USING HIGH THROUGHPUT SCREENING FOR PREDICTIVE MODELING OF REPRODUCTIVE TOXICITY

Matthew T. Martin

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.

Chapel Hill
2011

Approved by:

Ivan Rusyn, M.D. Ph.D.

David J. Dix, Ph.D

Richard S. Judson, Ph.D.

Alexander Tropsha, Ph.D.

Avram Gold, Ph.D.

ABSTRACT

MATTHEW T. MARTIN: Using High Throughput Screening For Predictive Modeling of Reproductive Toxicity
(Under the direction of Dr. David J. Dix)

Traditional reproductive toxicity testing is inefficient, animal intensive and expensive with under a thousand chemicals ever tested among the tens of thousands of chemicals in our environment. Screening hundreds of chemicals through hundreds high-throughput biological assays generated a validated model predictive of rodent reproductive toxicity with potential application toward large-scale chemical testing prioritization and chemical testing decision-making. Chemical classification for model development began with the uniform capturing of the available animal reproductive toxicity test information utilizing an originally developed relational database and reproductive toxicity ontology. Similarly, quantitative high-throughput screening data were consistently processed, analyzed and stored in a relational database with gene and pathway mapping information. Chemicals with high quality *in vivo* and *in vitro* data comprised the training, test, external and forward validation chemical sets used to develop and assess the predictive model based on eight selected features generally targeting known modes of reproductive toxicity action. In three case studies, the forward validated predictive model reduced the overall costs of reproductive toxicity testing by roughly twenty percent. The model provides a starting point for the future of reproductive toxicity testing.

ACKOWLEDGEMENTS

First I would like to thank my research advisor, David J. Dix, PhD., for his invaluable mentorship throughout my academic and professional career, as he has provided me countless opportunities for growth, advancement and independence while continuing to provide always needed guidance. I am also grateful to my academic advisor, Ivan Rusyn, M.D. Ph.D., for his patience, flexibility and willingness to mentor and guide me through the academic process even under somewhat unique circumstances. I would also like to thank Dr. Rusyn and the other members of my committee, Richard S. Judson, Ph.D., Alexander Tropsha, Ph.D., Avram Gold, Ph.D., for taking the time to provide knowledge, expertise and support as your involvement has helped shape my education and newly formed expertise. I would also like to thank my colleagues and managers at the U.S. Environmental Protection Agency's National Center for Computational Toxicology as they have been a continuing source of learning, challenge, entertainment, and enlightenment. It is a true pleasure to work with such an intelligent and motivated group of public servants.

The research described here has been partially supported by the U.S. Environmental Protection Agency. The views expressed in this paper are those of the authors and do not necessarily reflect the views or policies of the U.S. Environmental Protection Agency. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ACToR – Aggregated Chemical Toxicity Resource

C&L – Classification and Labeling

CEBS - Chemical Effects in Biological Systems

DER – Data Evaluation Record

DSSTox – Distributed Structure-Searchable Toxicity

EOGRTS - Extended One-Generation Reproductive Toxicity Study

FDA – Food and Drug Administration

FFDCA – Federal Food, Drug and Cosmetic Act

FIFRA – Federal Insecticide, Fungicide and Rodenticide Act

FSH - Follicle stimulating hormone

FQPA – Food Quality and Protection Act

GnRH - Gonadotropin-releasing hormone

hCG - Human chorionic gonadotropin

HPG - Hypothalamic-pituitary-gonadal

HCS – High Content Screening

HTS – High Throughput Screening

IRIS – Integrated Risk Information System

LH - Lutenizing hormone

LOAEL – Lowest Observed Adverse Effect Level

LOEL – Lowest Observed Effect Level

MGR – Multigeneration Reproductive Toxicity Study

MOA – Mode-of-Action

MOU – Memorandum of Understanding

NAS – National Academy of Sciences

NCGC – NIH Chemical Genomics Center

NIH – National Institutes of Health

NIEHS – National Institute of Environmental Health Sciences

NRC – National Research Council

NOAEL – No Observed Adverse Effect Level

NTP – National Toxicology Program

OECD – Organization for Economic Co-operation and Development

OPP – Office of Pesticide Programs

OPPTS – Office of Prevention, Pesticides and Toxic Substances

RED – Reregistration Eligibility Decision

SAR – Structure-Activity Relationship

ToxRefDB – Toxicity Reference Database

TSCA - Toxics Substance Control Act

USEPA – United States Environmental Protection Agency

CHAPTER 1

INTRODUCTION

Reproductive Physiology

Successful reproduction involves a complex orchestration of behavioral, physiological, biochemical, and molecular events requiring optimal timing and are driven primarily by the integration of signals along the hypothalamic-pituitary-gonadal (HPG) axis in both males and females (Perreault, 2008) (Figure 1.1). Normal male physiological processes revolve, primarily, around the testis which controls male development, sexual behavior, and reproductive capability. Specifically, the hypothalamus releases neuro-hormones including gonadotropin-releasing hormone (GnRH), stimulating the release of follicle stimulating hormone (FSH) and lutenizing hormone (LH) from the pituitary to regulate functions of seminiferous tubules (via Sertoli cells) and the interstitium (via the Leydig cell) in the testis, respectively (Johnson et al., 2010). Leydig cells in the testis are then responsible for the excretion of testosterone and secondary metabolites including estrogen, which provide regulatory feedback to the hypothalamus and pituitary, androgen- and estrogen-mediated effects on target organs, and regulate sexual development and function (Swerdloff et al., 2009). Spermatogenesis is one such biological process triggered by well orchestrated HPG signals as indicated by the balance of LH and testosterone in the serum. The mass balanced nature of the HPG axis, including both positive and negative feedback mechanisms, creates a compensatory system able to withstand ever changing dynamics from internal and external signals. However, chronic stressors and susceptible

developmental and life-stages can cause deleterious changes in sex hormone levels leading to sexual decrements.

Much of the same neuroendocrine control along the HPG axis regulates female reproduction and, in fact, the namesake for such hormones as FSH and LH were based on their function in the female reproductive system. The role of neuroendocrine function in females extends beyond that in males of reproductive tissue development, gametogenesis, and sexual behavior to also include menstrual cyclicity, offspring birthing and future development, and even milk production for lactation (Hoyer, 2010). Much like the testis, the ovary functions both as a gonad, developing and delivering oocytes, and as an endocrine gland producing important hormones such as inhibin, estradiol, and progesterone which allow the development of female secondary sexual characteristics and support pregnancy (McGee and Hsueh, 2000). Ovarian follicular development is induced by FSH followed by a surge of LH leading to the development of a mature follicle and release of the oocyte, but FSH is truly only indispensable to the final maturation process by rescuing a minority of follicles from atresia, or ovarian apoptosis (Chabbert Buffet et al., 1998; Richards et al., 1995). The majority of early ovarian follicular development have been shown to be independent of the larger HPG signaling processes with intra-ovarian paracrine signaling playing a major role (Richards and Pangas, 2010). Neuroendocrine regulation and ovarian function provide the necessary physiological environment with well choreographed processes to enable pregnancy, but the reproductive tract is the physical site of menstruation, implantation, pregnancy development and maintenance, and eventual birth. Alterations to the neuroendocrine system or reproductive tract can have profound effects on reproductive performance.

Chemically-induced Reproductive Toxicity

The HPG axis provides the key controls and functions to normal reproductive development and performance, although altering the HPG axis is not the only means to reproductive impairment and toxicity, including neuromuscular and systemic target organ toxicities. However, much of the research in molecular mechanism of reproductive toxicity in focused on the key molecular and hormonal targets within the HPG axis. Using animal models, a limited number of chemicals have well elucidated mechanisms of action in terms of effects on reproductive development or performance. For example, vinclozolin is a dicarboximide fungicide and recognized to exert antiandrogenic activity (Gray et al., 1994). The ability of vinclozolin to displace the androgen receptor's natural ligands, i.e., testosterone and dihydrotestosterone, leads to reduction in testosterone biosynthesis and disrupts androgen dependent mechanisms. The reduction in testosterone and disruption of androgen mediated mechanisms leads to reductions in sperm production, demasculinization, and sexual development, including delayed anogenital distance and preputial separation (Gray, et al., 1994; Kubota et al., 2003). As with many environmental chemicals, the reproductive toxicity profile includes additional endpoints related and unrelated to its antiandrogenic mode of action. Additionally, vinclozolin's antiandrogenic activity is not solely exerted in its parent form but generates metabolites with increased antiandrogenic potencies (Kelce et al., 1994). Vinclozolin provides a good example of the broad phenotypic consequences of disrupting key hormone levels that interact with many nodes in the HPG axis.

Prochloraz, a conazole fungicide, causes relatively specific effects on steroidogenesis and malformations in androgen-dependent tissues. In contrast to vinclozolin that causes

testosterone reduction via androgen receptor antagonism, prochloraz inhibits the conversion of progesterone to testosterone through the inhibition of CYP17, a key protein in the steroidogenic pathway (Blystone et al., 2007). Inhibition of CYP17 causes significant reductions in testosterone and prochloraz specifically inhibits steroidogenesis at the protein, not genomic, level (Laier et al., 2006; Vinggaard et al., 2002). The complexity and compensatory nature of the HPG axis make identifying the specific mechanism of action for particular chemicals difficult purely from a phenotypic perspective. The adverse outcomes of vinclozolin and prochloraz appear similar but in fact are driven primarily by two distinct mechanisms, androgen receptor antagonism and CYP17 inhibition, respectively. Androgen receptor antagonism appears to cause more direct effects on sexual development and differentiation with secondary effects on steroidogenesis, while inhibition of CYP17 and other steroidogenic pathway proteins cause primary effects on sperm production and secondary effects on sexual development via reduced testosterone.

Environmental chemicals have also been shown to cause male and female reproductive toxicity through direct cytotoxic actions in the reproductive tract. For example, 4-vinylcyclohexene and its diepoxide metabolite cause ovarian toxicity by damaging ovarian follicles. Additionally, 4-vinylcyclohexene causes ovarian cancer in long-term rodent studies and has demonstrated increased risk in humans (IARC, 1994). The primary mode of action for 4-vinylcyclohexene ovotoxicity is the induction of follicular apoptosis (Hoyer et al., 2001). 4-vinylcyclohexene is an example of a chemical acting primarily at a physiological level within the HPG axis. However, the relative sensitivity of 4-vinylcyclohexene can be mitigated through changes in metabolism, e.g., differences between rat and mouse, and through specific molecular oxidative stress sensors such as nuclear factor erythroid 2-related

factor 2 (Nrf2) (Hu et al., 2006). The wide range of modes and mechanisms of action leading to reproductive toxicity demonstrate the complexity of developing alternative toxicity tests that can efficiently evaluate a chemical complete reproductive toxicity potential. However, the diverse set of phenotypes that arise from the potentially finite numbers of mechanisms may make it possible to develop a battery of more efficient tests that adequately cover the known modes of action. As opposed to pure animal tests observing gross phenotypes, a battery of tests evaluating modes of action would better enable the evaluation of a chemical's human reproductive toxicity potential.

