09	0.030
Vitamin D3	67-97-0	rs75048966	28220491	1	C1orf38	7.70E-09	0.004
1,1,1,2-Tetrabromoethane	630-16-0	rs190646915	1410097	X	CSF2RA	7.71E-09	0.004
Fumaronitrile	764-42-1	rs6575909	102802858	14	ZNF839	7.72E-09	0.022
t-Butyl formate	66-81-9	rs62389351	163302180	5	MAT2B	7.92E-09	0.003
Cycloheximide	762-75-4	rs116857952	69222338	6	BAI3	7.92E-09	0.098
Amiloride hydrochloride	2016-88-8	rs45477793	88470954	12	CEP290	7.98E-09	0.017

dichloroacetonitrile	3018-12-0	rs114922911	186024997	4	SLC25A4	8.06E-09	0.004
p-Benzoquinone dioxime	105-11-3	rs116331855	182768941	2	SSFA2	8.08E-09	0.025
Vitamin D3	67-97-0	rs78073324	63616070	1	FOXD3	8.16E-09	0.004
1,8-Dihydroxy-4,5-dinitroanthraquinone	81-55-0	rs113434813	143835690	X	SPANXN1	8.24E-09	0.003
Nitazoxanide	55981-09-4	rs112997343	74356986	17	SPHK1	8.26E-09	0.002
Ergotamine tartrate	55981-09-4	rs4263901	50758525	X	BMP15	8.39E-09	0.002

Table 3.6. SNP set pathway

Chemical	CAS #	Set	Name	No. Genes	Zscore ^a	FWER-controlled <i>P</i> ^b
1,1,1,2-Tetrabromoethane	630-16-0	KEGG	Allograft rejection	30	4.31	0.002
		KEGG	Graft-versus-host disease	31	4.06	0.003
		KEGG	Asthma	25	3.31	0.019
2,2',4'-Trichloroacetophenone	4252-78-2	KEGG	Autoimmune thyroid disease	45	3.71	0.007
HC blue 2	33229-34-4	KEGG	Butirosin and neomycin biosynthesis	5	3.51	0.008
13-cis-Retinal	472-86-6	KEGG	Butirosin and neomycin biosynthesis	5	3.67	0.010
N-(1,3-Dimethylbutyl)-N'-phenyl-p-phenylenediamine	793-24-8	KEGG	Asthma	25	3.49	0.010
		KEGG	Allograft rejection	30	3.47	0.011
Azathioprine	446-86-6	GO.MF	transforming growth factor beta receptor, pathway-specific cytoplasmic mediator activity	5	5.19	0.015
2,3,4,5-Tetrachlorophenol	4901-51-3	KEGG	Natural killer cell mediated cytotoxicity	131	3.36	0.016
		KEGG	Antigen processing and presentation	63	3.15	0.022
Ziram	137-30-4	GO.BP	Regulation of chronic inflammatory response	7	5.26	0.017
8-Hydroxyquinoline	148-24-3	KEGG	Mismatch repair	23	3.26	0.018
		KEGG	Steroid hormone biosynthesis	52	3.24	0.020
		KEGG	Porphyrin and chlorophyll metabolism	39	3.04	0.038
Chlordane (technical grade)	12789-03-6	GO.CC	Central element	5	4.34	0.034
Permethrin	52645-53-1	GO.BP	Regulation of interleukin-2 production	34	5.08	0.037

^aZ-score computed by the gene_set_scan software. ^bFamily-wise error controlled by resampling per chemical for each pathway type investigated.

Table 3.7. LASSO prediction accuracy of expression vs. EC10 for chemicals with proportion of variance explained $R^2 > 0.01$. Prediction accuracy of expression vs. EC₁₀.

Drug_Name	CAS #	R2
Iodochlorohydroxyquinoline	130-26-7	0.058
Dieldrin	60-57-1	0.055
Triamterene	396-01-0	0.052
2,3,5-Trichlorophenol	933-78-8	0.048
m-Nitrobenzyl chloride	619-23-8	0.042
Azathioprine	446-86-6	0.040
Dexamethazone	50-02-2	0.038
Cadmium acetate dihydrate	4-4-5743	0.038
Diethylene glycol diacrylate	4074-88-8	0.034
Chlordecone (kepone)	143-50-0	0.029
Potassium dichromate	7778-50-9	0.028
Dichlorvos (Vapona)	62-73-7	0.026
Ethidium bromide	1239-45-8	0.024
Daunomycin HCL	23541-50-6	0.023
Diisobutyl phthalate	84-69-5	0.022
Mercuric chloride	7487-94-7	0.021
HC blue 2	33229-34-4	0.021
Sodium dichromate dihydrate (VI)	7789-12-0	0.020
N,N-Dimethyl-p-nitrosoaniline	138-89-6	0.020
Methylene bis(thiocyanate)	6317-18-6	0.019
Hexachlorophene	70-30-4	0.019
Di(2-ethylhexyl) phthalate	117-81-7	0.019
Systhane	88671-89-0	0.018
3,4-Diaminotoluene	496-72-0	0.018
N-(1,3-Dimethylbutyl)-N'-phenyl-	793-24-8	0.018
Colchicine	64-86-8	0.016
Ethacrynic acid	58-54-8	0.015
Azobenzene	103-33-3	0.014
t-Butylhydroquinone	1948-33-0	0.014
6-Mercaptopurine monohydrate	6112-76-1	0.014
Tamoxifen citrate	54965-24-1	0.014
4,4-Thiobis(6-tert-butyl-m-creso	96-69-5	0.013
Zinc pyrithione	13463-41-7	0.011
Methacrylonitrile	126-98-7	0.011
Domiphen bromide	538-71-6	0.010

F. REFERENCES

- National Research Council. 2007. Toxicity testing in the 21st century: A vision and a strategy. Washington, DC: National Academies Press.
- Collins FS, Gray GM, Bucher JR. 2008. Toxicology. Transforming environmental health protection. *Science* 319(5865): 906-907.
- Zeise L, Bois FY, Chiu WA, Hattis D, Rusyn I, Guyton KZ. 2013. Addressing human variability in next-generation human health risk assessments of environmental chemicals. *Environ Health Perspect* 121(1): 23-31.
- Tice RR, Austin CP, Kavlock RJ, Bucher JR. 2013. Improving the human hazard characterization of chemicals: a Tox21 update. *Environ Health Perspect* 121(7): 756-765.
- Huang R, Southall N, Cho MH, Xia M, Inglese J, Austin CP. 2008. Characterization of diversity in toxicity mechanism using in vitro cytotoxicity assays in quantitative high throughput screening. *Chem Res Toxicol* 21(3): 659-667.
- Lock EF, Abdo N, Huang R, Xia M, Kosyk O, O'Shea SH, et al. 2012. Quantitative high-throughput screening for chemical toxicity in a population-based in vitro model. *Toxicol Sci* 126(2): 578-588.
- O'Shea SH, Schwarz J, Kosyk O, Ross PK, Ha MJ, Wright FA, et al. 2011. In vitro screening for population variability in chemical toxicity. *Toxicol Sci* 119(2): 398-407.
- Xia M, Huang R, Witt KL, Southall N, Fostel J, Cho MH, et al. 2008. Compound cytotoxicity profiling using quantitative high-throughput screening. *Environ Health Perspect* 116(3): 284-291.
- Ginsberg G, Smolenski S, Neafsey P, Hattis D, Walker K, Guyton KZ, et al. 2009. The influence of genetic polymorphisms on population variability in six xenobiotic-metabolizing enzymes. *J Toxicol Environ Health B Crit Rev* 12(5-6): 307-333.
- Wheeler HE, Dolan ME. 2012. Lymphoblastoid cell lines in pharmacogenomic discovery and clinical translation. *Pharmacogenomics* 13(1): 55-70.
- Choy E, Yelensky R, Bonakdar S, Plenge RM, Saxena R, De Jager PL, et al. 2008. Genetic analysis of human traits in vitro: drug response and gene expression in lymphoblastoid cell lines. *PLoS Genet* 4(11): e1000287.
- Gamazon ER, Huang RS, Cox NJ, Dolan ME. 2010. Chemotherapeutic drug susceptibility associated SNPs are enriched in expression quantitative trait loci. *Proc Natl Acad Sci USA* 107(20): 9287-9292.
- Brown CC, Havener TM, Medina MW, Krauss RM, McLeod HL, Motsinger-Reif AA. 2012a. Multivariate methods and software for association mapping in dose-response genome-wide association studies. *BioData mining* 5(1): 21.

- Brown CC, Havener TM, Medina MW, Jack JR, Krauss RM, McLeod HL, et al. 2014. Genome-wide association and pharmacological profiling of 29 anticancer agents using lymphoblastoid cell lines. *Pharmacogenomics* 15(2): 137-146.
- Baynes RE. 2012. Quantitative risk assessment methods for cancer and noncancer effects. *Progress in molecular biology and translational science* 112: 259-283.
- 1000 Genomes Project Consortium, Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, et al. 2012. An integrated map of genetic variation from 1,092 human genomes. *Nature* 491(7422): 56-65.
- Pillai K. 1955. Some new test criteria in multivariate analysis. *Ann Math Stat* 26: 117-121.
- Abecasis GR, Cherny SS, Cookson WO, Cardon LR. 2002. Merlin - rapid analysis of dense genetic maps using sparse gene flow trees. *Nat Genet* 30(1): 97-101.
- Almasy L, Blangero J. 1998. Multipoint Quantitative-Trait Linkage Analysis in General Pedigrees. *Am J Hum Genet* 62(5): 1198-1211.
- Yang J, Lee SH, Goddard ME, Visscher PM. 2011. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet* 88(1): 76-82.
- Lee S, Zou F, Wright FA. 2010. Convergence and Prediction of Principal Component Scores in High-Dimensional Settings. *Annals of statistics* 38(6): 3605-3629.
- Li Y, Willer CJ, Ding J, Scheet P, Abecasis GR. 2010. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet Epidemiol* 34(8): 816-834.
- Wheeler HE, Gamazon ER, Stark AL, O'Donnell PH, Gorsic LK, Huang RS, et al. 2013. Genome-wide meta-analysis identifies variants associated with platinating agent susceptibility across populations. *Pharmacogenomics J* 13(1): 35-43.
- Innocenti F, Mirkov S, Nagasubramanian R, Ramirez J, Liu W, Bleibel WK, et al. 2009. The Werner's syndrome 4330T>C (Cys1367Arg) gene variant does not affect the in vitro cytotoxicity of topoisomerase inhibitors and platinum compounds. *Cancer chemotherapy and pharmacology* 63(5): 881-887.
- Stark AL, Zhang W, Mi S, Duan S, O'Donnell PH, Huang RS, et al. 2010. Heritable and non-genetic factors as variables of pharmacologic phenotypes in lymphoblastoid cell lines. *Pharmacogenomics J* 10(6): 505-512.
- Aksoy P, Zhu MJ, Kalari KR, Moon I, Pelleymounter LL, Eckloff BW, et al. 2009. Cytosolic 5'-nucleotidase III (NT5C3): gene sequence variation and functional genomics. *Pharmacogenetics and genomics* 19(8): 567-576.
- O'Donnell PH, Gamazon E, Zhang W, Stark AL, Kistner-Griffin EO, Stephanie Huang R, et al. 2010. Population differences in platinum toxicity as a means to identify novel genetic susceptibility variants. *Pharmacogenetics and genomics* 20(5): 327-337.

- Li L, Fridley BL, Kalari K, Jenkins G, Batzler A, Weinshilboum RM, et al. 2009. Gemcitabine and arabinosylcytosin pharmacogenomics: genome-wide association and drug response biomarkers. *PloS one* 4(11): e7765.
- Peters EJ, Motsinger-Reif A, Havener TM, Everitt L, Hardison NE, Watson VG, et al. 2011. Pharmacogenomic characterization of US FDA-approved cytotoxic drugs. *Pharmacogenomics* 12(10): 1407-1415.
- Huang RS, Gamazon ER, Ziliak D, Wen Y, Im HK, Zhang W, et al. 2011. Population differences in microRNA expression and biological implications. *RNA biology* 8(4): 692-701.
- Wheeler HE, Gorsic LK, Welsh M, Stark AL, Gamazon ER, Cox NJ, et al. 2011. Genome-wide local ancestry approach identifies genes and variants associated with chemotherapeutic susceptibility in African Americans. *PloS one* 6(7): e21920.
- Kulkarni H, Goring HH, Diego V, Cole S, Walder KR, Collier GR, et al. 2012. Association of differential gene expression with imatinib mesylate and omacetaxine mepesuccinate toxicity in lymphoblastoid cell lines. *BMC medical genomics* 5: 37.
- Brown CC, Havener TM, Medina MW, Auman JT, Mangravite LM, Krauss RM, et al. 2012b. A genome-wide association analysis of temozolomide response using lymphoblastoid cell lines shows a clinically relevant association with MGMT. *Pharmacogenetics and genomics* 22(11): 796-802.
- Schaid DJ, Sinnwell JP, Jenkins GD, McDonnell SK, Ingle JN, Kubo M, et al. 2012. Using the gene ontology to scan multilevel gene sets for associations in genome wide association studies. *Genet Epidemiol* 36(1): 3-16.
- Gamazon ER, Skol AD, Perera MA. 2012. The limits of genome-wide methods for pharmacogenomic testing. *Pharmacogenetics and genomics* 22(4): 261-272.
- van den Oord EJ, Sullivan PF. 2003. False discoveries and models for gene discovery. *Trends Genet* 19(10): 537-542.
- Halterman JA, Kwon HM, Wamhoff BR. 2012. Tonicity-independent regulation of the osmosensitive transcription factor TonEBP (NFAT5). *American journal of physiology Cell physiology* 302(1): C1-8.
- Njiaju UO, Gamazon ER, Gorsic LK, Delaney SM, Wheeler HE, Im HK, et al. 2012. Whole-genome studies identify solute carrier transporters in cellular susceptibility to paclitaxel. *Pharmacogenetics and genomics* 22(7): 498-507.
- Devlin B, Roeder K. 1999. Genomic control for association studies. *Biometrics* 55(4): 997-1004.
- Storey JD, Tibshirani R. 2003. Statistical significance for genomewide studies. *Proc Natl Acad Sci USA* 100(16): 9440-9445.

- Setou M, Seog DH, Tanaka Y, Kanai Y, Takei Y, Kawagishi M, et al. 2002. Glutamate-receptor-interacting protein GRIP1 directly steers kinesin to dendrites. *Nature* 417(6884): 83-87.
- Yamada K, Ono M, Perkins ND, Rocha S, Lamond AI. 2013. Identification and functional characterization of FMN2, a regulator of the cyclin-dependent kinase inhibitor p21. *Mol Cell* 49(5): 922-933.
- Sun P, Xia S, Lal B, Eberhart CG, Quinones-Hinojosa A, Maciaczyk J, et al. 2009. DNER, an epigenetically modulated gene, regulates glioblastoma-derived neurosphere cell differentiation and tumor propagation. *Stem Cells* 27(7): 1473-1486.
- Chan DW, Lee JM, Chan PC, Ng IO. 2008. Genetic and epigenetic inactivation of T-cadherin in human hepatocellular carcinoma cells. *Int J Cancer* 123(5): 1043-1052.
- Huang Y, Dai Z, Barbacioru C, Sadee W. 2005. Cystine-glutamate transporter SLC7A11 in cancer chemosensitivity and chemoresistance. *Cancer Res* 65(16): 7446-7454.
- Lappalainen T, Sammeth M, Friedlander MR, t Hoen PA, Monlong J, Rivas MA, et al. 2013. Transcriptome and genome sequencing uncovers functional variation in humans. *Nature* 501(7468): 506-511.
- Dennis G, Jr., Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, et al. 2003. DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol* 4(5): 3.
- Gatti DM, Barry WT, Nobel AB, Rusyn I, Wright FA. 2010. Heading down the wrong pathway: on the influence of correlation within gene sets. *BMC genomics* 11: 574.
- Zhou YH, Barry WT, Wright FA. 2013. Empirical pathway analysis, without permutation. *Biostatistics* 14(3): 573-585.
- Goldstein DB. 2009. Common genetic variation and human traits. *N Engl J Med* 360(17): 1696-1698.
- DeGorter MK, Xia CQ, Yang JJ, Kim RB. 2012. Drug transporters in drug efficacy and toxicity. *Annu Rev Pharmacol Toxicol* 52: 249-273.
- Zhang S, Lovejoy KS, Shima JE, Lagpacan LL, Shu Y, Lapuk A, et al. 2006. Organic cation transporters are determinants of oxaliplatin cytotoxicity. *Cancer Res* 66(17): 8847-8857.
- Marce S, Molina-Arcas M, Villamor N, Casado FJ, Campo E, Pastor-Anglada M, et al. 2006. Expression of human equilibrative nucleoside transporter 1 (hENT1) and its correlation with gemcitabine uptake and cytotoxicity in mantle cell lymphoma. *Haematologica* 91(7): 895-902.
- Lo M, Wang YZ, Gout PW. 2008. The x(c)- cystine/glutamate antiporter: a potential target for therapy of cancer and other diseases. *J Cell Physiol* 215(3): 593-602.