Human Reproductive Impairment

One in ten to one in six couples seek medical help due to subfertility with 20-25% of the problem being due to the male partner, 30-40% due to the female partner, roughly 30% due to potentially both partners, and roughly 15% with no specific factor identified (De Kretser and Baker, 1999). Infertility throughout most literature is described or defined as a couple who have never been able to become pregnant after at least one year of unprotected intercourse. Infertility statistics may be skewed based on the research demonstrating that about 25% of all pregnancies are lost prior to clinical recognition and were only detectable through urinary concentrations of human chorionic gonadotropin (hCG) (Wilcox et al., 1999; Wilcox et al., 1988). Therefore, the combination of male infertility, female infertility, and impaired fecundity (e.g., miscarriage or pregnancy loss) are all components in assessing overall human reproductive health decrements, but have diverse etiologies spanning genetic, disease, age, lifestyle, and environmental factors.

Estimating environmental/chemical-exposure contributions to human reproductive impairment and infertility has been shown to be difficult to assess due to many confounding

factors (Sharpe, 2000). For instance, male infertility, as compared to female infertility, may appear to be the more straightforward assessment because one can simply measure sperm count and quality as a major indicator of male infertility. Unfortunately, humans are naturally inefficient at spermatogenesis and have high inter- and intra-individual variability forcing studies to require high sample sizes and other study design considerations which make these studies less feasible and desirable (Sharpe, 2010). Nonetheless, lifestyle-related and environmental factors, such as smoking, obesity, traffic exhaust, dioxins, and combustion products, appear to negatively affect the perinatal and adult testes while there is suggestive, but generally unconfirmed, evidence that pesticide, food additive, persistent pollutant, and polychlorinated biphenyl exposure affect spermatogenesis in the general population (Sharpe, 2010). Many confounders, including age, lifestyle, and disease background must be accounted for in studying female fertility and fecundity levels due to environmental exposures. Recently a study of 1240 women found that environmental exposures to perfluorinated compounds significantly increased the time to pregnancy (Fei et al., 2009). These findings demonstrate that environmental relevant exposure levels to common industrial chemicals are linked to adverse reproductive performance in the general population. However, the ability to identify chemicals that affect human reproduction remains limited because of inadequate methodology spanning analytical chemistry to epidemiological measurements, small sample sizes, inappropriate endpoints, and the many confounders previously discussed and in many cases positive study findings are controversial and often are refuted in follow-up studies (Sharara et al., 1998). Epidemiological studies may be the most direct link to human relevant findings, but have shown limited success and only provide relevant information based on already exposed populations. The risk assessment

process attempts to characterize human risk of chemical exposure and the primary tool for doing such assessments are laboratory animal studies.

Reproductive Toxicity Testing

The chemical risk assessment process differs across regulatory bodies and chemical use classes. In the case of most environmental chemicals, multiple studies are used to assess the reproductive cycle in full (Figure 1.2). Prenatal developmental toxicity studies generally evaluate systemic toxicities to the pregnant female and fetal malformations and survival. Chronic and cancer long –term bioassays have the capacity to evaluate post-reproductive health, specifically male and female reproductive organ cancers. Specific and primary to the assessment of reproductive toxicity potential, the multigeneration reproductive toxicity study used by USEPA in the pesticide registration process assesses the performance and integrity of the male and female reproductive systems in the rat (USEPA, 1996) including assessment of gonadal function, the estrous cycle, mating behavior, conception, gestation, parturition, lactation, weaning, and on the growth and development of the offspring (Figure 1.3). The multigeneration study also provides information about the effects of the test substance on neonatal morbidity, mortality, target organs in the offspring, and data on prenatal and postnatal developmental toxicity. Two historical test guidelines have been used for the multigeneration studies. Multigeneration studies according to the 1982 Reproductive and Fertility Effects guideline (USEPA, 1982) have been conducted on over 700 chemicals and submitted to EPA. Multigeneration studies according to the newer 1998 guideline (USEPA, 1998) have been conducted on over 90 chemicals and submitted. Information on data submissions to EPA was drawn from the Office of Pesticide Programs (OPP) Information Network- the OPPIN database (Bouve, 2002). The 1998 guideline was harmonized by EPA's

Office of Pollution Prevention & Toxic Substances (OPPTS) to meet testing requirements of the EPA's Office of Pollution Prevention and Toxics (OPPT) and OPP, as well as international guidelines published by the Organization for Economic Cooperation and Development (OECD). Both of the guidelines call for a two-generation study in which continuously treated male and female rats are mated to produce first generation offspring, and in turn the adult offspring are mated to produce a second generation (Figure 1.3.A). In the United States, only food-use pesticide active ingredients are run routinely through a two-generation reproductive toxicity study.

Industrial chemicals, regulated under Toxics Substance Control Act (TSCA), rarely have a reproductive toxicity test performed and most studies found for industrial chemicals have been performed independent of TSCA (Aso et al., 2005; Fujii et al., 2005; Tyl et al., 2008; York et al., 2010). The European Union's Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH) regulation and potential TSCA reform may change the amount of reproductive toxicity tests routinely run on industrial chemicals, but the efficiency, practicality, and capacity to run thousands of industrial chemicals through such time-consuming and expensive tests has been questioned (Hartung and Rovida, 2009). The pre-clinical chemical safety assessment of pharmaceuticals with respect to reproductive toxicity have been historically divided into three segments with segment one and three addressing fertility and pre-/post-natal toxicity, respectively (Lumley, 1991). The International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) has evaluated the three-segment strategy and numerous proposals for revised testing protocols (Barrow, 2009; Lumley, 1991). A similar effort to revise the two-generation reproductive protocol used for most environmental

chemicals has been under evaluation and attempts to reduce animal use by going to an extended one-generation protocol and using cohorts to evaluate developmental neurotoxicity and immunotoxicity potential (Piersma et al., 2010). The two-generation study data has been used to retrospectively evaluate the ability of a one-generation protocol to adequately assess a chemical's reproductive toxicity potential by comparing classes of effects across the first and second generations (Figure 1.3B). Adapting and progressing reproductive toxicity testing study designs will reduce, to some extent, animal use and provide additional toxicological information. However, these new study designs do not and will not address the chemical testing bottleneck. The United States Environmental Protection Agency (EPA) has identified roughly 10,000 environmentally relevant chemicals that have been or may need to be assessed for their human exposure and toxicity potential (Judson et al., 2009). Past and current traditional reproductive toxicity testing has tested roughly 1,000 environmental chemicals in forty years. This testing strategy trajectory cannot be maintained, while claiming to be protective of human health and is being recognized around the world with regulatory changes such as the European Union's Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH) regulation and research initiatives such as the interagency program called Tox21 (Collins et al., 2008). The basic premise is to have all chemicals evaluated for toxicity potential, but the methods for arriving at this goal is less well defined in the regulatory context. Research initiatives, such as Tox21 and the EPA's ToxCast program, have proposed large scale high-throughput screening assays to probe biological pathways to ascertain a chemical's toxicity potential using computational modeling.

Predictive Toxicology using Computational and Alternative Methods

The medical and health community have been in search for indicators of human disease since the dawn of medicine. From demons and bad omens, to fevers and rash, to key measurement devices like the blood pressure cuff or x-ray machine, to the heart attack decision tree, these indicators of disease range from being associative, symptomatic, diagnostic, or prognostic. Just as twentieth century medicine was not focused on prevention, human disease indicators were not generally predictive. Two key scientific advancements have begun to change this. Advancements in molecular biology and computer science provide physicians, scientists, and researchers the tools to become preventative and predictive. Much like what the blood pressure cuff did for accurate measurement of blood pressure and thus the ability to causally associate high blood pressure with increased risk of heart disease and other adverse health outcomes, the mapping of the human genome has produced an entire generation of biological assays that allow for the accurate measurement of gene, protein, and cellular changes. These advancements have changed all health research fields, including drug discovery, epidemiology and toxicology.

Toxicology has the appearance of being reluctant to incorporating advances in molecular biology and computer science into its everyday practice and from a regulatory perspective this may be the case. On the other hand, toxicology does not have the luxury of testing adverse chemicals effects to humans in most cases and highly observational animal studies have been the primary measurement tool of toxicologists. Therefore, species extrapolation and mechanistic understanding have been primary themes in toxicology leading to the use of advanced molecular biology tools including genomics. Additionally, there has been an emphasis on reducing animal use in testing and increasing testing efficiency, which

has led to the use of advanced computational tools such as machine learning in structure activity relationship (SAR) modeling. The major difference between toxicology and other medical science disciplines is that toxicology has not taken advantage of the combined power of molecular biology and computer science. The focus from observational and descriptive toxicology has changed to an emphasis on mechanistic, predictive and high-throughput toxicology. Incorporating information and lessons from the last fifty years of toxicology into modern computing and biological tools include traditional toxicology studies being captured in databases, mechanistic studies, dose response modeling, and structure-based modeling.

Toxicity Databases

Therefore, computational toxicology is an emerging field that combines *in vitro* and computationally-generated data on chemicals, information on biological targets (genes, proteins), pathways and processes, and informatics methods to model and understand the mechanistic basis of chemical toxicity (Judson, 2011). One of the key informatics methods applied is relational database and knowledgebase creation capable of storing, aggregating and analyzing large previously unsearchable data and information. The amount of reference toxicity information on environmental chemicals, including primary studies, study reviews and summarized reports, quickly diminishes beyond pesticide active ingredients, Integrated Risk Information System (IRIS) chemicals, NTP nominated chemicals and a few other sources (Judson et al., 2008; Judson, et al., 2009). Efforts to digitize and structure the vast stores of open literature and unpublished industry-submitted studies have provided the information in a context amenable to using the data as anchoring endpoints for predictive toxicology applications. The currently available chemical-induced toxicity databases vary widely in breadth and depth of information (Bitsch et al., 2006; Knudsen et al., 2009;

Martin et al., 2009a; Martin et al., 2009b; Richard et al., 2008; Yang et al., 2006). IRIS is a good example of a database that has large content, covering greater than 500 chemicals and multiple toxicities, but lacks the standardization and detailed relational structure to provide accurate and efficient read-across (USEPA, 2011). ToxML, and related Food and Drug Administration (FDA) databases, ToxRefDB (U.S. EPA's Toxicity Reference Database), and REPDOSE are examples of relational formats that currently house hundreds of chemicals and multiple study types in a standardized format using controlled vocabularies (Bitsch, et al., 2006; Martin, et al., 2009a; Yang, et al., 2006). Web accessible toxicological data sources have been previously characterized (Felsot, 2002; Gold et al., 1991; Russom, 2002; Wolfgang and Johnson, 2002). These internet resources range from food and drug toxicity to environmental and ecological toxicity. Some of the internet sources provide fairly detailed summaries from cancer-related and genotoxicity studies. However, the information from these various sources is dispersed across the internet and in a wide variety of formats. Systems such as TOXNET, DSSTox, ACToR and PUBCHEM have made many of these resources available in a compiled format able to be searched based on chemical structure (Judson, et al., 2008; USEPA, 2007). The utility of chemical toxicity databases range from regulatory retrospective analyses (Piersma, et al., 2010), toxicity data gap analyses (Judson, et al., 2009), species concordance (Gold, et al., 1991), and toxicity classification modeling (Judson et al., 2010a; Zhu et al., 2009). Chemical toxicity information from structured, relational and searchable databases has provided a renewable resource for previously unachievable, unrepeatable, or non-updatable analyses, as well as setting the stage for measurable impacts that computational toxicology can make on chemical toxicity testing.

Chemical Structure-based Modeling in Toxicology

In addition to and in conjunction with chemical toxicity database efforts, computational approaches using structure-based classification and regression methods, including quantitative structure activity relationship (QSAR) models, have pushed the boundary on the amount of data and information required to assess a chemical's toxicity potential. Can a chemical's toxicity potential be predicted only using chemical structure? Historically the answer has be 'no'; QSAR and other structure-based modeling approaches have had limited success in gaining regulatory acceptance or application due to three primary reasons; a lack of well annotated reference chemical toxicity information, the complexity and multi-modality of toxicity endpoints, and limited focus on modeling best practices especially with regard to model quality and applicability metrics. Recently, efforts to develop modeling best practices and workflow have enabled appropriate evaluation of QSAR models before and after publication. QSAR modeling efforts have too often focused on "introspective" indicators of model accuracy and quality (Tropsha and Golbraikh, 2007) and in most cases were explanatory in nature.

QSAR publications often internal measures of model accuracy and quality, but until recently, the "modern QSAR" era, little effort or emphasis has been placed on rigorous model validation, including external validation, determination of the domain of applicability, and targeted chemical selection (Golbraikh et al., 2003; Golbraikh and Tropsha, 2002; Tropsha and Golbraikh, 2007). There are many potential reasons for the increased attention to external validation, including larger chemical sets, expanded applications and journal publication standards, but nonetheless modern QSAR model validation should and has begun to focus on external validation. In general, current practices generally include a clear

definition of the training and test set, external validation set, measures of accuracy at all levels of model development, and a clearly defined domain of applicability.

Structure-activity relationships (SAR) have been used for a number of regulatory actions and decisions in the past, including but not limited to: chemical groupings (e.g., High Production Volume chemical categories), bridging of toxicity data (e.g., antimicrobial re-registration; iodine and iodine complexes were all registered together with a single reference dose using data on representative compounds), supporting in vivo outcome calls (e.g., cancer assessment, using results of chemicals in same class to bolster argument for cancer classification), and study design requirements (e.g., organophosphates required to perform AChE testing). These applications of SAR have been fairly limited to simplistic analogy models or based solely on structure similarity and are not generalizable or externally predictive of biological or toxicological activity. QSAR models, alone, may never be broadly used for predictive toxicology, but the experiences gained from years of research in developing classification models of toxicity, developing best modeling practices and identifying data gaps where predictive toxicology is most needed has and will continue to push forward computational toxicology.

Experimental Predictive Toxicology

Approaches for predicting toxicity using computational methods and alternative test data has advanced our understanding of the molecular basis of toxicity. Iconix's Drug Matrix® stored experimental information from genomic studies including detailed pathology and developed genomic signatures or classifiers predictive of toxicity (Fielden et al., 2005; Fielden and Kolaja, 2008; Ganter et al., 2005) and showed promise in predicting toxicities of environmental chemicals (Martin et al., 2007). Importantly, the use of reference toxicity

information was used in the development of the classifiers in all studies. Similar governmental efforts to create the data management tools for storing genomic and phenotypic information has created the computational environments for the analysis of large genomic datasets with corresponding toxicity or phenotypic data. NIEHS's Chemical Effects in Biological Systems (CEBS) has been developed to store diverse biological information resulting from various toxicity and biological studies (Waters et al., 2003; Waters et al., 2008; Xirasagar et al., 2006). Systems-based toxicology in the world of drug discovery and drug safety assessment has begun to take hold and used as a viable approach both early in the discovery process and later in assessing toxicological information (Mayne et al., 2006). Additional informatics approaches including pathway-based analysis linking *in vitro* assay results to drug label information and adverse effect data provided mechanistic insight into purported toxicities and side-effects of drugs (Fliri et al., 2005). Similar system-based and pathway-based approaches for toxicity prediction to limit the high attrition rate of pharmaceuticals in the pipeline have produced other tools and products (Apic et al., 2005). These analytical tools required extensive curation of the biological literature, expensive laboratory-based data generation, and resulted in large databases for storing the information. Systems-biology approaches, including *in vivo* genomic studies, can provide detailed biological mechanistic information, but often lack in throughput and animal efficiency. In contrast, QSAR has nearly unlimited throughput and requires no animals, but lacks the mechanistic underpinnings often required for regulatory acceptance. To begin to obtain the advantages of both worlds, computational toxicology has elicited the use of high-throughput *in vitro* biological assays, previously used heavily in lead drug discovery within the pharmaceutical industry (Bleicher et al., 2003; Mayr and Bojanic, 2009). The toxicological

community has turned the paradigm upside down by attempting to broadly profile the biological activity of a chemical to then develop fingerprints of activity predictive of toxicities and to better understand a chemical's mechanism of action (Dix et al., 2007; Kavlock et al., 2008). Combining *in vitro* biological assay data into the QSAR modeling framework has shown improvement in predicting *in vivo* toxicological outcomes (Zhu et al., 2008) and serves as a good example of integrating data across many domains.

In Vitro Assays in Toxicology

In vitro biological assays have existed for many years and have been run routinely in the toxicological testing process, including the Ames mutagenicity assay (Ames, 1973; Ames et al., 1973a; Ames et al., 1973b). *In vitro* assay systems have advanced primarily due to innovations and advancements in gene/protein sequencing, computer /robotic systems, and cell culturing capacity. These innovations have enabled the creation of two focus areas of *in vitro* systems, high throughput screening (HTS) and high content screening (HCS). HTS generally focuses on techniques that efficiently measure chemical interactions with proteins or cells using 96-, 384-, or 1536-well plates in single concentration screening or concentration response formats. HTS assays can be characterized on being cell-free or cell-based, the cell type when applicable, the intended target or targets, and the readout. HTS is often intended to provide a specific response output covering a specific gene, protein, pathway or cell while testing hundreds to thousands of chemicals. The National Institutes of Health (NIH) Chemical Genomics Center (NCGC) took HTS one step further by created a robotic and assay platform system capable of screening thousands of chemicals in 12-15 concentration response format, coined Quantitative HTS (qHTS) and using further analysis automation to process the outputted data (Inglese et al., 2006; Inglese et al., 2007). In

contrast, HCS focuses on the dynamics (e.g., measurements over time and space) of the biological system and are almost exclusively cell-based or model organism-based (e.g., zebrafish) systems primarily using imaging techniques (Bullen, 2008; Giuliano et al., 2005; Giuliano et al., 2006). As computational power and robotics systems become cheaper and better, HTS and HCS will continue to merge with HTS investigating highly dynamic systems across time and producing highly multiplexed results and HCS increasing the throughput capacity and the molecular specificity of the target readouts. In addition to running assays in concentration response, the U.S. EPA's ToxCast research program has focused on covering as much biology as possible with commercially available assays and those assays available through internal and external collaborations (Judson, et al., 2010a). The product of running hundreds of diverse assays across thousands of chemicals in full concentration response format is millions of data points that require management, analysis and synthesis.

The field of toxicology has specialized in dose response modeling and has characterized dose response relationships across many toxicities and chemicals (Andersen et al., 2005; Cox, 1987; Leroux et al., 1996; Setzer et al., 2001; Swenberg et al., 2008), but the vast amount of data generated in HTS and the need for generalizable and extensible approaches across datasets remains a challenge for the toxicological community. Limited information is available on HTS dose response modeling best practices and appropriate dose response modeling techniques to be used for large HTS datasets (Parham et al., 2009). The U.S. EPA ToxCast program has released the fifty-percent activity concentrations (AC50), full concentration response, and summary calls for over 500 assays and provided both detailed and summary data analysis methodologies (Judson et al., 2010b). Generally, a four-parameter hill model can be fit to the data producing a model that estimates the baseline

response (B), maximal or top response (T), the slope of the response (W), and the AC50. Pre-processing, parameter constraining, confounding, statistical and biological considerations all play a role into the final set of conditions used to define active or inactive concentrations of a particular chemical-assay combination. Specific considerations include, but are not limited to; data normalization methods, positive and negative control performance and availability, baseline response adjustments, plate or well variations, response saturation (e.g., full versus partial agonist), active concentration extrapolation, outlier detection and removal, response directionality, non-monotonic response detection, cytotoxicity filtering and interpretation, fluorescence and other artifact detection, and statistical versus biological significance. Taking into account many of these considerations enables summary statistical outputs to be compared appropriately across assays and assay technologies and to be used in downstream modeling applications.

HTS and HCS, in the toxicological modeling context, requires the integration of results across a broad biological activity profile as compared to specific targeted HTS assay results. There have been limited efforts toward the development of models predictive of toxicity solely using bioactivity profiling data, primarily due to the fact that public research programs such as ToxCast and Tox21 programs are still in early stages of data generation and subsequent model development. Additionally, the assay coverage for specific biological pathways linked to toxicities, where known, varies when using commercially available vendors, because many of the assays were developed in the pharmaceutical industry and, for instance, have greater coverage of cancer pathways due to drug discovery efforts in cancer therapeutics. One focus area in environmental toxicology that helped progress assay development has been in the area of endocrine disruption, due in large part to the regulatory

creation of the Endocrine Disrupter Screening Program (EDSP) within the U.S. EPA based on the 1996 Food Quality and Protection Act (FQPA) and the Safe Drinking Water Act (SDWA). EDSP mandates environmental chemicals be evaluated for endocrine disruption and therefore a testing battery was developed resulting in large scale assay development and validation efforts spanning *in vitro* and *in vivo* tests (Goldman et al., 2000; Harding et al., 2006). Assay coverage of biological pathways is ever increasing and may enable computational modeling in areas that have either been understudied, such as immunotoxicity, or have proven to be difficult from a modeling perspective, such as reproductive toxicity.