Pham AN, Blower PE, Alvarado O, Ravula R, Gout PW, Huang Y. 2010. Pharmacogenomic approach reveals a role for the x(c)- cystine/glutamate antiporter in growth and celestrol resistance of glioma cell lines. *J Pharmacol Exp Ther* 332(3): 949-958.

Januchowski R, Zawierucha P, Andrzejewska M, Rucinski M, Zabel M. 2013. Microarray-based detection and expression analysis of ABC and SLC transporters in drug-resistant ovarian cancer cell lines. *Biomedicine & pharmacotherapy = Biomedecine & pharmacotherapie* 67(3): 240-245.

Motaghed M, Al-Hassan FM, Hamid SS. 2014. Thymoquinone regulates gene expression levels in the estrogen metabolic and interferon pathways in MCF7 breast cancer cells. *Int J Mol Med* 33(1): 8-16.

Kinoshita H, Okabe H, Beppu T, Chikamoto A, Hayashi H, Imai K, et al. 2013. Cystine/glutamic acid transporter is a novel marker for predicting poor survival in patients with hepatocellular carcinoma. *Oncology reports* 29(2): 685-689.

Drayton RM, Dudzic E, Peter S, Bertz S, Hartmann A, Bryant HE, et al. 2014. Reduced Expression of miRNA-27a Modulates Cisplatin Resistance in Bladder Cancer by Targeting the Cystine/Glutamate Exchanger SLC7A11. *Clin Cancer Res* 20(7): 1990-2000.

Rusyn I, Daston GP. 2010. Computational toxicology: realizing the promise of the toxicity testing in the 21st century. *Environ Health Perspect* 118(8): 1047-1050.

CHAPTER 4: IN VITRO SCREENING FOR INTER-INDIVIDUAL AND POPULATION VARIABILITY IN TOXICITY OF PESTICIDE MIXTURES

A. ABSTRACT

Population-based human *in vitro* models offer exceptional opportunities for evaluating the potential hazard and mode of action of chemicals, as well as variability in response. Challenges remain that require further assessment to increase the utility of the information obtained from *in vitro* models. This study was designed to address the potential challenges of screening and assessing the cytotoxicity of complex mixtures. We selected 146 lymphoblast cell lines from 4 ancestrally and geographically diverse populations based on the availability of genome sequence and basal RNA-seq data. Cells were exposed to two pesticide mixtures (an organochlorine pesticide environmental mixture extracted from a passive surface water sampling device, and a mixture of 36 currently used pesticides) at 8 concentrations and were then evaluated for cytotoxicity. qHTS screening in the genetically-defined populations produced robust and reproducible results. On average, the two mixtures exhibited a similar range of *in vitro* cytotoxicity and showed considerable inter-individual variability across the screened cell lines. However, *in vitro*-to-*in vivo* extrapolation (IVIVE), which was performed by reverse pharmacokinetics, suggested a significantly lower oral equivalent dose for the chlorinated pesticide mixture compared to the current-use pesticide mixture. Multivariate genome-wide association mapping revealed an association between the current use-pesticide mixture and a

polymorphism in rs1947825 in C17orf54. Moreover, genetic pathway analysis showed a significant association between metabolism pathways and the cytotoxicity of the chlorinated pesticide mixture. We concluded that, together with IVIVE, an efficient *in vitro* experimental design that incorporates population variability and comparative population genomics can effectively enable the quantification of human health hazard in the most sensitive individuals to environmental mixtures. Additionally, such approaches can lead to generation of testable hypotheses regarding potential toxicity mechanisms.

B. INTRODUCTION

Pesticides are compounds that are used to kill, repel, or control certain forms of plant or animal life that are considered to be pests (Krieger 2011). Adverse health effects of pesticides can range from mild skin and mucous membrane irritation to more severe outcomes such as neurotoxicity and cancer (Rother 2014; Bassil et al. 2007; Sanborn et al. 2007). Moreover, sensitivity to exposure is higher among relatively vulnerable populations, including women, children, the elderly, the immune-compromised and the malnourished (Perry et al. 2014; Jurewicz and Hanke 2008). There are several challenges in the evaluation of the human health hazard of pesticides. First, pesticides have variable Modes of Action (MOA) dependent on use and activity, and are meant to be harmful and toxic to pests, but not humans. Second, because they are widely used in agricultural and household settings, people are frequently exposed to pesticide residues. Third, pesticides can be dispersed as mixtures, creating complexity in hazard evaluation (Manikkam et al. 2012).

While safety testing of the individual pesticides is conducted according to established regulatory guidelines, evaluation of the toxicity of mixtures is less structured (OPP, 2002). The cumulative risk assessment approach is conducted for individual chemicals with common mechanisms of toxicity, even though the data is usually available only for individual chemicals (OPP 2002). Indeed, current toxicity testing paradigms have been questioned for their failure to consider commonly occurring co-exposures to environmental agents and the magnitude of human population variability in response to chemicals (National Research Council 2009).

Whole animal testing is difficult to employ for testing chemical mixtures. In contrast, *in vitro* testing could allow grouping of chemicals according to their effects on key biologic pathways or their real human co-exposures over a broad range of concentrations in a rapid and

inexpensive manner (Andersen and Krewski 2009). The resulting data could enable an informed and focused approach to the problem of assessing risk in human populations exposed to mixtures. Furthermore, with an experimental *in vitro* design that represents a human population, we are allowed to not only explore the hazard, but also the intrinsic variability that is associated with it across different dose ranges (Lock et al. 2012; O'Shea et al. 2010). Such information would be valuable to inform regulatory decisions that could more fully protect public health and sensitive subpopulations (National Research Council 2009).

In the present study, we addressed the hypothesis that comparative population genomics with efficient *in vitro* experimental design can be used for evaluation of the potential for hazard, mode of action, and the extent of population variability in responses to chemical mixtures. We screened 146 lymphoblast cell lines (LCLs) from four ancestrally and geographically diverse populations with publicly available genotypes and sequencing data from the 1000 Genomes Project (1000 Genomes Project Consortium 2010). Cells were exposed to two pesticide mixtures (organochlorine pesticide environmental mixture extracted from a passive surface water sampling device, and a mixture of 36 currently used pesticides) at 8 concentrations. Cell viability was evaluated with CellTiter-Glo® (Promega) assay which evaluates ATP production in a 96-well plate format. Cytotoxic response was assessed using an effective concentration threshold of 10% (EC₁₀) (Abdo et al., in preparation), designed to be relevant to the dose-response evaluation commonly used in quantitative risk assessment practice and to meaningfully capture ranges of variation in response across chemicals. Genome-wide association mapping, gene set scan, and pathway analyses were performed to evaluate the genetic determinants of susceptibility. Furthermore, *in vitro*-to-*in-vivo* extrapolation by reverse pharmacokinetics and real cumulative exposures were utilized to predict xenobiotic steady state pharmacokinetics.

C. MATERIALS AND METHODS

Experimental Design

Cell lines. A set of 146 immortalized LCLs was acquired from Coriell Cell Repositories (Camden, NJ). The 146 cell lines represent 4 ancestrally and geographically diverse populations: Utah residents with Northern & European ancestry (CEU); Tuscan in Italy (TSI); Yoruban in Ibadan, Nigeria (YRI); and British from England & Scotland (GBR) (see table 4.1 for number and percent of each population and gender). The cell lines were chosen based on the availability of dense genotyping information and RNA-Seq expression data. Screening was conducted in two batches, and cell lines were randomly divided into the batches without regard to family structure but with equal representation of population and gender in each batch. Cells were suspended in flasks in an upright position in RPMI 1640 media (Gibco, Carlsbad, CA) supplemented with 15% fetal bovine serum (HyClone, South Logan, UT) and 1% penicillin-streptomycin (Gibco) and cultured at 37°C with 5% CO₂. Media was changed every 3 days. Cell count and viability were assessed once a day for five days for all cell lines using Cellometer Auto T4 Plus (Nexcelom Bioscience, Lawrence, MA). Cells were grown to a concentration up to 10⁶ cells/ml, volume of at least 100 ml, and viability of > 85% before treatment. After centrifugation, the cells were resuspended in fresh media. Cells (100 µl containing 10⁴ cells) were aliquoted to each well in a 96-well treatment plate (following the addition of the chemicals) and mixed using the Biomek 3000 robot. Plates were incubated for 24 h after treatment at 37°C and 0.5% CO₂. To increase the robustness of the data and to evaluate reproducibility, each cell line was seeded in at least two plates so that each compound would be screened in each cell line on 2 or more plates.

Chemical Mixtures. Cells were exposed to two extracts of environmental chemical mixtures: the first mixture, Chlorinated Pesticide Mixture (CPM), a real environmental sample obtained from a

universal passive sampling device deployed for 30 days in surface water next to a chlorinated pesticide storage facility (10 pesticides were present in detectable quantities in the post-collection laboratory analysis) (see table 4.2 for complete list of pesticide chemicals identified by mass spectrometry); and the second mixture, Current-Use Pesticide Mixture (CUPM), being a mixture of 36 currently used pesticides that mimics real exposure amounts of eastern North Carolina (see table 4.2 for complete list of pesticide chemicals). Chemicals were dissolved with dimethyl sulfoxide (DMSO) into 8 different stock concentrations. Final concentrations ranged from 0.032 to 370.4 μM for current-use pesticide mixture and from 0.022 to 65.7 μM for chlorinated pesticide mixture. The mixtures were aliquoted to 96-well plate format using the Biomek 3000 robot. The negative control was DMSO at 0.5% vol/vol; the positive control was tetra-octyl ammonium bromide at 46 μM .

Cytotoxicity profiling. The CellTiter-Glo Luminescent Cell Viability (Promega Corporation, Madison, WI) assay was used to assess intracellular ATP concentration, a marker for cytotoxicity, 40 h post treatment. Time points were selected based on previous experiments at the National Institutes of Health Chemical Genomics Center (Xia et al. 2008). A ViewLux plate reader (PerkinElmer, Shelton, CT) was used to detect luminescent intensity in each well of the assay plates.

Data Processing

Cytotoxicity EC_{10} estimation and outlier detection. Cytotoxicity data were normalized relative to positive/negative controls as described elsewhere (Xia et al. 2008). We derived an effective concentration 10th percentile (EC_{10}) to provide a single cytotoxicity dose summary per chemical and cell line. The derivation of EC_{10} was based on the logistic model:

$$\log\left(\frac{y - \theta_{min}}{\theta_{max} - \theta_{min}}\right) = \beta_0 + \beta_1 d + \varepsilon,$$

with $\varepsilon \sim N(0, \sigma^2)$, where y is the observed normalized signal representing proportion of surviving cells (which we term the “cytotoxicity value”), d is the log(concentration) for each chemical, and θ_{max} is the mean cytotoxicity value on the logit scale for the zero concentration. θ_{min} was set to zero, to avoid difficulties in estimating the minimum cytotoxicity value for chemicals with low cytotoxicity. An exception was made for chemicals in which the cytotoxicity value at the highest concentration was higher than 0.4, as a very few number of plates/chemicals did not reliably reach maximum cytotoxicity. In those instances the cytotoxicity value was set at the observed cytotoxicity at the maximum concentration. Inspection of these data revealed good fits in such instances. Although in principle θ_{max} should have been 1.0, a number of plates exhibited a drift from this value, and thus the parameter was estimated from the data.

Fitting for the parameters $[\beta_0, \beta_1, \sigma^2, \theta_{max}]$ proceeded by maximum likelihood using numerical optimization in *R* v2.15. An automatic outlier detection algorithm was devised by considering the impact of dropping each concentration value in succession, and removing those values for which the maximum likelihood improved by a factor of 10 or more and refitting the model using the non-outlying observations.

Normalizing batch effects. Batch effects were evaluated by running principal component analysis. EC_{10} values were adjusted for batch effect using the ComBat method (Johnson et al. 2007).

Concentration response for populations and individuals. For each pesticide mixture, the three-parametric logistic regression described above in EC_{10} estimation was fit to concentration-response data for each cell line. The variation in the EC_{10} estimates was used as illustrative of population variation in true EC_{10} values, although additional sampling variation underlies each

EC₁₀ estimate. An overall logistic concentration-response curve was fit to the aggregated data across all individuals. See Figure 4.1

Reproducibility and correlation between mixtures. Pearson and Spearman correlation coefficients (r) between pairs of replicate plates were used to assess experimental reproducibility and the correlation between the two mixtures. For this analysis, the two replicate plates were selected for each mixture and cell line pair. See Figure 4.1

Chemical/Mixture Specific Adjustment Factor (CSAF). Variability in response for each mixture across the 146 cell lines was obtained by obtaining the ratio of the 50th percentile to the 5th percentile. The WHO CSAF guideline for toxicokinetics study is to obtain the ratio of the 95th to the 50th percentile as the uncertainty factor for human variability in toxicokinetics. Both of our mixtures were skewed to the left. To be on the conservative side, we obtained the uncertainty factor for human variability in toxicodynamics as 50thpercentile/5thpercentile or the bigger tail (WHO 2005).

Chemical descriptors. Chemical descriptors were calculated using Dragon version 5.5. Constant and near constant descriptors as well as highly correlated descriptors were excluded and descriptor values were normalized on a scale from 0 to 1.

Differences in cytotoxicity across different populations. *Analysis of Variance (ANOVA)* was performed to assess population differences in cytotoxicity between the four screened populations for each mixture. See Figure 4.2.

Genotypes. The primary source of genotypes was obtained as described in Abdo et al., 2014. SNPs with a call rate below 99%, minor allele frequency (MAF)<0.05, or HWE p-value<1X10⁻³ were excluded.

Multivariate Association Analysis (MAGWAS). The MAGWAS multivariate analysis of covariance model (Brown et al. 2012) was used for primary association mapping (http://www4.stat.ncsu.edu/~motsinger/Lab_Website/Software.html). The approach allows for use of the full concentration-response profile, as opposed to a univariate summary (such as EC₁₀) as a single response, with the advantage of robustness and power under a wide variety of association patterns (Pillai, 1955). The model used for association for the j th individual and genotype i for the chemical/SNP was

$$Y_{ij} = X_{ij}\beta + \mu_i + e_{ij}$$
$$e_{ij} \sim N(\mathbf{0}, \Sigma),$$

where Y_{ij} is the vector of responses (across the eight concentrations) for the j^{th} individual having genotype i , X_{ij} is the design matrix of covariates, including sex, indicator variables for laboratory batch, and the first ten genotype principal components, and μ_i is the eight-vector of parameters modeling the effects of genotype i on the response. The model assumes that the error terms are multivariate normally distributed, with mean vector $\mathbf{0}$ and variance-covariance matrix Σ , allowing for dependencies in the observations. P -values were obtained using Pillai's trace (Pillai 1955). Because this method makes use of asymptotic theory, markers with fewer than 20 individuals representing any genotype were removed, leaving 692,013 SNPs for analysis.

In vitro to in vivo extrapolation. *In vitro* Pharmacokinetic Assays were applied to chemicals as described previously (Wetmore et al. 2012). Plasma protein binding was determined for each chemical using the rapid equilibrium dialysis (RED) method (Wetmore et al. 2012). The rate of hepatic metabolism of the parent compound was determined using the substrate depletion

approach as previously (Wetmore et al. 2012; Rotroff et al. 2010).