Reproductive Toxicity Alternative Testing Methods

Reproductive toxicity is a prime example of a toxicity that has had limited effort and success in predictive toxicity, due in part to the lack of reference data in which to model and the physiological complexity of general reproductive impairment. One resource for predictive models have come from structure based methods (i.e., Quantitative Structure Activity Relationship (QSAR) models) and the accuracy and predictivity of the resultant models has been limited. A comprehensive effort toward the prediction of reproductive and developmental toxicity was undertaken by the FDA (Matthews *et al.*, 2007). The resultant QSAR models were developed for endpoints such as sperm effects, female reproductive toxicity and male reproductive toxicity and were generally highly specific models with an average specificity across all generated models being 88%. Based on the summary statistics extracted from the publications, the average balanced accuracy across all models was 58%, with the maximum balanced accuracy for any single model being 68% for trans-species female reproductive toxicity. It is difficult to assess the true accuracy and forward

predictivity of the models based solely on the summary statistics. Most likely, the limitation lies in the physiological complexity of reproductive toxicity and structural diversity of reproductive toxicants. Additional international efforts are underway with the goal of using alternative testing approaches in the detection of reproductive toxicants and on limited chemical sets have shown promise (Schenk *et al.*, 2010). Previous international efforts such as ReproTect have shed light on the need for alternative test methods and predictive models for reproductive toxicity (Hareng et al., 2005), which focuses both on developmental and reproductive toxicity. The following table summarizes alternative test methods for assessing reproductive toxicity, specifically areas of fertility and fecundity, male and female reproductive tract, and reproductive development, as is studied in the two-generation reproductive test most often used for environmental chemical reproductive toxicity assessment (Table 1.1). Currently, no *in vitro* test or set of tests would be considered an adequate alterative to the two-generation reproductive study and, in most cases, would not drastically increase testing efficiency. The goal of the following research is to develop a predictive model of general reproductive toxicity using bioactivity profiling data from HTS assays for applications towards chemical testing prioritization and with the eventual capacity to be used in an integrated testing strategy for reproductive toxicity testing.

Tables

Table 1.1. Alternative reproductive toxicity test methods, their target gender, current maximal assay throughput and example reference(s) are summarized showing the limited high-throughput assay systems currently available.

Alternative Test Method	Target Gender	System Type	Through-put	Reference
In Vitro Fertilisation assay (IVF)	Female	Ex Vivo/ In Vitro	Low	(Berger et al., 2000)
Ovarian Follicle BioAssay (FBA)	Female	In Vitro	Low	(Wan et al., 2010)
Granulosa and Theca Cell Culture Systems	Female	In Vitro	Low	(Albertini and Akkoyunlu, 2010)
In Vitro Maturation assay (IVM)	Female	In Vitro	Low	(Luciano et al., 2010)
Mouse Ovarian Tumor Cell Assay	Female	In Vitro	Low	(Blumenthal et al., 1988)
Ovarian/Follicle Culture Systems	Female	In Vitro	Low	(Cortvrindt and Smitz, 2002)
Human Ovarian Granulose-Like Tumor Cell Line (KGN) culture assay	Female	In Vitro	Medium	(Morinaga et al., 2004)
Uterotrophic Assay	Female	In Vivo	Low	(Kim et al., 2005)
Binding Assays (Estrogen and Androgen Receptor)	Male & Female	In Vitro	High	(Freyberger et al., 2010)
Transactivaton Assays (Cell-based Hormone Receptor Activation)	Male & Female	In Vitro	High	(Peekhaus et al., 2003)
Amphibian metamorphosis assay	Male & Female	In Vivo	Low	(Gutleb et al., 2007)
Estrogen-/Androgen-responsive cell proliferation assays (e.g., MCF-7)	Male & Female	In Vitro	High	(Lykkesfeldt and Sorensen, 1992)
Extended one-generation	Male & Female	In Vivo	Low	(Piersma, et al., 2010)

Pubertal Assay (Male and Female)	Male & Female	In Vivo	Low	(Marty et al., 1999; Marty et al., 2001)
Sperm Chromatin Structure Assay and related *in vitro* DNA damage assays	Male	Ex Vivo/ In Vitro	Low	(Evenson and Wixon, 2005)
Sperm Activation Assay	Male	Ex Vivo/ In Vitro	Low	(Sawyer and Brown, 1995)
Sperm Motility Assay	Male	Ex Vivo/ In Vitro	Low	(Betancourt et al., 2006)
Neutral Comet Assay in Sperm (ReProComet)	Male	In Vitro	Low	(Cordelli et al., 2007)
Ishikawa cell test	Male	In Vitro	Medium	(Schaefer et al., 2010)
H295R Steroidogenesis assay	Male	In Vitro	Medium	(Hecker and Giesy, 2008)
Leydig/Sertoli/Seminiferous Tubule/Germ Cell Culture & Co-culture systems	Male	In Vitro	Medium	(Brun et al., 1991)
Hershberger Assay	Male	In Vivo	Low	(Shin et al., 2007)

Figures

Figure 1.1 Overview of the key controls and functions of the hypothalamic-pituitary-gonadal axis, including gonadotropin-releasing hormone (GnRH), follicle stimulating hormone (FSH), and lutenizing hormone (LH) controls of testosterone (T), progesterone and estradiol production leading to gametogenesis in the female ovary and male testis.

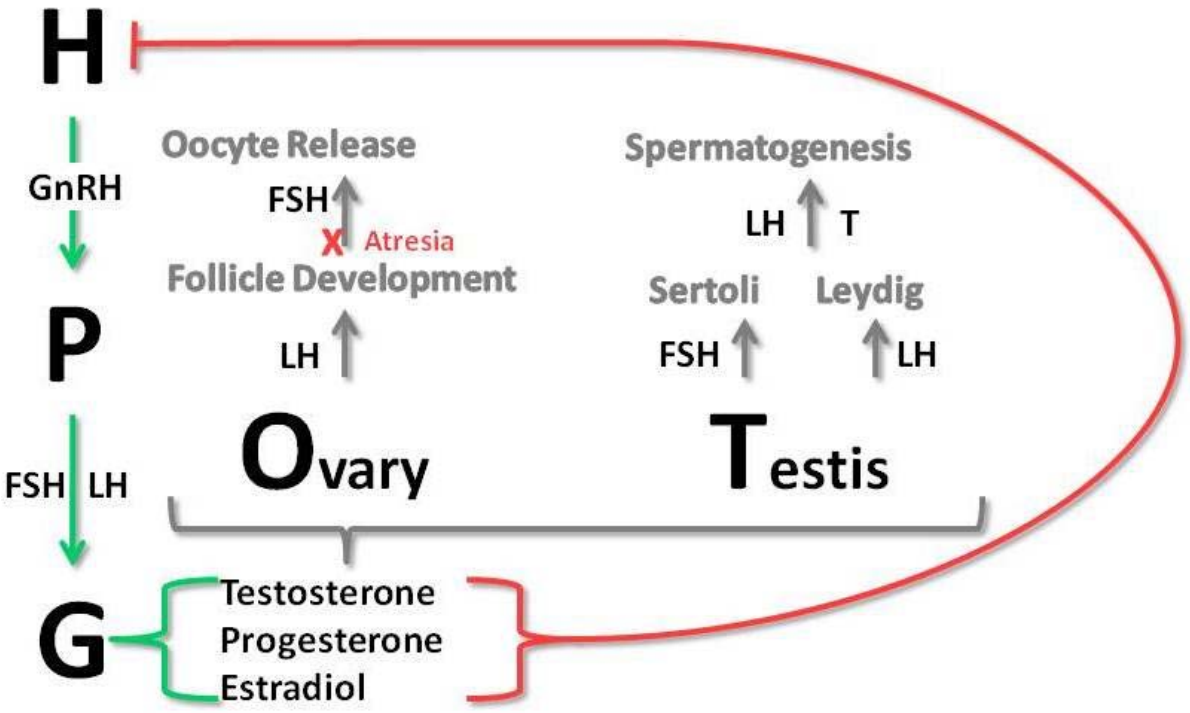


Figure 1.2. Schematic of the reproductive cycle in comparison to guideline toxicity studies performed most often for environmental chemical regulation. The multigeneration reproductive toxicity test spans portion of fetal development through juvenile development and back to the mating male and female and the pregnant female. In contrast, the prenatal developmental toxicity test focuses on the pregnant female and fetal development, particularly malformations. The chronic/cancer bioassay does identify systemic and carcinogenic reproductive organ effects.