Estimation of C_{ss} using In Vitro-to-In Vivo Extrapolation (IVIVE) and Monte Carlo Simulation.

The chemical steady-state blood concentrations (C_{ss}) were estimated as previously described (Wetmore et al. 2012) (Wetmore et al. in preparation) with modification. The base equation used to calculate static C_{ss} is based on constant uptake of a daily oral dose and factors in hepatic clearance and non-metabolic renal clearance:

$$C_{ss} = \frac{ko}{\left(\frac{(Q_H \times F_{ub} \times Cl_{intH})}{(Q_H + F_{ub} \times Cl_{intH})}\right) + (GFR \times F_{ub})}$$

where ko = chemical exposure rate set to 0.042 µg/kg/hr (i.e., 1 µg/kg/day); Q_H = hepatic blood flow (90 L/hr; Davies and Morris, 1993), F_{ub} = unbound fraction of parent compound in the blood; Cl_{intH} = hepatic intrinsic metabolic clearance; and GFR = glomerular filtration rate. The F_{ub} was calculated based on the experimentally measured F_{u plasma} divided by the blood:plasma ratio (B:P). The right side of the denominator considers non-metabolic renal clearance (GFR × F_{ub}), with GFR (6.7 L/hr) back calculated based on the serum creatinine Cockcroft-Gault equation (Cockcroft and Gault 1976). The CL_{int} values were derived using the following equation, which scales CL_{u in vitro} (µL/min/million cells) experimentally measured in hepatocytes to represent whole organ clearance with units of L/hr:

$$Cl_{int} = Cl_{u\ in\ vitro} \times HPGL \times V_L \times \frac{L}{10^6\ \mu l} \times \frac{60\ min}{1\ \square}$$

Where HPGL = hepatocytes per gram liver 110 million cells per g liver; (Barter et al. 2007) (HPGL) and V_L = liver volume (1596 g; (Johnson et al. 2005).

A correlated Monte Carlo approach was employed (Jamei et al. 2009) using Simcyp (Simcyp V. 13; Certara, Sheffield, UK) to simulate variability across a population of 10,000 individuals equally comprised of both genders, 20-50 years of age. A coefficient of variation of 30% was used for intrinsic and renal clearance. The median, upper and lower fifth percentiles for the C_{ss} were obtained as output.

Calculation of oral equivalent dose values. In conventional use, pharmacokinetic models are used to relate exposure concentrations to a blood or tissue concentration. This is typically referred to as “forward dosimetry.” In contrast, the models can also be reversed to relate blood or tissue concentrations to an exposure concentration, which is referred to as “reverse dosimetry” (Tan et al. 2007). Based on the principal of reverse dosimetry, the median, upper and lower 5th percentiles for the C_{ss} were used as conversion factors to generate oral equivalent doses according to the following formula:

$$\text{Oral equivalent} \left(\frac{\mu\text{g}}{\text{kg}} \right) = [EC]_{10} \text{ (nM)} \cdot ((1 \text{ mg/kg})/\text{day}) / (C_{ss} \text{ (nM)})$$

In the equation above, the oral equivalent dose value is linearly related to the *in vitro* EC_{10} and inversely related to C_{ss} . This equation is valid only for first-order metabolism that is expected at ambient exposure levels. An oral equivalent value was generated for each chemical-cell line combination and summed to provide a cumulative oral equivalent value for each cell line.

Predicted exposure limits. Pesticide specific predicted exposure limits were obtained as previously detailed in (Wambaugh et al. 2013). The pesticide specific exposure limit was available for 35 out of the 36 pesticides in the current-use pesticide mixture and for 6 out of 10 pesticides in the chlorinated pesticide mixture. Missing values were replaced by the highest exposure within each mixture. Then, a cumulative exposure was computed for each mixture from the upper 95th percentile. See flow chart in Figure 4.3.

D. RESULTS

Cytotoxicity of pesticide mixtures in vitro.

Screening was conducted in a 96-well plate format using a robotic system. The 146 cell lines were randomly assigned to two batches with gender and population blocking. Each cell line was dispensed by the robot to two separate plates where both pesticide mixtures were dispensed at 8 different concentrations ranging from (0.032 to 370.4 μM) for current-use pesticide mixture and from (0.022 to 65.7 μM) for chlorinated pesticide mixture. Positive and negative controls were aliquoted to the same plate. Normalization to the control for each plate was performed as described in the Materials and Methods section for each cell line separately. EC_{10} s were derived on \log_{10} scale as described in materials and methods, batch-corrected and averaged across replicated for each cell line.

The availability of cytotoxicity screening on 146 individuals, with the assay performed under controlled conditions, enables sensitive investigation of variation in individual dose-response profiles (National Research Council 2009). To visualize “individual” vs. “population” response to each pesticide mixtures, we fitted our 3-parametric logistic regression, described in the materials and methods, to each cell line’s concentration-response illustrated in Figure 4.1a and 1b corresponding to chlorinated- and current-used- pesticide mixtures respectively. For each concentration-response EC_{10} was estimated and shown as the inset red histogram to show the populations variability in response to each mixture. The mean of these EC_{10} values offers a population-wide summary of the cytotoxicity, of a mixture and is very similar to the EC_{10} produced when the data are first pooled for all individuals and then fit using a single concentration-response curve (red-dashed curve in Figure 4.1a and 1b). Both mixtures demonstrated considerable inter-individual variability in cytotoxicity response (Figure 4.1). However, aggregation across the population ignores the variability in toxic susceptibility and the

variability across replicates, and the EC₁₀ estimated fifth percentile may be used to illustrate the concept of a “vulnerable” subpopulation.

To increase the robustness of the cytotoxicity measurement, duplicate plates were run. To evaluate the reproducibility of our EC₁₀ estimates, pair-wise Pearson and Spearman correlations among replicate plate pairs using log₁₀(EC₁₀) values for each mixture were calculated. Highly significant correlations were seen in each pesticide mixture ($p < 0.0001$). For current pesticides mixtures $r[\text{Pearson's}] = 0.62$ and $r[\text{Spearman}] = 0.55$ (See Figure 4.1). For chlorinated pesticides mixture, $r[\text{Pearson's}] = 0.65$ and $r[\text{Spearman}] = 0.56$ (see Figure 4.1). Overall reproducibility for both mixtures was also significant ($p < 0.0001$) with $r[\text{Pearson's}] = 0.62$ and $r[\text{Spearman}] = 0.54$. Overall, Duplicate measures revealed excellent experimental reproducibility (See Figure 4.1).

Comparative analysis of cytotoxicity of pesticide mixtures using population-based model.

Both the mean and the median cytotoxicity (EC₁₀) for current use pesticide mixture were slightly lower than chlorinated pesticide mixture (see Figure 4.4a and 4.4b and table 4.4 for summary statistics) indicating slightly more potency for current-used pesticide mixture. The median EC₁₀ for current-use pesticide mixture was almost 12 μM and 13 μM for chlorinated pesticide mixture. However, there was no significant difference in their mean cytotoxicity between the two mixtures. Interestingly, the current pesticide mixture demonstrated a slightly wider distribution across the population than chlorinated pesticide mixture (see Figure 4.4a and 4.4b). The toxicodynamic uncertainty factor (UFD) for human variability was around 3 fold for each mixture (see table 4.4).

In order to translate our cytotoxicity measures into a meaningful potential health risk hazard, we computed oral equivalent doses for both mixtures using reverse toxicokinetics

approach (Wetmore et al. 2012; Rotroff et al. 2010). In vitro pharmacokinetic data (see complete list at: http://comptox.unc.edu/Toxcast_I_II_RTK_List.pdf) were available for 31 of the 36 chemicals present in the current use pesticides laboratory mixture, and for 4 of the 10 chemicals in the chlorinated pesticide environmental mixture. Review of the C_{ss} values predicted for the 31 current use pesticides revealed a similar distribution as observed across the 180 ToxCast Phase II chemicals similarly assessed for in vitro PK: a median C_{ss} $<1 \mu\text{M}$; 95th percentile $\approx 200 \mu\text{M}$. Two of the chemicals assessed had very high C_{ss} values: ethalfluralin ($350 \mu\text{M}$) and flumetralin ($277 \mu\text{M}$); the rest were below $8 \mu\text{M}$. The distribution of data for 4 chlorinated pesticides was different from the laboratory mixture: the max C_{ss} value estimated = $58.46 \mu\text{M}$. Since there are no standard PK approach for dealing with mixtures, we assumed for this analysis that there is not interaction or potentiation between chemicals in terms of toxicity and pharmacokinetics (PK) modeling. For the purposes of this work, for the PK modeling, we assumed that the PK of each chemical will not be significantly impacted by the presence of other chemicals present at these low levels anticipated in the environment. Given this, we expect that the *in vitro* PK that was derived using clearance and plasma protein binding data measured for individual chemicals will provide an adequate estimate of PK behavior for this assessment. Our data involved only one assay measured across 146 individuals, and so a separate oral equivalent was calculated for each individual based on the percentage of chemical in the mixture. To be conservative, missing *in vitro* pharmacokinetic data were assigned a value based on the most conservative simulation assuming no hepatic clearance, high blood binding and only renal clearance, which we defined as the “worst case scenario” (See Figure 4.4C).

The oral equivalent (OE) dose was computed based on four different scenarios in which we substituted the missing C_{ss} with either the median C_{ss} of known chemicals or worst-case-

scenario value, where OE was calculated with and without weighting of the EC₁₀ by the percentage of chemical in the mixture (Figure 4.5) for an illustrative flowchart of OE calculations). Supplemental Figure 4.6 shows OE doses for each scenario for each pesticide mixture across the 146 cell lines.

The simulations were run using Simcyp software (Simcyp Ltd, 2001), which incorporates Monte Carlo Simulation to capture population variability in PK. The C_{ss} values were derived using a population of healthy volunteers (Northern European, 20-50 years of age, equally mixed gender). The simulation was run using 10 trials, 1000 volunteers per trial. The upper 95th percentile values were used to determine the oral equivalents (Figure 4.4C), as this approach results in a reasonably conservative value. Across the four different scenarios, these *in vitro*-to-*in vivo* extrapolation data show that a significantly ($p < 0.001$) lower dose of chlorinated pesticide would lead to the internal concentrations that are equal to the EC₁₀ values that elicited cytotoxicity (See Figure 4.4C and Figure 4.6). Oral Equivalent doses for both mixtures were not substantially altered (<0.5 fold difference) when the median C_{ss} value was used instead of the worst case scenario. However, oral equivalent dose was remarkably shifted for both mixtures (>1.2 fold change) with non-weighted EC₁₀ vs. EC₁₀ weighted by the percentage of chemicals used in each mixture (See Figure 4.6). However, the relation between the two mixtures were maintained in all scenarios, in which the chlorinated pesticide mixture was more toxic the current use pesticide mixture.

To better evaluate the human health risk of exposure to such mixtures, we examined the relationship of our calculated oral-dose-equivalent with actual real-human-exposures to such mixtures. We computed a cumulative exposure value for each mixture based on individual exposure estimates for each chemical obtained from ExpoCast (Wambaugh et al. 2013). The

ExpoCast framework has created an estimated human exposure potential for 1936 chemicals (Wambaugh et al. 2013), which incorporated both biomonitoring data and uncertainty factors in their predictions. Predicted estimates were available for 35 of the 36 chemicals present in the current use pesticides laboratory mixture, and 6 of the 10 chemicals in the chlorinated pesticide environmental mixture. We presumed the highest predicted exposure for chemicals with no predicted exposure value and calculated the cumulative exposure for each mixture from the upper 95th percentile to be conservative. The actual real human exposure estimates were lower than our calculated oral dose equivalent indicating no real human hazard (See Figure 4.4c). Our oral equivalent dose was ~1-fold higher for current-use pesticide mixture and ~6 fold higher for chlorinated pesticide mixture than their corresponding real human exposure estimate (See Table 4.6 and Table 4.7). This indicates a wider margin of safety for the chlorinated pesticide mixture than the current-use pesticide mixture.

Similarities between pesticide mixtures.

We wanted to assess the similarity of cytotoxic response for the pesticide mixtures across different cell lines. A significant correlation (Spearman $r=0.25$ $p<0.01$) was observed between the two mixtures, illustrating concordance in individual cell line's responses to both tested mixtures (see Figure 4.7a). These findings are interesting, considering that none of the individual chemicals appear in both mixtures. Furthermore, the results might suggest potential shared mechanisms of susceptibility to toxicity. There were no suggestive patterns of population clustering in the correlation between the two mixtures. To make sure that the significant correlation was not substantially influenced by outlying points, we removed the three most

outlying points and rederived the correlation. The correlation still remained significant ($p < 0.05$, spearman $r = 0.2$).

None of the individual chemicals within each mixture overlapped with the other mixture, but there was a significant correlation between cytotoxic responses to the mixtures, we investigated similarities in their chemical structure. Principal Component Analysis (PCA) (Hotelling 1933) showed great similarities between single chemical compounds in both mixtures in their chemical descriptor space (see Figure 4.7b). This observation may partially explain the significant correlation between the two mixtures and the lack of significant difference in their mean cytotoxicity. We explored the potential of chemical descriptors in clustering the chemicals within each mixture by their pesticide mode of action (See Table 4.2 and 4.3 for a complete list of pesticide MOA (Wood 2014). We did not find any meaningful clusters. Only 4 of 5 chemical that acted as nematicides clustered together in the 1st principal component (see Figure 4.7b).

We were further interested in assessing the strength of the correlation between the two mixtures when compared to any other correlation between two random chemicals. Previously, we had screened 1086 LCLs with 179 diverse chemicals (Abdo et al, in preparation). The Spearman correlation between the two pesticide mixtures ($r = 0.25$) was slightly above the median correlation of a randomly picked correlation chosen from 15931 possible correlations in the previous cytotoxicity experiment (see Figure 4.7c).

Differences in cytotoxicity across ancestral populations.

The regulation of pesticides is a high priority in many countries due to their widespread use. Therefore, investigating susceptible sub-populations is of great interest. The current use pesticide mixture exhibited marginally significant differences among the populations tested ($p =$

0.058), while chlorinated pesticide mixture did not show any significant differences (see Figure 4.2). Interestingly, GBR (British from England & Scotland) population was the most sensitive population, while YRI (Yoruban in Ibadan, Nigeria) population appeared to be the least sensitive in both mixtures. Moreover, within-population variability was larger within the current use pesticide mixture compared to the chlorinated pesticide mixture, especially when comparing the range of the upper quartile to the lower quartile within each population (see Figure 4.2).

Susceptibility loci, genes, and pathways.

The availability of densely genotyped data from the 1000 Genomes Project for the LCLs used in this study allows for exploration of possible genetic determinants of cytotoxicity. Despite being relatively underpowered, with a sample size of 146, in comparison with modern disease genome-wide association studies, we attempted to identify possible loci, genes, or pathways associated with cytotoxicity. After careful frequency and genetic pruning and quality control (SNPs with a call rate below 99%, minor allele frequency (MAF) < 0.05, or HWE p-value < 1×10^{-3} were excluded), 1,015,304 SNPs were included in the analysis. Concentration-responses were subjected to quality control as described in (Abdo et al., in preparation), in which outliers were removed and subjected to fitted smoothing as described in (Abdo et al., in preparation). Sex, experimental batch and date, population, and the first ten genotype principal components were included as covariates for running Multivariate ANCOVA Genome-Wide Association Software (MAGWAS) on the curated concentration-responses. MAGWAS is a sensitive method in identifying any pattern of variation of cytotoxicity measurements due to genotype (Brown et al. 2012). Even though in this study we used a relatively small population of 146 cell lines, we were able to observe a finding of genome-wide suggestive significance. The cytotoxicity

measurement of current used pesticide was associated ($p < 6.5e^{-08}$) with a locus on Chr17 (Figure 4.8a and 4.8b), which is near the commonly used GWAS threshold of 5×10^{-8} (Dudbridge and Gusnanto 2008). Figure 4.8a shows the corresponding MAGWAS Manhattan plot, with a regional LocusZoom plot (Figure 4.8b) (Pruim et al. 2010). The associated SNP (rs1947825), is located in an open reading frame C17orf54 (See Figure 4.8b). We examined the cytotoxicity patterns for each genotype in Figure 4.5c. Interestingly, while the measured cytotoxicity of the heterozygous genotype (AT) consistently was in the middle between the measurements of the homozygous genotypes across all concentrations, the homozygous genotypes elicited an interesting pattern: AA was higher than TT at lowest concentration, but dropped dramatically faster than TT at higher concentrations (See Figure 4.8c).