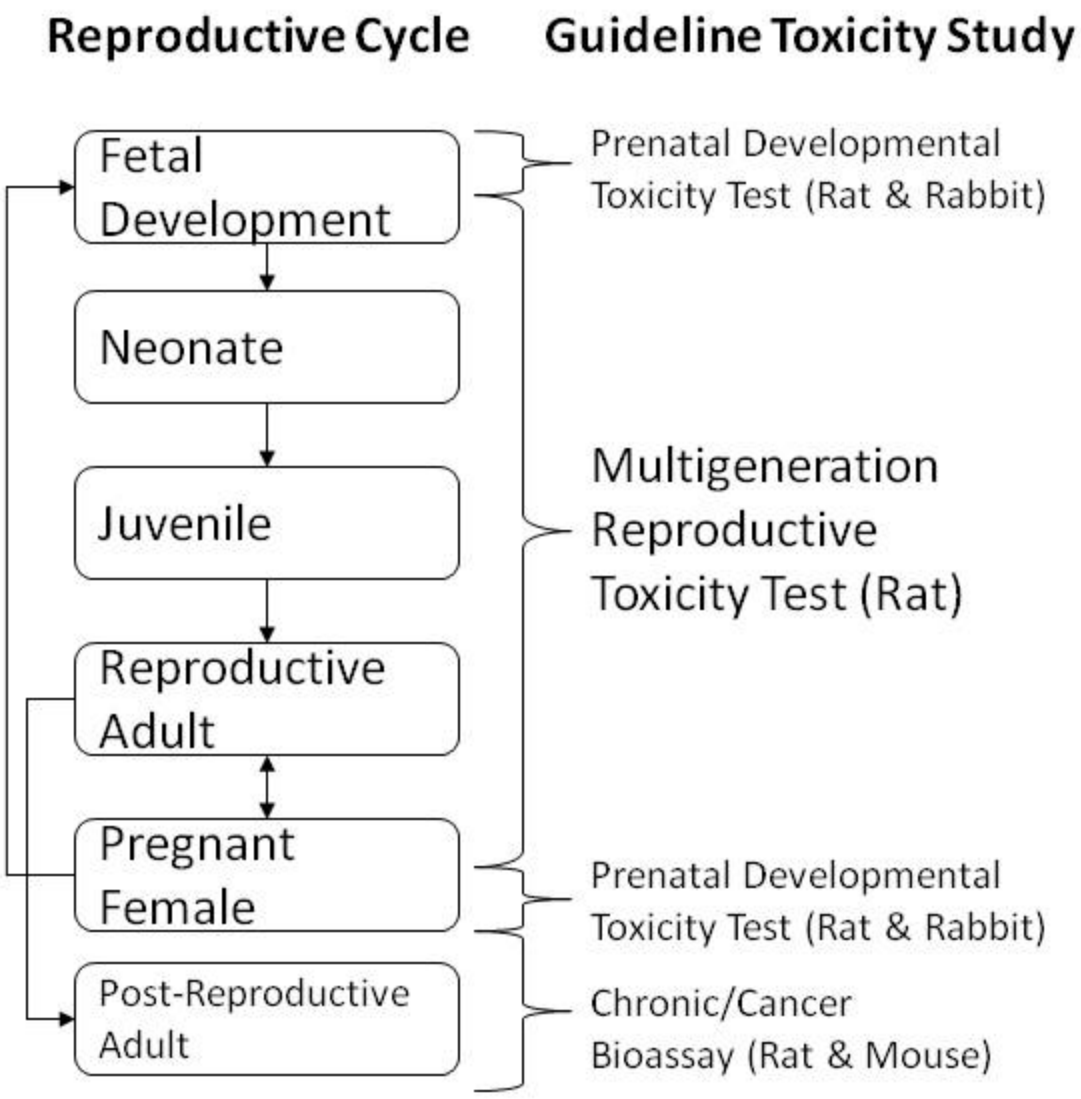
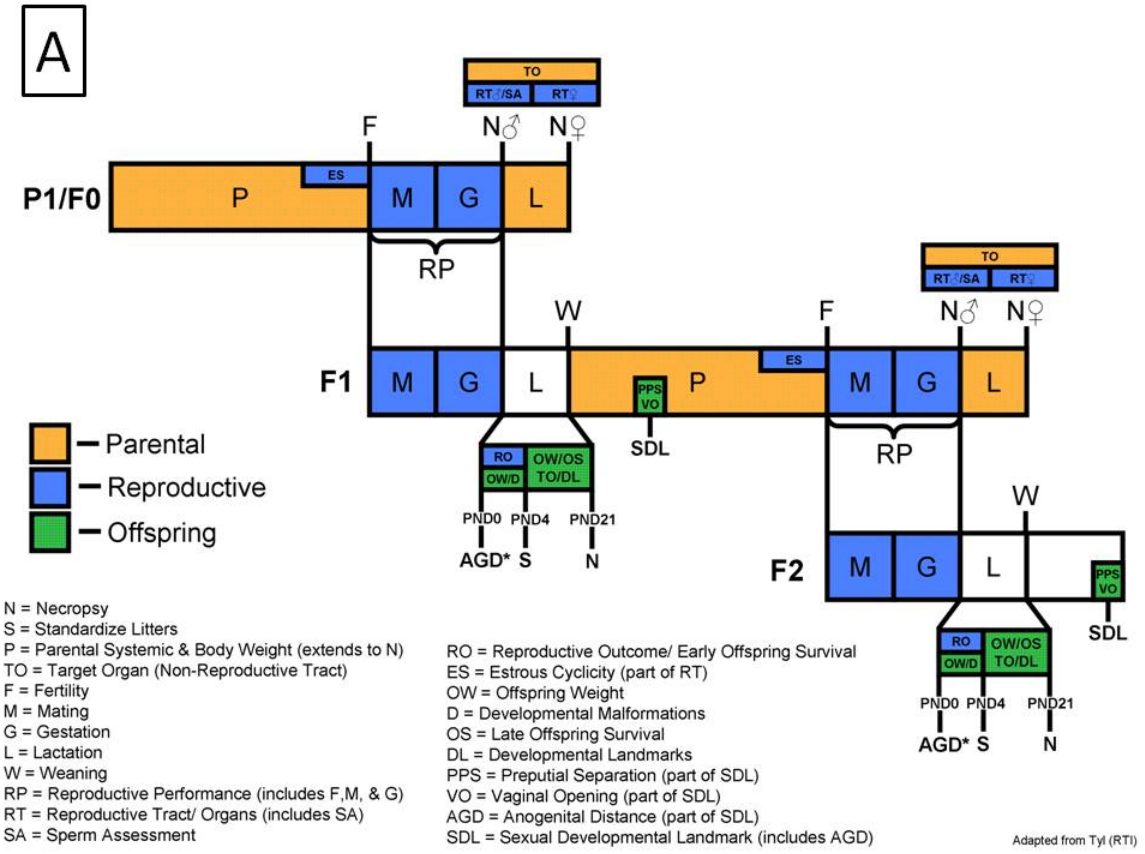
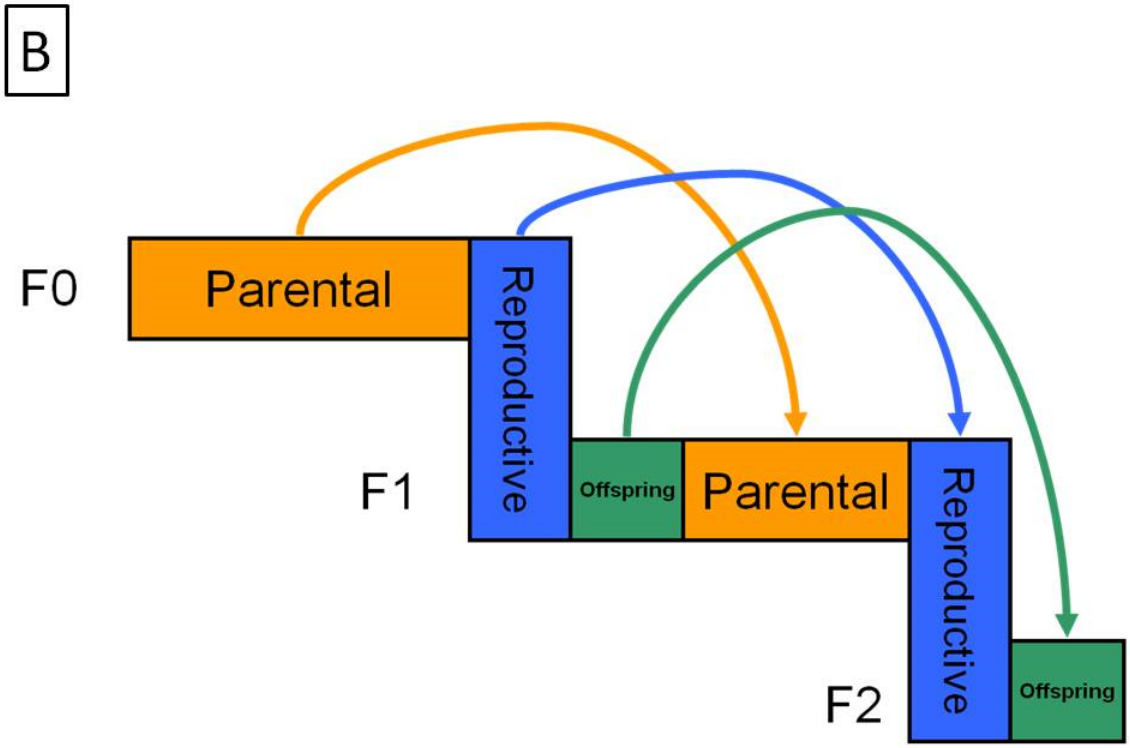


Figure 1.3. (A) Schematic of the two-generation reproductive study design mapped to terms used in ToxRefDB, in particular the definition and separation of parental, offspring and reproductive effects. (B) To simplify downstream analyses, terms were often binned based on the high level categories and compared across generations.


A
TO
RT♂/SA
RT♀
F
N♂
N♀
P1/F0
P
ES
M
G
L
RP
W
F1
PPS
VO
SDL
RO
OW/OS
OW/D
TO/DL
PND0
PND4
PND21
AGD*
S
N
F2
Parental
Reproductive
Offspring
N = Necropsy
S = Standardize Litters
P = Parental Systemic & Body Weight (extends to N)
TO = Target Organ (Non-Reproductive Tract)
F = Fertility
M = Mating
G = Gestation
L = Lactation
W = Weaning
RP = Reproductive Performance (includes F,M, & G)
RT = Reproductive Tract/ Organs (includes SA)
SA = Sperm Assessment
RO = Reproductive Outcome/ Early Offspring Survival
ES = Estrous Cyclicity (part of RT)
OW = Offspring Weight
D = Developmental Malformations
OS = Late Offspring Survival
DL = Developmental Landmarks
PPS = Preputial Separation (part of SDL)
VO = Vaginal Opening (part of SDL)
AGD = Anogenital Distance (part of SDL)
SDL = Sexual Developmental Landmark (includes AGD)
Adapted from Tyl (RTI)


B
F0
Parental
Reproductive
F1
Offspring
Parental
Reproductive
F2
Offspring

References

Albertini, D. F. and Akkoyunlu, G. (2010). Ovarian follicle culture systems for mammals. *Methods Enzymol* **476**, 107-21, S0076-6879(10)76007-9 [pii] 10.1016/S0076-6879(10)76007-9.

Ames, B. N. (1973). Carcinogens are mutagens: their detection and classification. *Environ Health Perspect* **6**, 115-8.

Ames, B. N., Durston, W. E., Yamasaki, E. and Lee, F. D. (1973a). Carcinogens are mutagens: a simple test system combining liver homogenates for activation and bacteria for detection. *Proc Natl Acad Sci U S A* **70**, 2281-5.

Ames, B. N., Lee, F. D. and Durston, W. E. (1973b). An improved bacterial test system for the detection and classification of mutagens and carcinogens. *Proc Natl Acad Sci U S A* **70**, 782-6.

Andersen, M. E., Thomas, R. S., Gaido, K. W. and Conolly, R. B. (2005). Dose-response modeling in reproductive toxicology in the systems biology era. *Reprod Toxicol* **19**, 327-37, S0890-6238(04)00217-5 [pii] 10.1016/j.reprotox.2004.12.004.

Apic, G., Ignjatovic, T., Boyer, S. and Russell, R. B. (2005). Illuminating drug discovery with biological pathways. *FEBS Lett* **579**, 1872-7, S0014-5793(05)00225-5 [pii] 10.1016/j.febslet.2005.02.023.

Aso, S., Ehara, H., Miyata, K., Hosyuyama, S., Shiraishi, K., Umano, T. and Minobe, Y. (2005). A two-generation reproductive toxicity study of butyl benzyl phthalate in rats. *J Toxicol Sci* **30 Spec No.**, 39-58, JST.JSTAGE/jts/30.S39 [pii].

Barrow, P. C. (2009). Reproductive toxicity testing for pharmaceuticals under ICH. *Reprod Toxicol* **28**, 172-9, S0890-6238(09)00072-0 [pii] 10.1016/j.reprotox.2009.03.013.

Berger, T., Miller, M. G. and Horner, C. M. (2000). In vitro fertilization after in vivo treatment of rats with three reproductive toxicants. *Reprod Toxicol* **14**, 45-53, S0890-6238(99)00062-3 [pii].

Betancourt, M., Resendiz, A. and Fierro, E. C. (2006). Effect of two insecticides and two herbicides on the porcine sperm motility patterns using computer-assisted semen analysis (CASA) in vitro. *Reprod Toxicol* **22**, 508-12, S0890-6238(06)00061-X [pii] 10.1016/j.reprotox.2006.03.001.

Bitsch, A., Jacobi, S., Melber, C., Wahnschaffe, U., Simetska, N. and Mangelsdorf, I. (2006). REPDOSE: A database on repeated dose toxicity studies of commercial chemicals--A multifunctional tool. *Regul Toxicol Pharmacol* **46**, 202-10, S0273-2300(06)00097-3 [pii] 10.1016/j.yrtph.2006.05.013.

Bleicher, K. H., Bohm, H. J., Muller, K. and Alanine, A. I. (2003). Hit and lead generation: beyond high-throughput screening. *Nat Rev Drug Discov* **2**, 369-78, 10.1038/nrd1086 nrd1086 [pii].

Blumenthal, R. D., McLaughlin, W. H., Jordan, J. J., Cryan, E. V. and Bloomer, W. D. (1988). Mouse ovarian tumor cells: an experimental model for progestin-mediated radiotherapy. *J Surg Oncol* **39**, 96-100.

Blystone, C. R., Lambright, C. S., Howdeshell, K. L., Furr, J., Sternberg, R. M., Butterworth, B. C., Durhan, E. J., Makynen, E. A., Ankley, G. T., Wilson, V. S., Leblanc, G. A. and Gray, L. E., Jr. (2007). Sensitivity of fetal rat testicular steroidogenesis to maternal prochloraz exposure and the underlying mechanism of inhibition. *Toxicol Sci* **97**, 512-9, kfm055 [pii] 10.1093/toxsci/kfm055.

Bouve, K. S. (2002). Electronic data submission: Pilot efforts in the Office of Pesticide Programs, US Environmental Protection Agency. *Acs Sym Ser* **824**, 188-197.

Brun, H. P., Leonard, J. F., Moronvalle, V., Caillaud, J. M., Melcion, C. and Cordier, A. (1991). Pig Leydig cell culture: a useful in vitro test for evaluating the testicular toxicity of compounds. *Toxicol Appl Pharmacol* **108**, 307-20, 0041-008X(91)90120-4 [pii].

Bullen, A. (2008). Microscopic imaging techniques for drug discovery. *Nat Rev Drug Discov* **7**, 54-67, nrd2446 [pii]
10.1038/nrd2446.

Chabbert Buffet, N., Djakoure, C., Maitre, S. C. and Bouchard, P. (1998). Regulation of the human menstrual cycle. *Front Neuroendocrinol* **19**, 151-86, S0091302298901674 [pii].

Collins, F. S., Gray, G. M. and Bucher, J. R. (2008). Toxicology - Transforming environmental health protection. *Science* **319**, 906-907, DOI 10.1126/science.1154619.

Cordelli, E., Fresegna, A. M., D'Alessio, A., Eleuteri, P., Spano, M., Pacchierotti, F. and Villani, P. (2007). ReProComet: a new in vitro method to assess DNA damage in mammalian sperm. *Toxicol Sci* **99**, 545-52, kfm191 [pii]
10.1093/toxsci/kfm191.

Cortvrindt, R. G. and Smitz, J. E. (2002). Follicle culture in reproductive toxicology: a tool for in-vitro testing of ovarian function? *Hum Reprod Update* **8**, 243-54.

Cox, C. (1987). Threshold dose-response models in toxicology. *Biometrics* **43**, 511-23.

De Kretser, D. M. and Baker, H. W. G. (1999). Infertility in men: Recent advances and continuing controversies. *J Clin Endocr Metab* **84**, 3443-3450.

Dix, D. J., Houck, K. A., Martin, M. T., Richard, A. M., Setzer, R. W. and Kavlock, R. J. (2007). The ToxCast program for prioritizing toxicity testing of environmental chemicals. *Toxicol Sci* **95**, 5-12, kfl103 [pii]
10.1093/toxsci/kfl103.

Evenson, D. P. and Wixon, R. (2005). Environmental toxicants cause sperm DNA fragmentation as detected by the Sperm Chromatin Structure Assay (SCSA). *Toxicol Appl Pharmacol* **207**, 532-7, S0041-008X(05)00241-3 [pii]
10.1016/j.taap.2005.03.021.

Fei, C., McLaughlin, J. K., Lipworth, L. and Olsen, J. (2009). Maternal levels of perfluorinated chemicals and subfecundity. *Hum Reprod* **24**, 1200-5, den490 [pii]
10.1093/humrep/den490.

Felsot, A. S. (2002). WEB resources for pesticide toxicology, environmental chemistry, and policy: a utilitarian perspective. *Toxicology* **173**, 153-66, S0300483X0200029X [pii].

Fielden, M. R., Eynon, B. P., Natsoulis, G., Jarnagin, K., Banas, D. and Kolaja, K. L. (2005). A gene expression signature that predicts the future onset of drug-induced renal tubular toxicity. *Toxicol Pathol* **33**, 675-83, T06867K6170G4R68 [pii]
10.1080/01926230500321213.

Fielden, M. R. and Kolaja, K. L. (2008). The role of early in vivo toxicity testing in drug discovery toxicology. *Expert Opin Drug Saf* **7**, 107-10, 10.1517/14740338.7.2.107.

Fliri, A. F., Loging, W. T., Thadeio, P. F. and Volkmann, R. A. (2005). Analysis of drug-induced effect patterns to link structure and side effects of medicines. *Nat Chem Biol* **1**, 389-97.

Freyberger, A., Wilson, V., Weimer, M., Tan, S., Tran, H. S. and Ahr, H. J. (2010). Assessment of a robust model protocol with accelerated throughput for a human recombinant full length estrogen receptor-alpha binding assay: protocol optimization and intralaboratory assay performance as initial steps towards validation. *Reprod Toxicol* **30**, 50-9, S0890-6238(10)00003-1 [pii]
10.1016/j.reprotox.2010.01.001.

Fujii, S., Yabe, K., Furukawa, M., Hirata, M., Kiguchi, M. and Ikka, T. (2005). A two-generation reproductive toxicity study of diethyl phthalate (DEP) in rats. *J Toxicol Sci* **30 Spec No.**, 97-116, JST.JSTAGE/jts/30.S97 [pii].

Ganter, B., Tugendreich, S., Pearson, C. I., Ayanoglu, E., Baumhueter, S., Bostian, K. A., Brady, L., Browne, L. J., Calvin, J. T., Day, G. J., Breckenridge, N., Dunlea, S., Eynon, B. P., Furness, L. M., Ferng, J., Fielden, M. R., Fujimoto, S. Y., Gong, L., Hu, C., Idury, R.,

Judo, M. S., Kolaja, K. L., Lee, M. D., McSorley, C., Minor, J. M., Nair, R. V., Natsoulis, G., Nguyen, P., Nicholson, S. M., Pham, H., Roter, A. H., Sun, D., Tan, S., Thode, S., Tolley, A. M., Vladimirova, A., Yang, J., Zhou, Z. and Jarnagin, K. (2005). Development of a large-scale chemogenomics database to improve drug candidate selection and to understand mechanisms of chemical toxicity and action. *J Biotechnol* **119**, 219-44, S0168-1656(05)00151-3 [pii]
10.1016/j.jbiotec.2005.03.022.

Giuliano, K. A., Cheung, W. S., Curran, D. P., Day, B. W., Kassick, A. J., Lazo, J. S., Nelson, S. G., Shin, Y. and Taylor, D. L. (2005). Systems cell biology knowledge created from high content screening. *Assay Drug Dev Technol* **3**, 501-14, 10.1089/adt.2005.3.501.

Giuliano, K. A., Johnston, P. A., Gough, A. and Taylor, D. L. (2006). Systems cell biology based on high-content screening. *Methods Enzymol* **414**, 601-19, S0076-6879(06)14031-8 [pii]
10.1016/S0076-6879(06)14031-8.

Golbraikh, A., Shen, M., Xiao, Z., Xiao, Y. D., Lee, K. H. and Tropsha, A. (2003). Rational selection of training and test sets for the development of validated QSAR models. *J Comput Aided Mol Des* **17**, 241-53.

Golbraikh, A. and Tropsha, A. (2002). Predictive QSAR modeling based on diversity sampling of experimental datasets for the training and test set selection. *J Comput Aided Mol Des* **16**, 357-69.

Gold, L. S., Slone, T. H., Manley, N. B., Garfinkel, G. B., Hudes, E. S., Rohrbach, L. and Ames, B. N. (1991). The Carcinogenic Potency Database: analyses of 4000 chronic animal cancer experiments published in the general literature and by the U.S. National Cancer Institute/National Toxicology Program. *Environ Health Perspect* **96**, 11-5.

Goldman, J. M., Laws, S. C., Balchak, S. K., Cooper, R. L. and Kavlock, R. J. (2000). Endocrine-disrupting chemicals: prepubertal exposures and effects on sexual maturation and thyroid activity in the female rat. A focus on the EDSTAC recommendations. *Crit Rev Toxicol* **30**, 135-96, 10.1080/10408440091159185.

Gray, L. E., Jr., Ostby, J. S. and Kelce, W. R. (1994). Developmental effects of an environmental antiandrogen: the fungicide vinclozolin alters sex differentiation of the male rat. *Toxicol Appl Pharmacol* **129**, 46-52, S0041-008X(84)71227-0 [pii]
10.1006/taap.1994.1227.

Gutleb, A. C., Schriks, M., Mossink, L., Berg, J. H. and Murk, A. J. (2007). A synchronized amphibian metamorphosis assay as an improved tool to detect thyroid hormone disturbance by endocrine disruptors and apolar sediment extracts. *Chemosphere* **70**, 93-100, S0045-6535(07)00832-6 [pii]
10.1016/j.chemosphere.2007.06.048.

Harding, A. K., Daston, G. P., Boyd, G. R., Lucier, G. W., Safe, S. H., Stewart, J., Tillitt, D. E. and Van Der Kraak, G. (2006). Endocrine disrupting chemicals research program of the U.S. Environmental Protection Agency: summary of a peer-review report. *Environ Health Perspect* **114**, 1276-82.

Hareng, L., Pellizzer, C., Bremer, S., Schwarz, M. and Hartung, T. (2005). The integrated project ReProTect: a novel approach in reproductive toxicity hazard assessment. *Reprod Toxicol* **20**, 441-52, S0890-6238(05)00092-4 [pii]
10.1016/j.reprotox.2005.04.003.

Hartung, T. and Rovida, C. (2009). Chemical regulators have overreached. *Nature* **460**, 1080-1081, Doi 10.1038/4601080a.

Hecker, M. and Giesy, J. P. (2008). Novel trends in endocrine disruptor testing: the H295R Steroidogenesis Assay for identification of inducers and inhibitors of hormone production. *Anal Bioanal Chem* **390**, 287-91, 10.1007/s00216-007-1657-5.

Hoyer, P. B. (2010). Female Reproductive Toxicology. pp. 339-345. Elsevier, Oxford.

Hoyer, P. B., Devine, P. J., Hu, X., Thompson, K. E. and Sipes, I. G. (2001). Ovarian toxicity of 4-vinylcyclohexene diepoxide: a mechanistic model. *Toxicol Pathol* **29**, 91-9.

Hu, X., Roberts, J. R., Apopa, P. L., Kan, Y. W. and Ma, Q. (2006). Accelerated ovarian failure induced by 4-vinyl cyclohexene diepoxide in Nrf2 null mice. *Mol Cell Biol* **26**, 940-54, 26/3/940 [pii]
10.1128/MCB.26.3.940-954.2006.