“Pathway” association analysis of gene sets/ontologies was performed for EC₁₀ phenotypes and the 1.0 million SNPs using gene set scan (Schaid et al. 2012), which performs resampling to compute significance of SNPs, genes, and ontologies (KEGG and Gene Ontologies) in a hierarchical manner. For each mixture and ontology, we applied family-wise error rate (FWER) control using 10,000 resamples, and report in Table 4.7 all of the ontology findings with $FWER < 0.3$. Several metabolism pathways were significantly associated with current-use pesticide mixture (Table 4.7). The top contributing eight genes within each of those pathways were mainly from the uridine diphosphate (UDP) glucuronosyltransferases (UGT) family. The UGT genes regulate of the UGT enzymes responsible for glucuronidation which is generally accountable for transforming compounds into water-soluble glucuronides for excretion in bile and urine (Burchell 2003).

E. DISCUSSION

With the recent shift in the focus of many toxicology studies from *in vivo* to *in vitro* methods, substantial advancements in high-throughput approaches to characterize *in vitro* biological activity have been implemented (Dix et al. 2007; Bucher et al, 2008). Nonetheless, several challenges remain in establishing meaningful human health risk assessments from *in vitro* endpoints. Additionally, a lack in comprehensive understanding of population variability in susceptibility to chemicals remains. Regulatory risk assessment incorporates multiple uncertainty factors that are based on assumptions. Furthermore, no clear framework has been set to determine the toxicity of chemical mixtures. Regulatory authorities have raised some concern about the risks posed by complex chemical mixtures in the environment (Kepner, 2004). Few environmental chemical mixtures have been evaluated, especially at environmentally relevant concentrations (Carvalho et al., 2014), with regulatory decisions primarily based on a single compound evaluation. However, potentiation and synergistic interactions of chemicals in mixtures is of great concern (Cedergreen, 2014). It has been shown that exposure to chemical mixtures, including pesticides, often occurs with each chemicals in the mixture present at respective safety limit concentrations (Carvalho et al., 2014). Moreover, evaluation of chemical mixtures with similar modes of action, without consideration of realistic exposure in the environment, might underestimate the toxicological risk associated with their exposure (Hadrup, 2014).

In response to these needs, we aimed to provide a quantitative experimental measurement for population-based *in vitro* toxicity that could be applied to regulatory risk assessments of environmentally-relevant concentrations of pesticide mixtures. This screening paradigm provides quantitative data on population-wide variability in toxicity, which may be used to establish data-

driven uncertainty estimates when extrapolating *in vitro* data to potential *in vivo* toxicity (Judson et al. 2011). Our results show that both of the pesticide mixtures we tested exhibited a similar, considerable inter-individual variation in the induction of toxicity. The toxicodynamic uncertainty factor (UF_d) for population variability was approximately 3-fold from the median to largest tail for both mixtures. Interestingly, this finding is consistent with the assumed uncertainty factor of 3.2 (100.5) that is used for toxicodynamic studies in risk assessment (WHO 2005). The calculated UF_d can be used to obtain a chemical-specific adjustment factor by adding the toxicokinetics uncertainty factor (UF_k) (WHO 2005). To our knowledge, this is the first study to examine inter-individual variability in response to chemical mixtures. Investigation of population variability to more than 100 individual chemicals has been previously investigated (Abdo et al., in preparation). Interestingly, the toxicodynamic uncertainty factor in our present study for pesticide mixtures was similar to the median inter-individual variability for the 179 individual chemicals previously tested.

On average, there was no significant difference between the *in vitro* cytotoxicity of the current-used pesticide mixture and the chlorinated pesticide mixture. However, the *in vitro*-to-*in vivo* extrapolation data showed that a significantly lower dose of chlorinated pesticides would lead to the internal concentration equal to the cytotoxicity-eliciting EC₁₀. This observation confirms that relying on the quantitative *in vitro* potencies alone for ranking chemical mixtures might not accurately reflect the potential risk associated with these chemicals, due to differences in bioavailability, clearance, and *in vivo* exposure (Blauboer, 2010). Incorporation of human dosimetry and predicted human exposure will contribute tremendously to the “presumed hazard” calculated by *in vitro* high throughput screening alone, and provides improved estimates for informed regulatory decisions (Blauboer, 2010; Cohen Hubal et al., 2010).

It is not surprising that the cumulative human predicted exposure limit is much higher for the current-use pesticide mixture compared to the chlorinated pesticide mixture, which mostly consisted of banned-used pesticides. The current-use pesticide mixture included 36 currently used pesticides and mimicked real exposure levels in Eastern North Carolina, with Atrazine (ATZ) pesticides being the most abundant. ATZ is the most highly applied pesticide (64-80 million pounds annually in the United States), and the second most widely used agricultural pesticide in the United States (Donaldson et al. 2002; Kiely et al. 2004; Barr et al., 2007). Therefore, the predicted exposure limit for the current-use pesticide mixture was expected to be high, and in our case it was very close to our calculated oral equivalent dose.

With the availability of genetic information for our screened cell lines from the 1000 Genomes Project (Durbin et al., 2010), we were able to establish genotype-phenotype associations. Recognizing the genetic underpinning of cytotoxicity may offer exceptional insight as to the underlying casual physiological variation and biologically associated pathways. One of the major challenges in the interpretation of GWAS results is posed by SNPs located in non-coding regions. While hundreds of loci have been identified from many GWAS studies for diverse diseases and quantitative phenotypes (Hindorff et al., 2012), more than 90% of them were in non-coding regions (Jones et al. 2012; Fraser 2013). However, while some of those locations were discovered to have a role in transcriptional regulatory mechanisms, including modulation of promoter and enhancer elements (Cookson et al., 2009; Pomerantz et al., 2009; Musunuru et al., 2010; Harismendy et al., 2011), and enrichment within expression quantitative trait loci (eQTL) (Cookson et al., 2009; Nicole et al., 2010; Denger et al., 2012), the roles of others have yet to be determined. While we see a suggestive association between cytotoxicity

and the rs1947825 polymorphisms, we do not yet know how this open reading frame affects the cytotoxicity phenotype.

Through pathway analyses we found that the chlorinated pesticide mixture is significantly associated with UGT metabolism enzymes. UGTs can reduce the toxic effects of pesticides and other drugs by facilitating their excretion in bile and urine through transforming them to less-toxic water-soluble glucuronides (Burchell, 2003; Ahmad & Forgash, 1976; Meech et al., 2012). Although glucuronidation typically produces less toxic compounds, it may activate xenobiotics to produce reactive acylglucuronides that can cause cytotoxicity (Wieland et al., 2000; Bailey & Dickinson, 2003; Stingl et al., 2014). This is particularly important because UGT enzymes are genetically polymorphic with more than 200 alleles (Stingl et al., 2014). Polymorphisms in UGT1 and UGT2 families can alter enzymatic role, cellular processes, or gene expression (Stingl et al., 2014), thereby possibly affecting individual's cytotoxic response. The majority of compounds are metabolized mainly by 1A1, 1A3, 1A4, 1A9 and 2B7 (Stingl et al., 2014), which were the top ten significant genes associated with the cytotoxicity of the chlorinated pesticide mixture (see Table 4.7). This finding suggests that variation in the genes coding for these enzymes may be particularly relevant in metabolizing chlorinated pesticide mixtures.

Like any toxicological model system, *in vitro* toxicity profiling using LCLs has a number of limitations for the extrapolation to intact humans, including a lack of metabolism, as well as the inability to establish cell-cell interactions, assess target organ adverse effects, investigate potential role of other environmental factors such as lifestyle, diet, or co-exposures, or evaluate chronic toxicity. Furthermore, a great deal of debate exists regarding how chemicals may interact with one another in mixtures, both in terms of PK and in terms of toxicity. Establishing a safety

assessment of mixtures may be somewhat constrained by the assumptions we have made. There remains a pressing need to screen individual pesticides, in addition to mixtures, in order to test these assumptions. In addition, our work highlights the need for a more complete assessment of oral equivalent and real human exposures for pesticides and other chemicals.

FIGURES

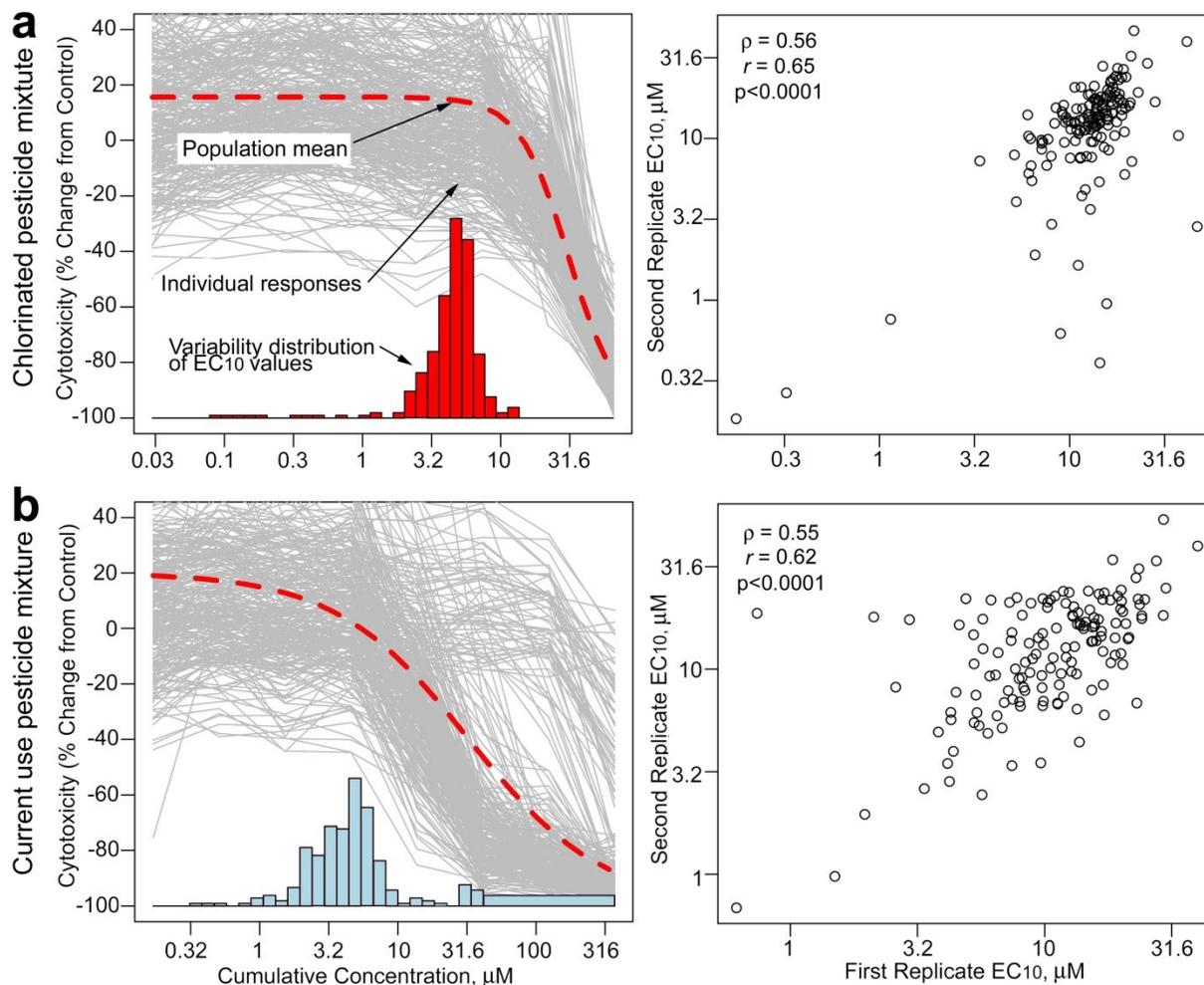


Figure 4.1. Inter-individuals and Population Variability and Reproducibility of the Mixtures. *Left panel*, a population concentration response was modeled using *in vitro* qHTS data using chlorinated pesticide mixture (panel a) and current-use pesticide mixture (panel b) data. Logistic concentration-response modeling was performed for each individual to the values shown in gray, providing individual 10% effect concentration values (EC₁₀). The red dashed line represents the logistic concentration-response for the population's mean (the data are first pooled for all individuals and then fit using a single concentration-response curve). The EC₁₀

obtained by performing the modeling on average assay values for each concentration (see frequency distribution) are shown in the inset.

Right panel, experimental reproducibility for cytotoxicity of chlorinated pesticide mixture (panel a) and current-use pesticide mixture (panel b). EC₁₀ cytotoxicity values for replicate pairs were plotted and Spearman and Pearson's correlations are shown.

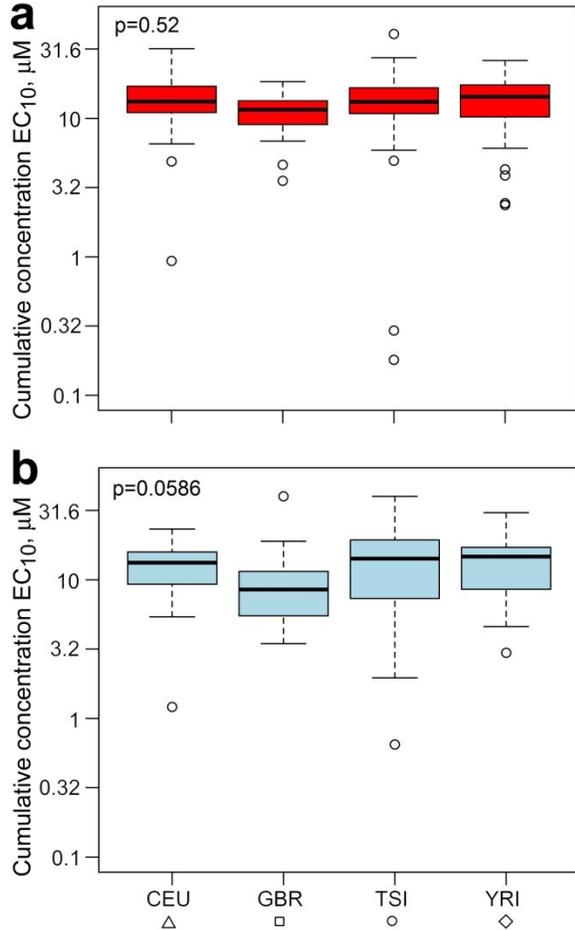


Figure 4.2. Boxplots of Population Differences for Each Mixture. Boxplots* of EC₁₀ values by population for chlorinated pesticide mixture (a) and chlorinated pesticide mixture (b), which showed marginally significant population differences by ANOVA for current-use pesticide mixture (p=0.058).

*The bottom, band inside the box, and top of the box are the first, second (the median) and third quartiles, The whiskers represent 1.5 the lowest and highest datum within 1.5 Inter Quartile Range (IQR). Small circles are outliers with >1.5 IQR above minimum or maximum datum.