IARC (1994). 4-vinylcyclohexene. *IARC Monogr Eval Carcinog Risks Hum* **60**, 347-59.

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. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 11473-8.

Inglese, J., Johnson, R. L., Simeonov, A., Xia, M., Zheng, W., Austin, C. P. and Auld, D. S. (2007). High-throughput screening assays for the identification of chemical probes. *Nature chemical biology* **3**, 466-79, nchembio.2007.17 [pii]
10.1038/nchembio.2007.17.

Johnson, L., Welsh Jr, T. H., Curley Jr, K. O. and Johnston, C. E. (2010). Anatomy and Physiology of the Male Reproductive System and Potential Targets of Toxicants. pp. 5-59. Elsevier, Oxford.

Judson, R., Richard, A., Dix, D., Houck, K., Elloumi, F., Martin, M., Cathey, T., Transue, T. R., Spencer, R. and Wolf, M. (2008). ACTOR - Aggregated Computational Toxicology Resource. *Toxicol Appl Pharm* **233**, 7-13, DOI 10.1016/j.taap.2007.12.037.

Judson, R., Richard, A., Dix, D. J., Houck, K., Martin, M., Kavlock, R., Dellarco, V., Henry, T., Holderman, T., Sayre, P., Tan, S., Carpenter, T. and Smith, E. (2009). The Toxicity Data Landscape for Environmental Chemicals. *Environ Health Persp* **117**, 685-695, Doi 10.1289/Ehp.0800168.

Judson, R. S., Houck, K. A., Kavlock, R. J., Knudsen, T. B., Martin, M. T., Mortensen, H. M., Reif, D. M., Rotroff, D. M., Shah, I., Richard, A. M. and Dix, D. J. (2010a). In Vitro Screening of Environmental Chemicals for Targeted Testing Prioritization: The ToxCast Project. *Environ Health Perspect* **118**, 485-92, 10.1289/ehp.0901392.

Judson, R. S., Houck, K. A., Kavlock, R. J., Knudsen, T. B., Martin, M. T., Mortensen, H. M., Reif, D. M., Rotroff, D. M., Shah, I., Richard, A. M. and Dix, D. J. (2010b). In Vitro Screening of Environmental Chemicals for Targeted Testing Prioritization: The ToxCast Project. *Environ Health Persp* **118**, 485-492, Doi 10.1289/Ehp.0901392.

Judson, R. S., Reif, D.M., Houck K. (2011). Computational Toxicology In Principles of Toxicology: Environmental and Industrial Applications, 3rd Edition [In Preparation].

Kavlock, R. J., Ankley, G., Blancato, J., Breen, M., Conolly, R., Dix, D., Houck, K., Hubal, E., Judson, R., Rabinowitz, J., Richard, A., Setzer, R. W., Shah, I., Villeneuve, D. and Weber, E. (2008). Computational toxicology - A state of the science mini review. *Toxicol Sci* **103**, 14-27, DOI 10.1093/toxsci/kfm297.

Kelce, W. R., Monosson, E., Gamcsik, M. P., Laws, S. C. and Gray, L. E., Jr. (1994). Environmental hormone disruptors: evidence that vinclozolin developmental toxicity is mediated by antiandrogenic metabolites. *Toxicol Appl Pharmacol* **126**, 276-85, S0041-008X(84)71117-3 [pii]
10.1006/taap.1994.1117.

Kim, H. S., Kang, T. S., Kang, I. H., Kim, T. S., Moon, H. J., Kim, I. Y., Ki, H., Park, K. L., Lee, B. M., Yoo, S. D. and Han, S. Y. (2005). Validation study of OECD rodent uterotrophic assay for the assessment of estrogenic activity in Sprague-Dawley immature female rats. *J Toxicol Environ Health A* **68**, 2249-62, G128388372141862 [pii]
10.1080/15287390500182354.

Knudsen, T. B., Martin, M. T., Kavlock, R. J., Judson, R. S., Dix, D. J. and Singh, A. V. (2009). Profiling the activity of environmental chemicals in prenatal developmental toxicity studies using the US EPA's ToxRefDB. *Reprod Toxicol* **28**, 209-219, DOI 10.1016/j.reprotox.2009.03.016.

Kubota, K., Ohsako, S., Kurosawa, S., Takeda, K., Qing, W., Sakaue, M., Kawakami, T., Ishimura, R. and Tohyama, C. (2003). Effects of vinclozolin administration on sperm production and testosterone biosynthetic pathway in adult male rat. *J Reprod Dev* **49**, 403-12.

Laier, P., Metzdorff, S. B., Borch, J., Hagen, M. L., Hass, U., Christiansen, S., Axelstad, M., Kledal, T., Dalgaard, M., McKinnell, C., Brokken, L. J. and Vinggaard, A. M. (2006). Mechanisms of action underlying the antiandrogenic effects of the fungicide prochloraz. *Toxicol Appl Pharmacol* **213**, 160-71, S0041-008X(05)00618-6 [pii] 10.1016/j.taap.2005.10.013.

Leroux, B. G., Leisenring, W. M., Moolgavkar, S. H. and Faustman, E. M. (1996). A biologically-based dose-response model for developmental toxicology. *Risk Anal* **16**, 449-58.

Luciano, A. M., Franciosi, F., Lodde, V., Corbani, D., Lazzari, G., Crotti, G., Galli, C., Pellizzer, C., Bremer, S., Weimer, M. and Modina, S. C. (2010). Transferability and inter-laboratory variability assessment of the in vitro bovine oocyte maturation (IVM) test within ReProTect. *Reprod Toxicol* **30**, 81-8, S0890-6238(10)00018-3 [pii] 10.1016/j.reprotox.2010.01.015.

Lumley, C. E. (1991). Proposal for international guidelines for reproductive and developmental toxicity testing for pharmaceuticals. *Adverse Drug React Toxicol Rev* **10**, 143-53.

Lykkesfeldt, A. E. and Sorensen, E. K. (1992). Effect of estrogen and antiestrogens on cell proliferation and synthesis of secreted proteins in the human breast cancer cell line MCF-7 and a tamoxifen resistant variant subline, AL-1. *Acta Oncol* **31**, 131-8.

Martin, M. T., Brennan, R. J., Hu, W. Y., Ayanoglu, E., Lau, C., Ren, H. Z., Wood, C. R., Corton, J. C., Kavlock, R. J. and Dix, D. J. (2007). Toxicogenomic study of triazole fungicides and perfluoroalkyl acids in rat livers predicts toxicity and categorizes chemicals based on mechanisms of toxicity. *Toxicol Sci* **97**, 595-613, DOI 10.1093/toxsci/kfm065.

Martin, M. T., Judson, R. S., Reif, D. M., Kavlock, R. J. and Dix, D. J. (2009a). Profiling Chemicals Based on Chronic Toxicity Results from the US EPA ToxRef Database. *Environ Health Persp* **117**, 392-399, Doi 10.1289/Ehp.0800074.

Martin, M. T., Mendez, E., Corum, D. G., Judson, R. S., Kavlock, R. J., Rotroff, D. M. and Dix, D. J. (2009b). Profiling the Reproductive Toxicity of Chemicals from Multigeneration Studies in the Toxicity Reference Database. *Toxicol Sci* **110**, 181-190, DOI 10.1093/toxsci/kfp080.

Marty, M. S., Crissman, J. W. and Carney, E. W. (1999). Evaluation of the EDSTAC female pubertal assay in CD rats using 17beta-estradiol, steroid biosynthesis inhibitors, and a thyroid inhibitor. *Toxicol Sci* **52**, 269-77.

Marty, M. S., Crissman, J. W. and Carney, E. W. (2001). Evaluation of the male pubertal onset assay to detect testosterone and steroid biosynthesis inhibitors in CD rats. *Toxicol Sci* **60**, 285-95.

Matthews, E. J., Kruhlak, N. L., Daniel Benz, R., Ivanov, J., Klopman, G. and Contrera, J. F. (2007). A comprehensive model for reproductive and developmental toxicity hazard identification: II. Construction of QSAR models to predict activities of untested chemicals. *Regul Toxicol Pharmacol* **47**, 136-55, S0273-2300(06)00174-7 [pii] 10.1016/j.yrtph.2006.10.001.

Mayne, J. T., Ku, W. W. and Kennedy, S. P. (2006). Informed toxicity assessment in drug discovery: systems-based toxicology. *Curr Opin Drug Discov Devel* **9**, 75-83.

Mayr, L. M. and Bojanic, D. (2009). Novel trends in high-throughput screening. *Curr Opin Pharmacol* **9**, 580-8, S1471-4892(09)00128-3 [pii] 10.1016/j.coph.2009.08.004.

McGee, E. A. and Hsueh, A. J. (2000). Initial and cyclic recruitment of ovarian follicles. *Endocr Rev* **21**, 200-14.

Morinaga, H., Yanase, T., Nomura, M., Okabe, T., Goto, K., Harada, N. and Nawata, H. (2004). A benzimidazole fungicide, benomyl, and its metabolite, carbendazim, induce aromatase activity in a human ovarian granulose-like tumor cell line (KGN). *Endocrinology* **145**, 1860-9, 10.1210/en.2003-1182 en.2003-1182 [pii].

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**, 1216-27, 1087057109349355 [pii] 10.1177/1087057109349355.

Peekhaus, N. T., Ferrer, M., Chang, T., Kornienko, O., Schneeweis, J. E., Smith, T. S., Hoffman, I., Mitnaul, L. J., Chin, J., Fischer, P. A., Blizzard, T. A., Birzin, E. T., Chan, W., Inglese, J., Strulovici, B., Rohrer, S. P. and Schaeffer, J. M. (2003). A beta-lactamase-dependent Gal4-estrogen receptor beta transactivation assay for the ultra-high throughput screening of estrogen receptor beta agonists in a 3456-well format. *Assay Drug Dev Technol* **1**, 789-800, 10.1089/154065803772613426.

Perreault, S. D., Klinefelter G.R., Clegg E. (2008). Assessment of Male Reproductive Toxicity. Informa Healthcare, USA.

Piersma, A. H., Rorije, E., Beekhuijzen, M. E., Cooper, R., Dix, D. J., Heinrich-Hirsch, B., Martin, M. T., Mendez, E., Muller, A., Paparella, M., Ramsingh, D., Reaves, E., Ridgway, P., Schenk, E., Stachiw, L., Ulbrich, B. and Hakkert, B. C. (2010). Combined retrospective analysis of 498 rat multi-generation reproductive toxicity studies: on the impact of parameters related to F1 mating and F2 offspring. *Reprod Toxicol*, S0890-6238(10)00351-5 [pii] 10.1016/j.reprotox.2010.11.013.

Richard, A. M., Yang, C. and Judson, R. S. (2008). Toxicity data informatics: supporting a new paradigm for toxicity prediction. *Toxicol Mech Methods* **18**, 103-18, 10.1080/15376510701857452.