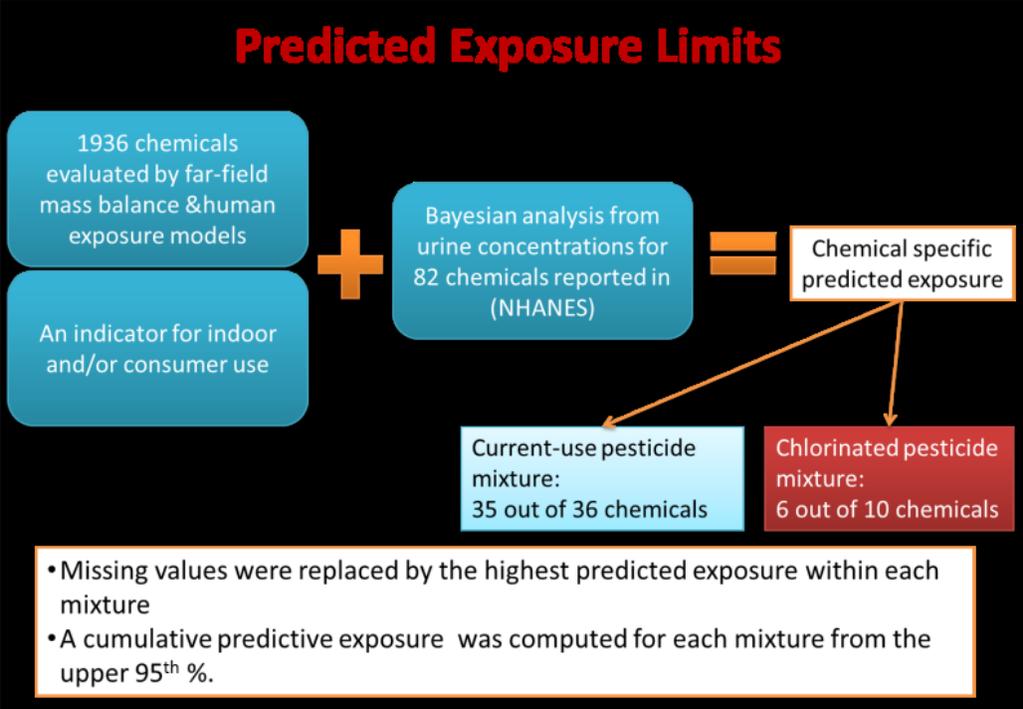


Figure 4.3 Illustrative Flow Chart of Predicted Exposure Limits Calculations. Chemical specific predicted exposure was obtained as previously described in Wambaugh et al., 2013.

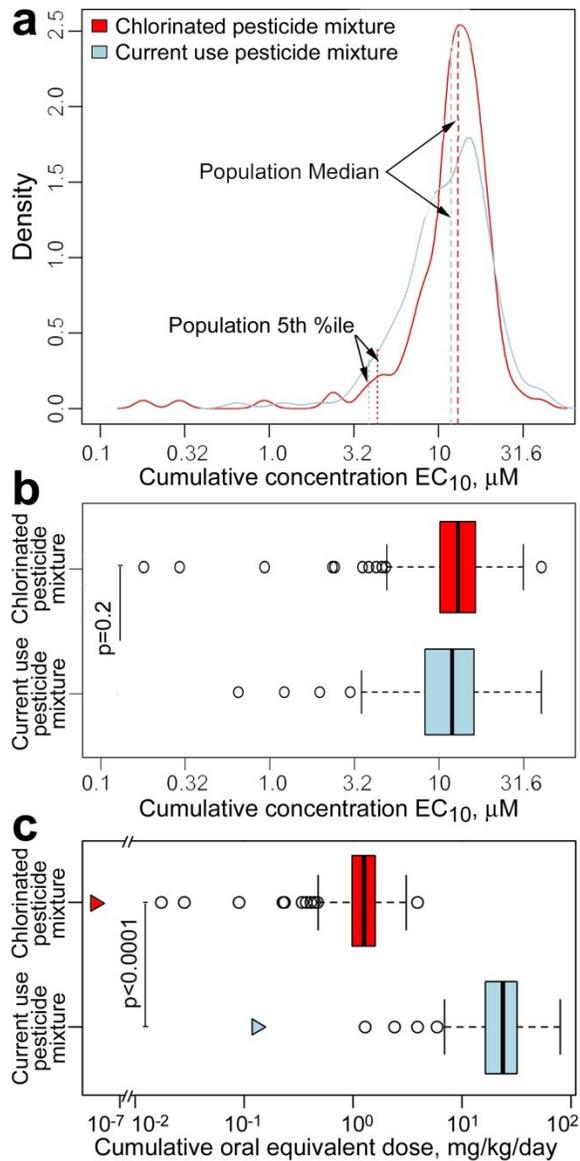


Figure 4.4 Distribution of EC_{10} s and Oral Equivalent Dose Across 146 cell lines for each mixture. *Panel a*, represents a density plot for the distribution and mean of EC_{10} of each pesticide mixture^ϕ across 146 cell lines. *Panel b*, boxplots* of EC_{10} values for each of pesticide mixtures^ϕ across 146 cell lines. *Panel c*, boxplots* of the cumulative oral doses in mg/kg/day for each of pesticide mixtures^ϕ across the 146 cell lines. The triangles represent the computed predicted exposure limit for each pesticide mixture^ϕ.

*The bottom, band inside the box, and top of the box are the first, second (the median) and third quartiles, The whiskers represent 1.5 the lowest and highest datum within 1.5 Inter Quantile Range (IQR). Small circles are outliers with >1.5 IQR above minimum or maximum datum.

♠ blue: Current Pesticide Mixture; red: Chlorinated Pesticide Mixture

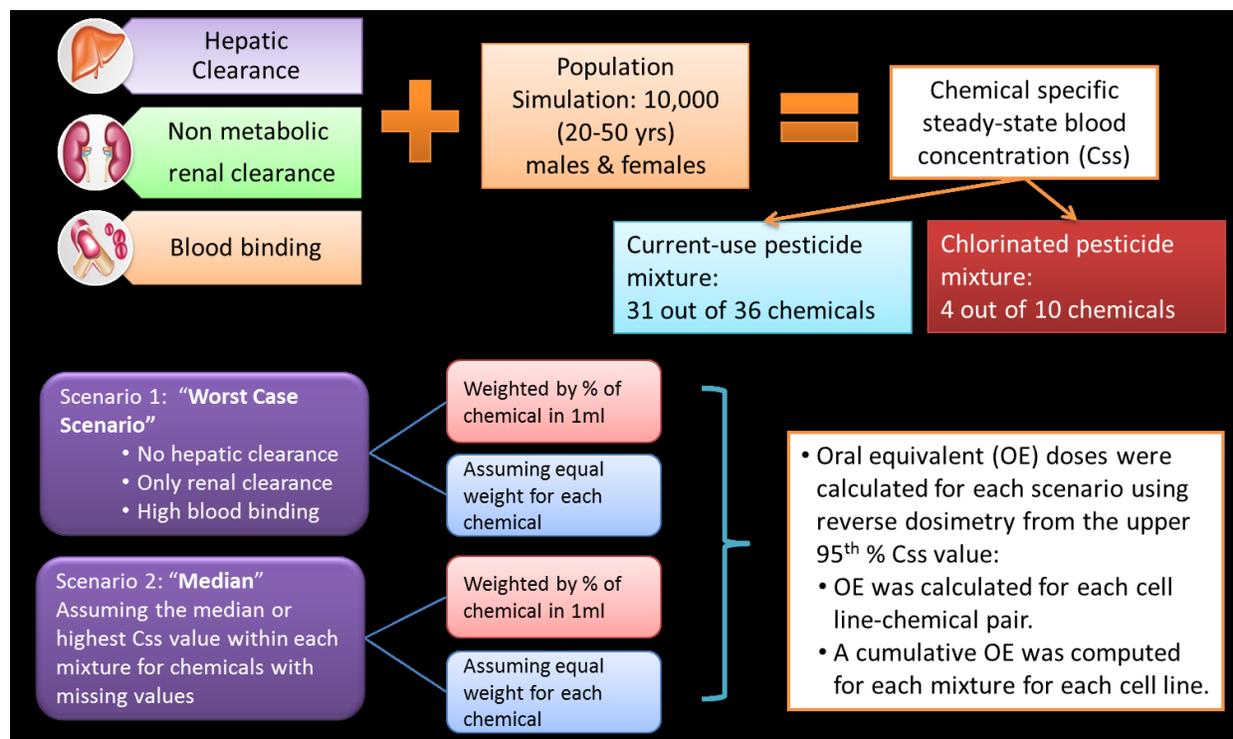


Figure 4.5. Illustrative Flow chart of Oral Equivalent Dose Calculations. Chemical specific steady-state values were obtained as previously described in Wetmore et al. 2012.

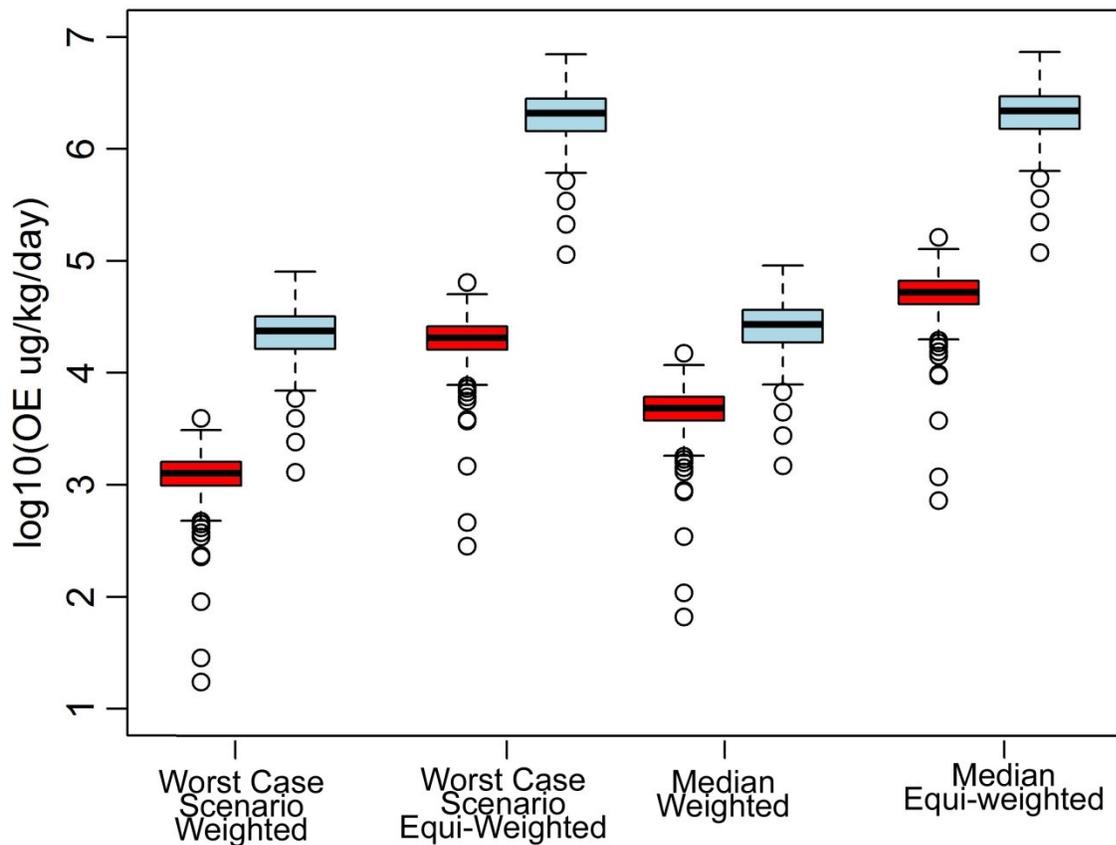


Figure 4.6 Oral Equivalent Doses with the Four Different Scenarios. Boxplots* of the cumulative oral doses in $\log_{10}(\mu\text{g}/\text{kg}/\text{day})$ for each of pesticide mixtures (red: chlorinated pesticide mixture, and blue: current pesticide mixture) across the 146 cell lines in four different scenarios: weighted by chemical percentage in mixture or not, and assuming worst case scenario (WCS) vs median for missing values.

*The bottom, band inside the box, and top of the box are the first, second (the median) and third quartiles, The whiskers represent 1.5 the lowest and highest datum within 1.5 Inter Quartile Range (IQR). Small circles are outliers with >1.5 IQR above minimum or maximum datum.

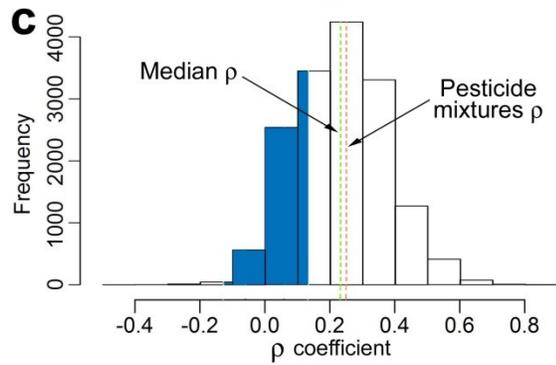
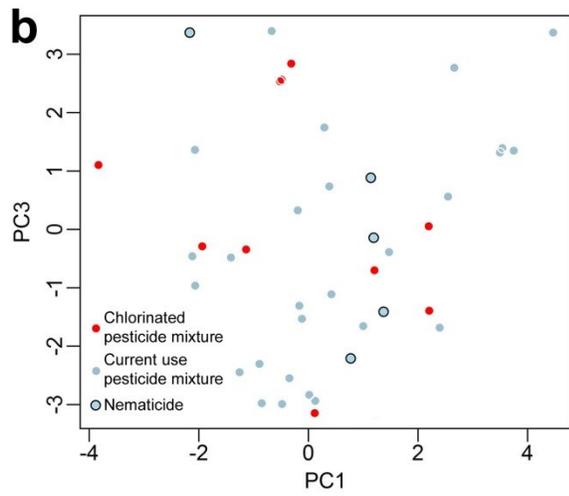
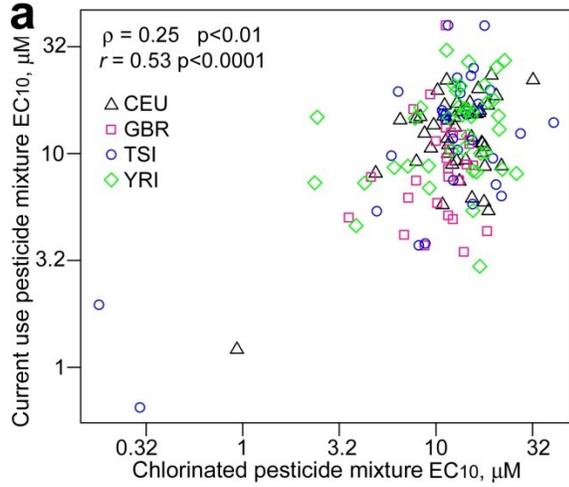


Figure 4.7. Inter-individual and Population Variability and Reproducibility of the

Mixtures. *Panel a*, a scatter plot comparison of EC_{10} values for each pesticide mixture. Each

symbol represents one of the four populations. Pearson and Spearman correlations are shown at

the top left side. *Panel b*, a scatter plot of 1st and 3rd principal components of individual pesticide

chemical structures within both mixtures colored blue for current-use pesticide mixture and red for chlorinated pesticide mixture. Five pesticides were Nematicides and were marked by an outer black circle. *Panel c*, represents a frequency histogram of 15931 spearman correlations produced pairwise correlations of 179 random chemicals. The green dashed line represents the median r value for all correlations, and the red dashed line represents the pairwise correlation of pesticide mixtures in comparison to all chemicals. Blue shading represents non-significant correlations with FDR correction.

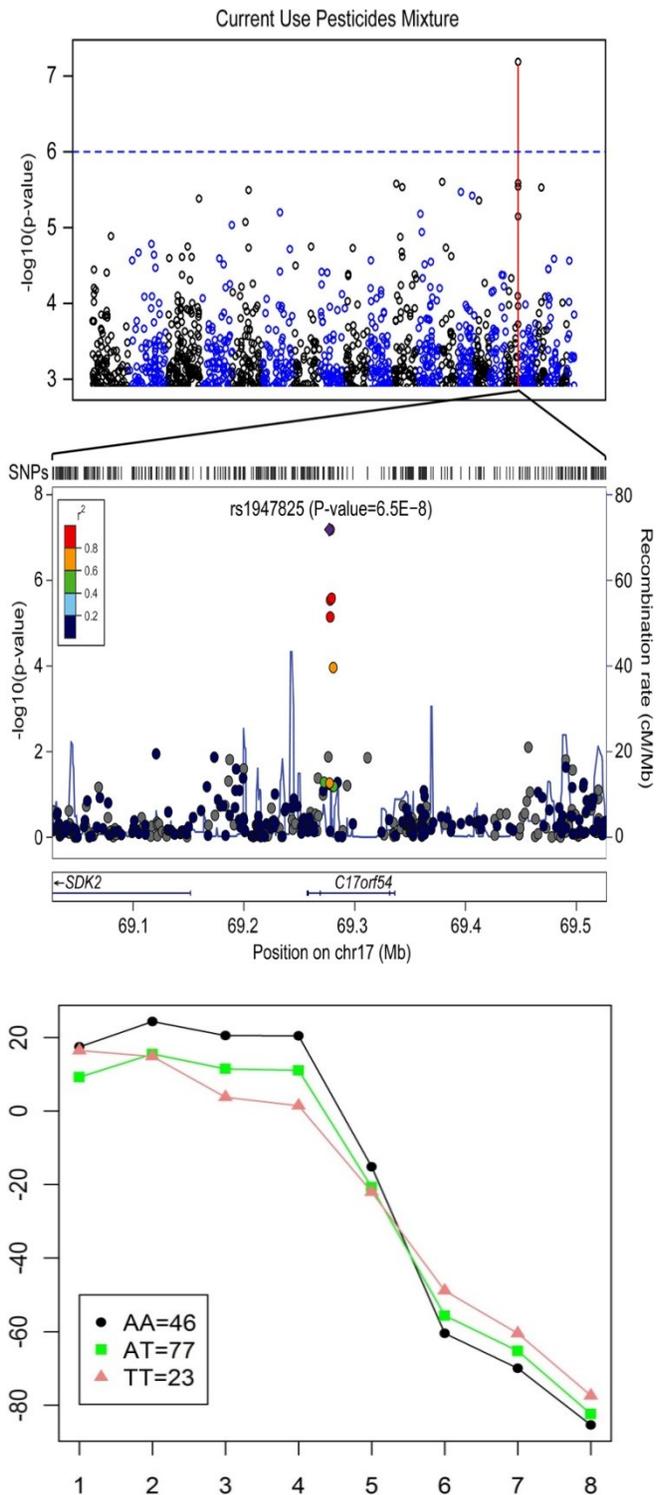


Figure 4.8 MAGWAS Results for Current-Use Pesticide Mixture. *Upper panel.* Manhattan plot of MAGWAS $-\log_{10}(P)$ vs. genomic position, for association of genotype and cytotoxicity

to current-use pesticide mixture. The line of suggestive association (expected once per genome scan) is in dashed blue. *Middle panel*, a LocusZoom plot of the most significant region, SNP rs1947825 was the most significant ($P=6.5\times 10^{-8}$). *Lower panel*, Average concentration-response profiles of cytotoxicity of current-use pesticide mixture plotted for each rs1947825 genotype.

Table 4.1. The Screened Population's Distribution

Population	# of Cell lines Screened	% of Total	N males	N females
CEU: Utah residents with Northern & Western European ancestry	47	32.2%	24	23
YRI: Yoruban in Ibadan, Nigeria	40	27.4%	19	21
TSI: Tuscan in Italy	32	21.9%	16	16
GBR: British from England & Scotland	27	18.5%	14	13
Total	146	100%	73	73

Table 4.2. Individual Pesticides in Chlorinated Pesticide Mixture.

Organochlorine pesticides					
PSD extract from waste site	MW	CAS #	µg in 1 mL	µmoles in 1 mL	% in 1mL
alpha-BHC	290.8	319-84-6	107	0.368	5.60
beta-BHC	290.8	319-85-7	55	0.189	2.88
gamma-BHC (lindane)	290.8	58-899/ 55963-79-6	151	0.519	7.90
delta-BHC	290.8	319-86-8	41	0.141	2.15
cis-chlordane	409.8	5103-71-9	18	0.044	0.67
trans-chlordane	409.8	5103-74-2	15	0.037	0.56
4,4'-DDD	320.1	72-54-8	293	0.915	13.94
4,4'-DDE	318.0	72-55-9	1193	3.75	57.11
4,4'-DDT	354.5	50-29-3	176	0.496	7.56
Dieldrin	380.9	60-57-1	41	0.108	1.64
Cumulative Concentration			2090	6.57	100

Table 4.3. Individual Pesticides Current-use Pesticide Mixture.

Current-Use Pesticide Mixture	MW	CAS #	µg in 1 mL	µmoles in 1 mL	% in 1mL
2,6-Diethylaniline	149.2	579-66-8	1259	8.44	19.76
Aldicarb	190.3	116-06-3	92	0.484	0.74
Benfluralin	335.3	1861-40-1	76	0.227	1.31
Butylate	217.4	2008-41-5	193	0.888	1.35
Carbaryl	201.2	63-25-2	106	0.527	0.39
Carbofuran	221.3	1563-66-2	85	0.384	0.50
Desethyl atrazine	187.6	6190-65-4	1352	7.21	0.37
Atrazine-desisopropyl	173.6	1007-28-9	1271	7.32	1.04
Ethalfuralin	333.3	55283-68-6	82	0.246	0.63
Flumetralin	421.7	62924-70-3	81	0.192	0.37
Metribuzin	214.3	21087-64-9	98	0.457	0.89
Napropamide	271.4	15299-99-7	37	0.136	0.12
Pebulate (Tilliam)	203.4	1114-71-2	56	0.275	0.61
Pendimethalin	281.3	40487-42-1	33	0.117	0.29
Tebuthiuron	228.3	34014-18-1	65	0.285	14.74
Trifluralin	335.5	1582-09-8/ 75635-23-3	78	0.232	2.40
Alachlor	269.8	15972-60-8	37	0.137	1.23
Atrazine	215.7	1912-24-9	1178	5.46	0.35
Chlorothalonil	265.9	1897-45-6	29	0.109	2.00
Chlorpyrifos (Dursban)	350.6	2921-88-2	71	0.202	0.48
Cyanazine	240.7	21725-46-2/ 11096-88-1	31	0.129	0.55
Dacthal	332.0	65862-98-8/ /1861-32-1	15	0.045	0.37
Tribufos (DEF 6)	314.5	78-48-8	41	0.130	0.26
Diazinon	304.4	333-41-5	89	0.292	0.79
Disulfoton	274.4	298-04-4	26	0.095	0.77
Fonofos (Dyfonate)	246.3	944-22-9	32	0.130	0.32
Ethoprop	242.3	13194-48-4	45	0.186	1.09
Fenamiphos	303.4	22224-92-6	54	0.178	0.27
Methyl parathion	263.2	298-00-0	36	0.137	0.66
Metolachlor	283.8	94449-58-8/ 51218-45-2	115	0.405	22.77
Molinate	187.3	2212-67-1	139	0.742	19.45
Permethrin	391.3	52645-53-1	39	0.100	0.52
Prometon	225.3	1610-18-0	74	0.328	1.42
Prometryne	241.4	7287-19-6/ 83653-07-0	42	0.174	0.47
Simazine	201.7	122-34-9	101	0.501	0.35
Terbufos	288.4	13071-79-9	42	0.146	0.35
Cumulative Concentration	261.3		7200	37.0	100

Table 4.4. Summary Statistics of EC₁₀ for each mixture.

Pesticide Mixture	Mean	STD⁺	Range	Median	Q05*	Q95^φ	Ufd^λ
Chlorinated Pesticide Mixture	11.6	1.96	(0.180-40.6)	13.1	4.36	21.7	3.00
Current Pesticide Mixture	11.1	1.85	(0.649-39.9)	11.9	3.89	24.7	3.05

⁺ The standard deviation of EC₁₀

^{*} The value corresponding the 5th percentile of EC₁₀ across 146 averaged values for each individual

^φ The value corresponding the 95th percentile of EC₁₀ across 146 averaged values for each individual

^λ The population toxicodynamic uncertainty factor corresponding to each pesticide.

Table 4.5. Margin of Exposure for Current-Use Pesticide Mixture.

Margin of Exposure for Current-Use Pesticide Mixture				
Scenario		Margin of Exposure*		
		Minimum	5th percentile	Median
Weighted by chemical %	Worst Case Scenario	1.0	1.8	2.3
	Median	1.1	1.8	2.3
Equally Weighted	Worst Case Scenario	2.9	3.7	4.2
	Median	3.0	3.7	4.2

*Margin of exposure is measure by obtaining the fold difference for each of the 4 scenarios of oral equivalent dose and the predicted exposure limits (minimum, 5th percentile, and median)

Table 4.6. Margin of Exposure for Chlorinated Pesticide Mixture.

Margin of Exposure for Chlorinated Pesticide Mixture				
Scenario		Margin of Exposure		
		Minimum	5th percentile	Median
Weighted by chemical %	Worst Case Scenario	5.9	7.2	7.7
	Median	6.4	7.8	8.3
Equally Weighted	Worst Case Scenario	7.1	8.5	8.9
	Median	7.5	8.9	9.3

*Margin of exposure is measure by obtaining the fold difference for each of the 4 scenarios of oral equivalent dose and the predicted exposure limits (minimum, 5th percentile, and median)

Table 4.7: Top Results from Pathway Analysis of Pesticide Mixtures

compound	ID	Term	N gene s	pval.f wer	Top 7 Genes for each pathway
Chlorinated Pesticides	KE GG	Ascorbate and aldarate metabolism	22	0.009	UGT2B11 UGT2B7 UGT1A3 UGT1A7 UGT1A4 UGT1A5 UGT1A6
Chlorinated Pesticides	KE GG	Starch and sucrose metabolism	48	0.034	UGT2B11 UGT2B7 UGT1A3 UGT1A7 UGT1A4 UGT1A5 UGT1A6
Chlorinated Pesticides	KE GG	Porphyrin and chlorophyll metabolism	39	0.06	EARS2 UGT2B11 UGT2B7 BLVRA UGT1A3 UGT1A7 UGT1A4
Chlorinated Pesticides	KE GG	Pentose and glucuronate interconversions	28	0.08	UGT2B11 UGT2B7 UGT1A3 UGT1A7 UGT1A4 UGT1A5 UGT1A6
Chlorinated Pesticides	KE GG	Nitrogen metabolism	23	0.08	CA6 GLUL CA2 CA4 HALCTH CA5A
Chlorinated Pesticides	KE GG	p53 signaling pathway	68	0.185	DDB2 CCNE2 CHEK1 TP73 CD82 SFN SERPINE1
Chlorinated Pesticides	KE GG	Other types of O-glycan biosynthesis	42	0.201	UGT2B11 UGT2B7CHST10 UGT1A3 UGT1A7 UGT1A4 UGT1A5
Chlorinated Pesticides	KE GG	Prion diseases	35	0.204	IL1B C8A PRKACA C8B IL1A IL6 C8G
Chlorinated Pesticides	KE GG	Vitamin digestion and absorption	24	0.246	BTD MMACHC SLC19A2 APOA4 PLB1 SLC19A1 APOA1
Chlorinated Pesticides	KE GG	Drug metabolism - other enzymes	48	0.288	UGT2B11 UGT2B7 IMPDH2 UGT1A3 UGT1A7 UGT1A4 UGT1A5
Current-use Pesticides	KE GG	Sulfur relay system	10	0.218	MOCS2 TRMU MPST CTU2 CTU1 TST MOCS3
Current-use Pesticides	KE GG	Other types of O-glycan biosynthesis	42	0.299	RFNG CHST10 UGT1A10 UGT1A8 UGT1A7 UGT1A9 UGT1A6
Chlorinated Pesticides	GO	regulation of triglyceride biosynthetic process	5	0.286	NR1H3 NR1H2 FITM2 GPAM ACSL5
Chlorinated Pesticides	GO	regulation of lipoprotein lipase activity	21	0.3	NR1H3 NR1H2 SORT1 APOA4 APOC1 APOC3 APOH

F. REFERENCES

- Krieger R. 2010. *Hayes' Handbook of Pesticide Toxicology*. Imprint: Academic Press. Volume 1. ISBN: 978-0-12-374367-1.
- Rother HA. 2014. Communicating pesticide neurotoxicity research findings and risks to decision-makers and the public. *Neurotoxicology*.
- Bassil KL, Vakil C, Sanborn M, Cole DC, Kaur JS, Kerr KJ. 2007. Cancer health effects of pesticides: systematic review. *Canadian family physician Medecin de famille canadien* 53(10): 1704-1711.
- Sanborn M, Kerr KJ, Sanin LH, Cole DC, Bassil KL, Vakil C. 2007. Non-cancer health effects of pesticides: systematic review and implications for family doctors. *Canadian family physician Medecin de famille canadien* 53(10): 1712-1720.
- Perry L, Adams RD, Bennett AR, Lupton DJ, Jackson G, Good AM, et al. 2014. National toxicovigilance for pesticide exposures resulting in health care contact - An example from the UK's National Poisons Information Service. *Clin Toxicol (Phila)* 52(5): 549-555.
- Jurewicz J, Hanke W. 2008. Prenatal and childhood exposure to pesticides and neurobehavioral development: review of epidemiological studies. *International journal of occupational medicine and environmental health* 21(2): 121-132.
- Manikkam M, Tracey R, Guerrero-Bosagna C, Skinner MK. 2012. Pesticide and insect repellent mixture (permethrin and DEET) induces epigenetic transgenerational inheritance of disease and sperm epimutations. *Reproductive toxicology* 34(4): 708-719.
- OPP (Office of Pesticide Program). 2002. *Guidance on Cumulative Risk Assessment of Pesticide Chemicals That Have a Common Mechanism of Toxicity*. U.S. Environmental Protection Agency. January 14, 2002.
- National Research Council. 2009. *Science and Decisions: Advancing Risk Assessment*. Washington, DC: National Academies Press.
- Andersen ME, Krewski D. 2009. Toxicity testing in the 21st century: bringing the vision to life. *Toxicol Sci* 107(2): 324-330.
- Lock EF, Abdo N, Huang R, Xia M, Kosyk O, O'Shea SH, et al. 2012. Quantitative high-throughput screening for chemical toxicity in a population-based in vitro model. *Toxicol Sci* 126(2): 578-588.
- O'Shea SH, Schwartz J, Kosyk O, Ross PK, Wright F, Tice R, et al. 2010. In vitro screening for population variability in chemical toxicity. In: *Society of Toxicology Annual Meeting*. Salt lake City, UT.
- 1000 Genomes Project Consortium. 2010. A map of human genome variation from population-scale sequencing. *Nature* 467(7319): 1061-1073.

Xia M, Huang R, Witt KL, Southall N, Fostel J, Cho MH, et al. 2008. Compound cytotoxicity profiling using quantitative high-throughput screening. *Environ Health Perspect* 116(3): 284-291.

Johnson WE, Li C, Rabinovic A. 2007. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* 8(1): 118-127.

WHO. 2005. Chemical-Specific Adjustment Factors for interspecies differences and human variability: Guidance document for use of data in dose-concentration-response assessment. Harmonization Project Document No. 2(No. 2).

Brown CC, Havener TM, Medina MW, Krauss RM, McLeod HL, Motsinger-Reif AA. 2012. Multivariate methods and software for association mapping in dose-response genome-wide association studies. *BioData mining* 5(1): 21.

Pillai K. 1955. Some new test criteria in multivariate analysis. *Ann Math Stat* 26: 117-121.

Wetmore BA, Wambaugh JF, Ferguson SS, Sochaski MA, Rotroff DM, Freeman K, et al. 2012. Integration of dosimetry, exposure, and high-throughput screening data in chemical toxicity assessment. *Toxicol Sci* 125(1): 157-174.

Rotroff DM, Wetmore BA, Dix DJ, Ferguson SS, Clewell HJ, Houck KA, et al. 2010. Incorporating human dosimetry and exposure into high-throughput in vitro toxicity screening. *Toxicol Sci* 117(2): 348-358.

Simcyp Ltd, 2001. <http://www.simcyp.com/>

Cockcroft DW, Gault MH. 1976. Prediction of creatinine clearance from serum creatinine. *Nephron* 16(1): 31-41.

Barter ZE, Bayliss MK, Beaune PH, Boobis AR, Carlile DJ, Edwards RJ, et al. 2007. Scaling factors for the extrapolation of in vivo metabolic drug clearance from in vitro data: reaching a consensus on values of human microsomal protein and hepatocellularity per gram of liver. *Current drug metabolism* 8(1): 33-45.

Johnson TN, Tucker GT, Tanner MS, Rostami-Hodjegan A. 2005. Changes in liver volume from birth to adulthood: a meta-analysis. *Liver transplantation : official publication of the American Association for the Study of Liver Diseases and the International Liver Transplantation Society* 11(12): 1481-1493.

Jamei M, Marciniak S, Feng K, Barnett A, Tucker G, Rostami-Hodjegan A. 2009. The Simcyp((R)) Population-based ADME Simulator. *Expert Opin Drug Metab Toxicol*.

Tan YM, Liao KH, Clewell HJ, 3rd. 2007. Reverse dosimetry: interpreting trihalomethanes biomonitoring data using physiologically based pharmacokinetic modeling. *J Expo Sci Environ Epidemiol* 17(7): 591-603.

- Wambaugh JF, Setzer RW, Reif DM, Gangwal S, Mitchell-Blackwood J, Arnot JA, et al. 2013. High-throughput models for exposure-based chemical prioritization in the ExpoCast project. *Environmental science & technology* 47(15): 8479-8488.
- Wood A. 2014. Compendium of Pesticide Common Names. <http://www.alanwood.net/pesticides/>
- Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24, 417-441, and 498-520.
- Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, et al. 2010. LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* 26(18): 2336-2337.
- Dudbridge F, Gusnanto A. 2008. Estimation of significance thresholds for genomewide association scans. *Genet Epidemiol*;32(3):227-34.
- Schaid DJ, Sinnwell JP, Jenkins GD, McDonnell SK, Ingle JN, Kubo M, et al. 2012. Using the gene ontology to scan multilevel gene sets for associations in genome wide association studies. *Genet Epidemiol* 36(1): 3-16.
- Burchell B. 2003. Genetic variation of human UDP-glucuronosyltransferase: implications in disease and drug glucuronidation. *American journal of pharmacogenomics : genomics-related research in drug development and clinical practice* 3(1): 37-52.
- 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.
- Bucher, J. R. Guest editorial: NTP: New initiatives, new alignment. *Environ. Health Perspect.* 2008, 116 (1), A14–A15.
- Hadrup N. 2014. Evidence from pharmacology and pathophysiology suggests that chemicals with dissimilar mechanisms of action could be of bigger concern in the toxicological risk assessment of chemical mixtures than chemicals with a similar mechanism of action. *Regul Toxicol Pharmacol.* 2014 May 17. pii: S0273-2300(14)00089-0.
- Kepner J. 2004. Synergy: The Big Unknowns of Pesticide Exposure. *Pesticides and You* 23: 17–20.
- Carvalho RN, Arukwe A, Ait-Aissa S, Bado-Nilles A, Balzamo S, Baun A, et al. 2014. Mixtures of Chemical Pollutants at European Legislation Safety Concentrations: How Safe are They?. *Toxicol Sci.* 2014 Jun 23. pii: kfu118.
- Cedergreen N. 2014. Quantifying synergy: a systematic review of mixture toxicity studies within environmental toxicology. *PLoS One.* 2;9(5):e96580.

Judson RS, Kavlock RJ, Setzer RW, Cohen Hubal EA, Martin MT, Knudsen TB, et al. 2011. Estimating Toxicity-Related Biological Pathway Altering Doses for High-Throughput Chemical Risk Assessment. *Chem Res Toxicol* 24(4): 451-462.

Blaauboer BJ. 2010. Biokinetic modeling and in vitro-in vivo extrapolations. *J Toxicol Environ Health B Crit Rev* 13(2-4): 242-252.

Cohen Hubal EA, Richard A, Aylward L, Edwards S, Gallagher J, Goldsmith MR, et al. 2010. Advancing exposure characterization for chemical evaluation and risk assessment. *J Toxicol Environ Health B Crit Rev* 13(2-4): 299-313.

Durbin RM, Abecasis GR, Altshuler DL, Auton A, Brooks LD, Gibbs RA, et al. 2010. A map of human genome variation from population-scale sequencing. *Nature* 467(7319): 1061-1073.

Hindorff L, et al. 2012. A Catalog of Published Genome-Wide Association Studies. available at <http://www.genome.gov/gwastudies>.

Jones FC, Grabherr MG, Chan YF, Russell P, Mauceli E, Johnson J, et al. 2012. The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* 484(7392): 55-61.

Fraser HB. 2013. Gene expression drives local adaptation in humans. *Genome Res* 23(7): 1089-1096.

Pomerantz MM, , Ahmadiyeh N, Jia L, Herman P, Verzi MP, Doddapaneni H, et al. 2009. The 8q24 cancer risk variant rs6983267 shows long-range interaction with MYC in colorectal cancer. *Nat Genet.* 2009;41:882–884.

Musunuru K, Strong A, Frank-Kamenetsky M, Lee NE, Ahfeldt T, Sachs KV et al. 2010. From noncoding variant to phenotype via SORT1 at the 1p13 cholesterol locus. *Nature.* 2010;466:714–719.

Harismendy O, Notani D, Song X, Rahim NG, Tanasa B, Heintzman N, et al. 2011. 9p21 DNA variants associated with coronary artery disease impair interferon- γ signalling response. *Nature.* 2011;470:264–268.

Donaldson D, Kiely T, Grube A. Pesticides Industry Sales and Usage Report. Washington, DC: U.S. Environmental Protection Agency; 2002. 1998 and 1999 Market Estimates.

Kiely T, Donaldson D, Grube A. 2004. Pesticide industry sales and usage: 2000 and 2001 market estimates. Washington, DC: U.S. Environmental Protection Agency; 2004.

Barr D., Panuwet P, Nguyen J, Udunka J, and Needham L. 2007. Assessing Exposure to Atrazine and Its Metabolites Using Biomonitoring. *Environ Health Perspect.* Oct 2007; 115(10): 1474–1478.

Hindorff L, et al. 2012. A Catalog of Published Genome-Wide Association Studies. available at <http://www.genome.gov/gwastudies>.

Cookson W, Liang L, Abecasis G, Moffatt M, Lathrop M. 2009. Mapping complex disease traits with global gene expression. *Nat Rev Genet* 10(3): 184-194.

Nicolae DL, Gamazon E, Zhang W, Duan S, Dolan ME, Cox NJ. 2010. Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. *PLoS Genet*. 6:e1000888.

Degner JF, Degner JF, Pai AA, Pique-Regi R, Veyrieras JB, Gaffney DJ, Pickrell JK, et al. 2012. DNase I sensitivity QTLs are a major determinant of human expression variation. *Nature*. 2012;482:390–394.

Ahmad S, Forgash AJ. 1976. Nonoxidative enzymes in the metabolism of insecticides. *Annals of clinical biochemistry* 13(3): 141-164.

Meech R, Miners JO, Lewis BC, Mackenzie PI. 2012. The glycosidation of xenobiotics and endogenous compounds: versatility and redundancy in the UDP glycosyltransferase superfamily. *Pharmacol Ther* 134(2): 200-218.

Stingl JC, Bartels H, Viviani R, Lehmann ML, Brockmoller J. 2014. Relevance of UDP-glucuronosyltransferase polymorphisms for drug dosing: A quantitative systematic review. *Pharmacol Ther* 141(1): 92-116.

Bailey MJ, Dickinson RG. 2003. Acyl glucuronide reactivity in perspective: biological consequences. *Chem Biol Interact* 145(2): 117-137.

Wieland E, Shipkova M, Schellhaas U, Schutz E, Niedmann PD, Armstrong VW, et al. 2000. Induction of cytokine release by the acyl glucuronide of mycophenolic acid: a link to side effects? *Clinical biochemistry* 33(2): 107-113.

CHAPTER 5: GENERAL DISCUSSION

A. SUMMARY AND CONCLUSIONS.

Quantitative assessment of both the hazard and the degree of inter-individual biological variability in the human population is critical for proper evaluation of chemicals for potential adverse human health outcomes (Zeise et al. 2013). A comprehensive characterization of human genome sequence variation is important for understanding observed inherited variation in toxicity phenotypes. Genetic polymorphisms can have a profound influence on risk and should be considered for human risk assessment (Baynes 2012). However, such characterization and assessment is difficult to quantitatively evaluate using current *in vivo* animal test systems or *in vitro* methods with established cell lines.

The availability of genetically-defined, genetically-diverse renewable sources of human cells, such as lymphoblasts from the International HapMap and 1000 Genomes projects, enables *in vitro* testing at the population level. As the focus of risk assessment processes shifts toward *in vitro* data, the quantitative assessment of inter-individual variability in response to chemicals, and an understanding of the underlying genetic causes are necessary for regulatory decisions to be based on scientific data rather than on default assumptions. Our population-based quantitative high-throughput model is a valuable tool in evaluating both the hazard and the degree of inter-individual variability in response to chemicals.

1. QUANTITATIVE HIGH-THROUGHOUT SCREENING FOR CHEMICAL TOXICITY IN A POPULATION-BASED IN VITRO MODEL

With the recent shift in toxicity testing from *in vivo* to *in vitro*, several quantitative high-throughput screening (qHTS) approaches for computational toxicology were developed to prioritize compounds, reveal new mechanisms, and enable predictive modeling. We strived in this study to design predictive *in vitro* models of chemical-induced toxicity with a focus on inter-individual genetic variability. We exposed 81 LCLs from 27 Centre d'Etude du Polymorphisme Humain (CEPH) trios to 240 chemical substances at 12 concentrations and evaluated for cytotoxicity and apoptosis. We demonstrated the feasibility of LCLs in creating an *in vitro* model system to evaluate inter-individual and population-wide variability of chemical-induced toxicity phenotypes. Indeed, the *in vitro* genetics–anchored human model system used in the study has not only been fruitful in measuring toxicity in a population-level, but also it carried the potential to recognize candidate genetic susceptibility for individual variability. Our *in vitro* population model has proven to produce robust reproducible toxicity values for a wide variety of chemicals across the different cell lines.

Regardless of toxicity phenotype used (cytotoxicity or apoptosis); variability across individual cell lines varied from one chemical to another in magnitude but was consistent for each chemical. Through combining our toxicity data with publicly available genetic information for LCLs, we were able to generate plausible hypotheses for the associations identified by either the genotypes or the RNA-expression levels. We recognize that our sample size combined with the likely small genomic effect size will probably hinder our ability for detecting meaningful biological associations. However, we consider this study as a successful proof of concept and as

a motivation to carry out a larger scale population based screening to better understand human and genetic variability.

2. GENETIC MAPPING OF IN VITRO SUSCEPTIBILITY TO CYTOTOXIC COMPOUNDS-THE 1000 GENOMES HIGH THROUGHPUT SCREENING STUDY

There is a significant lack of understanding of human variation in response to toxic environmental chemicals. We strived in this study to address this critical gap in next generation risk assessment. We used 1086, representing 9 populations from 5 continents, drawn from the 1000 Genomes Project to assess variation in cytotoxic response to 179 chemicals. We ranked chemicals by average response and assessed population variation and heritability. Genome-wide association mapping was also performed, with attention to phenotypic relevance to human exposures. This study provides an example of how a large-scale systems biology experiment that integrates toxicity phenotyping with genetic mapping can aid in translation to public health protection. For example, our data did not only provided quantitative assessments of hazard for hundreds of chemicals, but most importantly, the hazard was identified in the most sensitive individuals. The next few paragraphs will illustrate some of the conclusions drawn from the second aim.

Some, but not all chemicals elicit inter-individual variation in cytotoxicity responses. The degree of inter-individual variability varies from one chemical to another. The degree of inter-individual variability varies from one chemical to another and was less than 10 fold for about 2/3 of the compounds. However, some compounds exhibited more than 100-fold range in variability. Overall, the median inter-individual variability across all chemicals is similar to the default assumption for toxicodynamic study, even in chemical mixtures. However, assuming a 3-fold

uncertainty factor for toxicodynamic assessment might be overestimating or underestimating the actual inter-individual variation for some chemicals. The degree of inter-individual variability, in addition to the degree of cytotoxicity, may aid in prioritization of utilized compounds for further testing using additional *in vitro* or *in vivo* approaches.

Inter-individual variability in cytotoxicity could be within and/or between the populations tested in these experiments. While certain populations were more sensitive than others to certain chemicals, the variation tended to be modest. Furthermore, the variation within populations was generally greater than across populations. The pattern for variation across population was unique for each chemical. Consequently, no population was sensitive to all or most of the chemicals. There was some evidence of hierarchical clustering by continental ancestry, indicating that on average certain populations might exhibit similar sensitivity to the same chemicals.

Combining toxicity data and publicly available genotyping and RNA-Seq information enabled the possibility to probe and select candidate susceptibility genes and pathways/networks. Genetic mapping suggested important roles for variation in membrane and trans-membrane genes with a number of chemicals showing association with rs13120371 in the solute carrier SLC7A11, which has been implicated in chemo-resistance. Analysis of public RNA-sequencing profiles on the same cell lines provided evidence of association between basal transcription and cytotoxic response, with enrichment for genes with membrane localization.

3. IN VITRO SCREENING FOR INTER-INDIVIDUAL AND POPULATION VARIABILITY IN TOXICITY OF PESTICIDE MIXTURES

Population-based human *in vitro* models offer exceptional opportunities for evaluating the potential hazard and mode of action of chemicals, as well as variability in response as shown in the previous two studies. Potential challenges of screening and assessing the cytotoxicity of complex mixtures requires further assessment to increase the utility of the information obtained from *in vitro* models. We strived in this study to explore the potential of a population-based *in vitro* model in evaluating chemical mixtures and the possibility of integrating real human exposures in assessing *in vitro* data. We used 146 lymphoblast cell lines from 4 ancestrally and geographically diverse populations that were densely genotyped and with publically available basal RNA-seq data. Cells were exposed to two pesticide mixtures (an organochlorine pesticide environmental mixture extracted from a passive surface water sampling device and a mixture of 36 currently used pesticides) at 8 concentrations and were evaluated for cytotoxicity. Cytotoxicity measures for replicates produced robust and reproducible results. On average, the two mixtures exhibited a similar range of *in vitro* cytotoxicity and showed considerable inter-individual variability across the screened cell lines. However, *in vitro*-to-*in vivo* extrapolation (IVIVE), which was performed by reverse pharmacokinetics, suggested a significantly lower oral equivalent dose for the chlorinated pesticide mixture compared to the current-use pesticide mixture. Multivariate genome-wide association mapping revealed an association between the current use-pesticide mixture and a polymorphism in rs1947825 in C17orf54. Moreover, genetic pathway analysis showed a significant association between metabolism pathways and the

cytotoxicity of the chlorinated pesticide mixture. Our model was not only valuable in quantitatively estimating hazard and understanding inter-individual variability for individual chemicals, but was also amenable to rapidly and efficiently test mixtures of chemicals as well. Moreover, through integration of our cytotoxicity data with proper dose estimation through PK modeling and real human exposure limits, we were able to advance our assessment of pesticide mixtures and understand the actual human health hazard associated with them.

Overall Conclusions

Our population-based toxicity screening that is based on genetically-defined and genetically diverse human *in vitro* model system is a more powerful approach than traditional *in vitro* approaches. First, it allowed for efficient quantitative assessment of both hazard and inter-individual variability in toxicodynamics for individual chemical and chemical-mixtures. From the concentration-response *in vitro*, we have been able to establish a quantitative toxicity phenotype (EC10) that is similar to the Lowest Observed Adverse Effect Level (LOAEL). This toxicity phenotype can serve as an *in vitro* point of departure where chemical exposures could be identified without major biological perturbations, which in our case is a meaningful depletion in ATP or cell death. Because testing was conducted from many cell lines from many human donors from representative populations, we were able to gauge population and inter-individual variability associated with the hazard for each chemical. Identifying both the hazard and inter-individual variability was useful in establishing meaningful prioritization of chemicals and exploring potential differences/similarities in modes of action between chemical substances.

Second, the availability of cell lines from many human donors representing several ancestral populations afforded identification of susceptible sub-populations. Third, combining toxicity data and publicly available genetic information for LCLs from 1000 Genomes and HapMap Projects offered the possibility to probe and select candidate susceptibility genes and pathways/networks. Consequently, we were able to understand of the genetic determinants of the inter-individual variability, the contribution of genetics to adverse toxicity phenotype, and generate testable hypotheses about toxicity pathways by leveraging genetic and genomic data. Furthermore, while most of the variability between different cell lines that is can be attributed to genetics is modest and similar to complex chronic diseases, the genetic variation identified may have a profound effect on differences between individual cell lines. Consequently, such variability can be quantified and used to generate testable hypotheses about the mechanisms of toxicity.

Finally, incorporation of reverse dosimetry and exposure estimates when exploring concentration-response relationships for individual chemicals aided tremendously in assessing human hazard. Because the *in vitro* assay endpoint does not incorporate metabolic clearance and plasma protein binding, ranking the chemical mixtures by nominal assay concentrations might result in over- or under- estimation of the steady-state of a chemical mixture.

In conclusion, our population-based high throughput model enabled us to capture a population-wide measure of uncertainty, where we are able to quantitatively estimate chemical-specific or mixture-specific range in individual variability and to understand contributing genetic variation associated with it.

B. SIGNIFICANCE, INNOVATION, AND IMPACT OF THIS STUDY

We have utilized a human lymphoblast cells representing diverse populations to understand the hazard and magnitude of inter-individual variability to different environmental chemicals and/or chemical mixtures. There are several advantages to make use of LCLs for toxicity screening, as opposed to other current *in vitro* approaches, that make our study innovative. First, LCLs are derived from healthy adult individuals as opposed to tumor driven cell lines that are being utilized by Tox21 and other current human *in vitro* screening paradigms. Consequently, LCLs provide a better understanding of toxicological effects and improved translation to a real human population when compared to cancer cell lines. Second, LCLs represent geographically diverse populations from different continents and ancestral backgrounds which facilitate the assessment of variability within and between populations. Third, the public availability of the genome-wide genotype and gene-expression data allows for investigating the molecular-genetic mechanistic underpinnings of chemical toxicity for no additional cost. This availability of data also presents the opportunity for Genome Wide Association Studies (GWAS) and identification of SNPs, genes, and/or pathways that are primarily associated with toxicity of chemicals. Such findings permit the generation of novel hypotheses of how chemicals cause toxicity, and/or validate our understanding of what is known about the mechanism of chemical toxicity. Fourth, the immortalization of LCLs grants a renewable source for repeated experiments with easy manipulation at no additional cost. In conclusion, all those qualities of LCLs provide the possibility for effective and innovative *in vitro* population-based screening of chemicals without *in vivo* confounders and surpasses other *in vitro* testing paradigms.

With the first aim, we were able to test hundreds of chemicals with LCLs at different concentrations, allowing for the identification of a concentration-specific toxicity in addition to

recognizing population variability. The study was first of its kind and proved the remarkable value of our model in achieving several gains. It served as a proof of principle that population *in vitro* screening with LCLs and careful experimental design can be used to assess both hazard and population variability, cluster chemicals for further prioritization, and help uncover potential genetic underpinnings.

With the second aim, we were able to conduct the largest scale study, in terms of number of cell lines and chemicals used, to address major gaps in current risk assessment. This study was innovative in its hypotheses, the approach and model system to be utilized, the combination of methodologies and analyses to be performed, the organizational structure, and the outstanding translational potential to human populations. We were able to quantitatively assess and address population based toxicological effects or hazard of environmental contaminants, determine the extent of human inter-individual variability in chemical toxicity, identify susceptible sub-populations or races, understand the genetic determinants of the inter-individual variability, generate testable hypotheses about toxicity pathways by leveraging genetic and genomic data from 1000 Genomes and HapMap Projects, use the data obtained from this research to build predictive *in silico* models, capture a population-wide measure of uncertainty in dose-response, and explore potential differences/similarities in modes of action between chemicals.

Furthermore, the second aim provided data for improved prioritization and clustering of chemicals according to a number of differing criteria. Clustering of chemicals can be done according to toxicity, variability between individuals, genetic mode of action, and similarity in toxicological pattern.

Finally, the high-throughput information derived from our 1000 Genomes Toxicity Screening Project was utilized to build models that can predict cytotoxicity based on either chemicals

structure or genomic profiles without the need of additional experiments. With the collaboration between Sage Bionetworks and DREAM, University of North Carolina (UNC), the National Institutes of Environmental Health Sciences (NIEHS), and the National Center for Advancing Translational Sciences (NCATS), the NIEHS-NCATS-UNC DREAM Toxicogenetics Challenge was launched in 2013. This challenge represents an innovative new track for toxicity testing and is intended to help comprehend how genetic variation affects individual response to exposure to environmental chemicals. The Toxicogenetics Challenge approached researchers to utilize the data obtained from the 1000 Genomes Toxicity Project in order to elucidate the extent to which adverse effects (e.g. cytotoxicity) of compounds can be inferred from genomic and/or chemical structure data. Participants are tasked with solving two related sub-challenges: (1) develop predictive models of cytotoxicity using genetic and genomic data to predict individual responses to compound exposure and (2) use chemical attributes to predict population-based cytotoxicity characteristics (median, variance) for a set of compounds. The challenge engaged 232 registered participants with 99 submissions from 34 teams for first sub-challenge and 91 submissions from 24 teams for the second sub-challenge. Successful models/participants were selected for each sub-challenge. The computational models built for each sub-challenge could be considered in certain decision-making contexts to inform government agencies as to which environmental chemicals and drugs are of the greatest potential concern to human health. Moreover, Nature Biotechnology will consider an overview paper describing the results and insights of successful models.

In the third aim, we wanted to expand our model to address remaining challenges in risk assessment of chemical mixtures. While high-throughput *in vitro* toxicity screening provides an efficient way to identify hazard for environmental and industrial chemicals while conserving

limited testing resources. It is hard to interpret cytotoxicity values without proper understanding of *in vivo* dose. Differences in clearance, protein binding, and other pharmacokinetic factors have tremendous consequence on the bioactivity of a chemical. With this particular study, we were able to quantify hazard and population variability in with respect to two pesticides mixtures *in vitro*, extrapolate the *in vitro* concentration to an *in vivo* associated dose, and compare its relevance to real human exposure amounts for a full assessment of the pesticide mixtures.

C. LIMITATIONS OF THIS STUDY

There are several disadvantages associated with utilizing LCLs for *in vitro* population-based screening that are related to all three studies/aims. While high-throughput *in vitro* testing offers many potential benefits, there are potential challenges in extrapolating a particular perturbing concentration in one cell type to a dose of a chemical that results in an observable change in the health or normal functioning of the whole human. First, *in vitro* testing does not necessarily cover the full biological pathway or all critical changes that be consequent to exposure to a chemical. Despite our ability to identify potential genetic pathways associated with exposure to a specific chemical, other factors such as epigenetics and cell-cell interactions might mitigate or enhance the observed a toxic response in LCLs, leading to underestimation or overestimation of the actual risk from chemical exposure. Second, while we are trying to account for human population diversity in response to chemicals by using LCLs that represent different populations, we still cannot account for other sources of population variability such as epigenomics, age, pre-existing health conditions, life style, and co-exposures. Third, while frequency, duration, and route of exposure to a chemical are key elements that need to be considered in toxicity assessment, those factors cannot be accounted in *in vitro* screening. Consequently, caution and thoughtful deliberation is required in translating of *in vitro* data to an intact organism to avoid over-interpretation or erroneous conclusions (Rothman, 2002).

The biggest shortcoming of *in vitro* toxicity testing is the lack of metabolism. When it comes to binning compounds to dichotomized categories (toxic vs non-toxic), false-positives or false-negatives might arise as consequence of absence of metabolism. LCLs are derived from B-lymphocytes whose main function is humoral adaptive immunity. Unlike hepatocytes, human lymphoblasts do not function to metabolize chemicals and therefore do not have the metabolic

capacity of the liver, or even that of freshly isolated hepatocytes. However, there is evidence that LCLs express a number of nuclear receptors, as well as most genes of the phase I and II metabolism, and transporters (Siest et al. 2008). Nevertheless, the expression of metabolism genes and nuclear receptor in lymphoblasts (Stranger et al. 2007) is about 10 to 100 times lower when compared to hepatocytes (Schadt et al. 2008a). While using high concentrations in experimental design could potentially overcome the low metabolism in LCLs, we still lack a clear understanding of metabolism in LCLs. Nonetheless, screening parent compounds with their major metabolites may offer an improved measure of true toxicity.

While LCLs have been a promising model for pharmacogenomic discoveries, their utility has been questioned because of concern with changes in cell biochemistry stemming from the immortalization process. Potential confounders that affect the utility of LCLs include baseline growth rates, EBV copy numbers and ATP levels (Choy et al. 2008). While growth rate of LCLs was associated with chemotherapeutic-induced cytotoxicity in one study (Stark et al. 2010), it was not associated with cytotoxicity of 100+ chemicals (Lock et al. 2012). Altered apoptosis responses have been observed as a result of EBV transformation in LCLs with cancer drugs (Liu et al. 2004).

The immortalization process or EBV transformation has been observed to affect gene-expression and promoter-methylation profiles of majority of genes compared to primary B cells (Caliskan et al. 2011). However, the difference in expression levels between the primary and immortalized cells was small in magnitude (<1.5 fold). Moreover, the inter-individual variability in gene expression was the same between primary B cells and LCLs (Caliskan et al. 2011). Furthermore, many expression quantitative trait loci eQTLs observed in LCLs were observed in

primary tissues like the liver, lung, and skin (Bullaughy et al. 2009; Ding et al. 2010; Schadt et al. 2008b).

In the *first specific aim*, we acknowledge that our sample size combined with the likely small genomic effect size will probably hinder our ability to detect meaningful associations. Moreover, in both *first and second aims*, further evaluation is needed to better understand the *in vitro* cytotoxicity values obtained from our model. *In vitro-to-in vivo* extrapolation might tremendously help get better comprehension of hazards associated with screened chemicals. However, such analysis is depended on availability of PK specific values for screened chemicals.

Other limitations inherent to the *third aim* are the strong assumptions we made in evaluating toxicity of mixtures. With mixtures, there is a great deal of debate in the toxicology community about how chemicals may interact with one another, both in terms of PK modeling and in terms of toxicity. While we assumed that the chemicals do not potentiate or inhibit the cytotoxicity of each other, we have no way of knowing if that assumption was entirely true. Furthermore, we assumed that the chemicals in each mixture are equipotent. However, we tried to extract cytotoxicity data for the chemicals component for each mixture from the ToxCast data. Furthermore, we had missing data in both human exposure estimates and PK modeling data for some the chemicals within each mixtures. While we tried to be conservative by considering different scenarios for substituted missing values and picking the worst case ones, we do not know if we are overestimating the actual exposure or oral equivalent dose. This is especially true, since some of the chemicals were contributing more heavily than others to oral equivalent dose and we do not know how the missing chemicals might be influencing our estimates. We need better estimates for oral equivalent dose and real human exposures for some of our pesticides that have missing data for an improved risk assessment.

D. FUTURE DIRECTIONS

There are several directions in which this research could proceed. First, we can functionally validate the discoveries made with GWAS and other analyses like RNA-Seq and pathway analyses. The confirmation of the potential association between a certain SNP or gene and the cytotoxicity in LCLs could be performed in several ways. One way is to knock down the gene of interest by SiRNA in LCLs and compare cytotoxicity across broad range of concentrations in “wild type” LCLs vs knocked down LCLs. Another potential way is to pick cell lines representing each genotype of the associated SNP and test them across a broad range of concentrations. Then statistically test the difference in cytotoxicity (either EC₁₀ or whole concentration response) between the three different genotypes.

Second, we could elucidate modes and mechanisms of toxicities of our cell lines by quantitatively assessing gene expression variation for chemicals across different concentrations. The evolving of the new technology that can quantify mRNA responses in thousands of genes, called RASL-Seq, has been rapidly growing to accommodate the high-throughput screening of many compounds screened in many cell lines across multiple concentrations at different endpoints. The gene expression in RASL-Seq technology is characterized by excellent reproducibility, high accuracy, excellent gene specificity, amenability for high-throughput multiplex economical approach (Fu et al., 2012). This technology can enable us assess quantitatively mRNA expression for our screened pesticide mixtures, drug-metabolites, and cadmium chloride at different concentrations. The gene expression profiles can illuminate our understanding of the mechanism of cytotoxicity.

Third, we can explore the potential differences/and similarities in modes of action between chemicals either on a population-wide compared to individual effects. Through our screenings,

we were able to cluster chemicals by their similarities in toxic response across different cell lines. We concluded that those similarities exist due to shared mechanism of action (MOA). Thousands of chemicals have been screened in Tox21 and ToxCast through multiple assays in different cell lines. We can leverage the data produced by ToxCast and Tox21 and try to see if shared MOA for certain chemicals across different assays was a factor in chemical's profiles across cell lines.

Finally, we can further understand cytotoxicity of mixtures by screening both the individual components and the mixtures. This approach will help us test and validate some of the assumptions we made with the mixtures study. By comparing the cytotoxicity of each chemical and comparing it to the cumulative cytotoxicity of the whole mixture, we can assess whether chemicals potentiate or inhibit the activity of other chemicals within a mixture, and we can evaluate the assumption of equal potency for chemical within each mixture.

E. REFERENCES

- Zeise L, Bois FY, Chiu WA, Hattis D, Rusyn I, Guyton KZ. 2013. Addressing human variability in next-generation human health risk assessments of environmental chemicals. *Environ Health Perspect* 121(1): 23-31.
- Baynes RE. 2012. Quantitative risk assessment methods for cancer and noncancer effects. *Progress in molecular biology and translational science* 112: 259-283.
- Siest G, Jeannesson E, Marteau JB, Samara A, Marie B, Pfister M, et al. 2008. Transcription factor and drug-metabolizing enzyme gene expression in lymphocytes from healthy human subjects. *Drug Metab Dispos* 36(1): 182-189.
- Stranger BE, Forrest MS, Dunning M, Ingle CE, Beazley C, Thorne N, et al. 2007. Relative Impact of Nucleotide and Copy Number Variation on Gene Expression Phenotypes. *Science* 315(5813): 848-853.
- Schadt EE, Molony C, Chudin E, Hao K, Yang X, Lum PY, et al. 2008a. Mapping the genetic architecture of gene expression in human liver. *PLoS biology* 6(5): e107.
- Choy E, Yelensky R, Bonakdar S, Plenge RM, Saxena R, De Jager PL, et al. 2008. Genetic analysis of human traits in vitro: drug response and gene expression in lymphoblastoid cell lines. *PLoSGenet* 4(11): e1000287.
- Stark AL, Zhang W, Mi S, Duan S, O'Donnell PH, Huang RS, et al. 2010. Heritable and non-genetic factors as variables of pharmacologic phenotypes in lymphoblastoid cell lines. *Pharmacogenomics J* 10(6): 505-512.
- Lock EF, Abdo N, Huang R, Xia M, Kosyk O, O'Shea SH, et al. 2012. Quantitative high-throughput screening for chemical toxicity in a population-based in vitro model. *Toxicol Sci* 126(2): 578-588.
- Liu MT, Chen YR, Chen SC, Hu CY, Lin CS, Chang YT, et al. 2004. Epstein-Barr virus latent membrane protein 1 induces micronucleus formation, represses DNA repair and enhances sensitivity to DNA-damaging agents in human epithelial cells. *Oncogene* 23(14): 2531-2539.
- Caliskan M, Cusanovich DA, Ober C, Gilad Y. 2011. The effects of EBV transformation on gene expression levels and methylation profiles. *Hum Mol Genet* 20(8): 1643-1652.
- Bullaughhey K, Chavarria CI, Coop G, Gilad Y. 2009. Expression quantitative trait loci detected in cell-lines are often present in primary tissues. *HumMol Genet*.
- Ding J, Gudjonsson JE, Liang L, Stuart PE, Li Y, Chen W, et al. 2010. Gene expression in skin and lymphoblastoid cells: Refined statistical method reveals extensive overlap in cis-eQTL signals. *Am J Hum Genet* 87(6): 779-789.
- Schadt EE, Molony C, Chudin E, Hao K, Yang X, Lum PY, et al. 2008b. Mapping the genetic architecture of gene expression in human liver. *PLoS Biol* 6(5): e107.