Richards, J. S., Fitzpatrick, S. L., Clemens, J. W., Morris, J. K., Alliston, T. and Sirois, J. (1995). Ovarian cell differentiation: a cascade of multiple hormones, cellular signals, and regulated genes. *Recent Prog Horm Res* **50**, 223-54.

Richards, J. S. and Pangas, S. A. (2010). The ovary: basic biology and clinical implications. *J Clin Invest* **120**, 963-72, 10.1172/JCI41350
41350 [pii].

Russom, C. L. (2002). Mining environmental toxicology information: web resources. *Toxicology* **173**, 75-88, S0300483X02000239 [pii].

Sawyer, D. E. and Brown, D. B. (1995). The use of an in vitro sperm activation assay to detect chemically induced damage of human sperm nuclei. *Reprod Toxicol* **9**, 351-7, 0890623895000212 [pii].

Schaefer, W. R., Fischer, L., Deppert, W. R., Hanjalic-Beck, A., Seebacher, L., Weimer, M. and Zahradnik, H. P. (2010). In vitro-Ishikawa cell test for assessing tissue-specific chemical effects on human endometrium. *Reprod Toxicol* **30**, 89-93, S0890-6238(10)00019-5 [pii] 10.1016/j.reprotox.2010.02.002.

Schenk, B., Weimer, M., Bremer, S., van der Burg, B., Cortvrindt, R., Freyberger, A., Lazzari, G., Pellizzer, C., Piersma, A., Schafer, W. R., Seiler, A., Witters, H. and Schwarz, M. (2010). The ReProTect Feasibility Study, a novel comprehensive in vitro approach to detect reproductive toxicants. *Reprod Toxicol* **30**, 200-218, DOI 10.1016/j.reprotox.2010.05.012.

Setzer, R. W., Lau, C., Mole, M. L., Copeland, M. F., Rogers, J. M. and Kavlock, R. J. (2001). Toward a biologically based dose-response model for developmental toxicity of 5-fluorouracil in the rat: a mathematical construct. *Toxicol Sci* **59**, 49-58.

Sharara, F. I., Seifer, D. B. and Flaws, J. A. (1998). Environmental toxicants and female reproduction. *Fertility and Sterility* **70**, 613-622, Doi: 10.1016/s0015-0282(98)00253-2.

Sharpe, R. M. (2000). Lifestyle and environmental contribution to male infertility. *Br Med Bull* **56**, 630-42.

Sharpe, R. M. (2010). Environmental/lifestyle effects on spermatogenesis. *Philos T R Soc B* **365**, 1697-1712, DOI 10.1098/rstb.2009.0206.

Shin, J. H., Moon, H. J., Kang, I. H., Kim, T. S., Lee, S. J., Ahn, J. Y., Bae, H., Jeung, E. B. and Han, S. Y. (2007). OECD validation of the rodent Hershberger assay using three

reference chemicals; 17alpha-methyltestosterone, procymidone, and p,p'-DDE. *Arch Toxicol* **81**, 309-18, 10.1007/s00204-006-0174-1.

Swenberg, J. A., Fryar-Tita, E., Jeong, Y. C., Boysen, G., Starr, T., Walker, V. E. and Albertini, R. J. (2008). Biomarkers in toxicology and risk assessment: informing critical dose-response relationships. *Chem Res Toxicol* **21**, 253-65, 10.1021/tx700408t.

Swerdloff, R. S., Wang, C. and Sinha Hikim, A. P. (2009). Hypothalamic-Pituitary-Gonadal Axis in Men. pp. 2357-2393. Academic Press, San Diego.

Tropsha, A. and Golbraikh, A. (2007). Predictive QSAR modeling workflow, model applicability domains, and virtual screening. *Curr Pharm Des* **13**, 3494-504.

Tyl, R. W., Myers, C. B., Marr, M. C., Sloan, C. S., Castillo, N. P., Veselica, M. M., Seely, J. C., Dimond, S. S., Van Miller, J. P., Shiotsuka, R. N., Beyer, D., Hentges, S. G. and Waechter, J. M., Jr. (2008). Two-generation reproductive toxicity study of dietary bisphenol A in CD-1 (Swiss) mice. *Toxicol Sci* **104**, 362-84, kfn084 [pii] 10.1093/toxsci/kfn084.

USEPA (1982). OPP Guideline 83–4: Reproductive and Fertility Effects. Pesticide Assessment Guidelines, Subdivision F, Hazard Evaluation: Human and Domestic Animals. Office of Pesticides and Toxic Substances, Washington, DC.EPA–540/9–82–025

USEPA (1996). Reproductive Toxicity Risk Assessment Guidelines. *Fed. Regist* **61**, 56274–56322.

USEPA (1998). Health Effects Test Guidelines
OPPTS 870.3800 Reproduction and Fertility Effects. Office of Pesticides and Toxic Substances, Washington, DC.EPA 712–C–98–208

USEPA (2007). Distributed Structure-Searchable Toxicity (DSSTox) Database Network.http://www.epa.gov/ncct/dsstox/index.html

USEPA (2011). Integrated Risk Information System (IRIS) Homepage.http://www.epa.gov/IRIS/

Vinggaard, A. M., Nellemann, C., Dalgaard, M., Jorgensen, E. B. and Andersen, H. R. (2002). Antiandrogenic effects in vitro and in vivo of the fungicide prochloraz. *Toxicol Sci* **69**, 344-53.

Wan, X., Zhu, J., Zhu, Y., Ma, X., Zheng, Y., Wang, F., Liu, Z. and Zhang, T. (2010). Rat ovarian follicle bioassay reveals adverse effects of cadmium chloride (CdCl2) exposure on follicle development and oocyte maturation. *Toxicol Ind Health* **26**, 609-18, 0748233710375949 [pii] 10.1177/0748233710375949.

Waters, M., Boorman, G., Bushel, P., Cunningham, M., Irwin, R., Merrick, A., Olden, K., Paules, R., Selkirk, J., Stasiewicz, S., Weis, B., Van Houten, B., Walker, N. and Tennant, R. (2003). Systems toxicology and the Chemical Effects in Biological Systems (CEBS) knowledge base. *EHP Toxicogenomics* **111**, 15-28.

Waters, M., Stasiewicz, S., Merrick, B. A., Tomer, K., Bushel, P., Paules, R., Stegman, N., Nehls, G., Yost, K. J., Johnson, C. H., Gustafson, S. F., Xirasagar, S., Xiao, N., Huang, C. C., Boyer, P., Chan, D. D., Pan, Q., Gong, H., Taylor, J., Choi, D., Rashid, A., Ahmed, A., Howle, R., Selkirk, J., Tennant, R. and Fostel, J. (2008). CEBS--Chemical Effects in Biological Systems: a public data repository integrating study design and toxicity data with microarray and proteomics data. *Nucleic Acids Res* **36**, D892-900, gkm755 [pii] 10.1093/nar/gkm755.

Wilcox, A. J., Baird, D. D. and Weinberg, C. R. (1999). Time of implantation of the conceptus and loss of pregnancy. *N Engl J Med* **340**, 1796-9, 10.1056/NEJM199906103402304.

Wilcox, A. J., Weinberg, C. R., Oconnor, J. F., Baird, D. D., Schlatterer, J. P., Canfield, R. E., Armstrong, E. G. and Nisula, B. C. (1988). Incidence of Early Loss of Pregnancy. *New Engl J Med* **319**, 189-194.

Wolfgang, G. H. and Johnson, D. E. (2002). Web resources for drug toxicity. *Toxicology* **173**, 67-74, S0300483X02000227 [pii].

Xirasagar, S., Gustafson, S. F., Huang, C. C., Pan, Q., Fostel, J., Boyer, P., Merrick, B. A., Tomer, K. B., Chan, D. D., Yost, K. J., 3rd, Choi, D., Xiao, N., Stasiewicz, S., Bushel, P. and Waters, M. D. (2006). Chemical effects in biological systems (CEBS) object model for toxicology data, SysTox-OM: design and application. *Bioinformatics* **22**, 874-82, btk045 [pii] 10.1093/bioinformatics/btk045.

Yang, C., Benz, R. D. and Cheeseman, M. A. (2006). Landscape of current toxicity databases and database standards. *Curr Opin Drug Disc* **9**, 124-133.

York, R. G., Kennedy, G. L., Jr., Olsen, G. W. and Butenhoff, J. L. (2010). Male reproductive system parameters in a two-generation reproduction study of ammonium perfluorooctanoate in rats and human relevance. *Toxicology* **271**, 64-72, S0300-483X(10)00094-6 [pii] 10.1016/j.tox.2010.03.005.

Zhu, H., Martin, T. M., Ye, L., Sedykh, A., Young, D. M. and Tropsha, A. (2009). Quantitative structure-activity relationship modeling of rat acute toxicity by oral exposure. *Chem Res Toxicol* **22**, 1913-21, 10.1021/tx900189p.

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**, 506-13, 10.1289/ehp.10573.

CHAPTER 2

PROFILING PHENOTYPIC ACTIVITY OF CHEMICALS FROM MULTIGENERATION REPRODUCTION STUDIES IN TOXREFDB[1]

Abstract

Multigeneration reproduction studies are used to characterize parental and offspring systemic toxicity, as well as reproductive toxicity of pesticides, industrial chemicals and pharmaceuticals. Results from 329 multigeneration studies on 316 chemicals have been digitized into standardized and structured toxicity data within the Toxicity Reference Database (ToxRefDB). An initial assessment of data quality and consistency was performed prior to profiling these environmental chemicals based on reproductive toxicity potential and generated toxicity endpoints for predictive modeling within the EPA's ToxCast™ research program. Unsupervised hierarchical clustering of the lowest effect levels for 75 effects generated chemical-phenotype relationships and provided chemical clusters for predictive modeling. Comparative analysis across the 329 studies identified chemicals with sensitive reproductive effects, based on comparisons to chronic and subchronic toxicity studies, as well as cross-generational comparisons within the multigeneration study. The unsupervised clustering and comparative analyses identified 19 parental, offspring and reproductive effects with a high enough incidence to serve as targets for predictive modeling in ToxCast. These

[1] Previously published as Martin, M. T., Mendez, E., Corum, D. G., Judson, R. S., Kavlock, R. J., Rotroff, D. M. and Dix, D. J. (2009b). Profiling the Reproductive Toxicity of Chemicals from Multigeneration Studies in the Toxicity Reference Database. *Toxicol Sci* **110**, 181-190, DOI 10.1093/toxsci/kfp080.

toxicity endpoints included specific reproductive performance indices, male and female reproductive organ pathologies, offspring viability, and parental systemic toxicities. Capturing this reproductive toxicity data in ToxRefDB supports ongoing retrospective analyses, test guideline revisions, and computational toxicology research.

Introduction

The U.S. Environmental Protection Agency (EPA) and other regulatory agencies are investigating novel approaches for predicting chemical toxicity, with the goal of rapidly screening the thousands of environmental chemicals with limited toxicity data (Judson et al., 2008). Building predictive models of chemical toxicity requires high quality *in vivo* toxicity data, in order to develop and validate new *in vitro* and *in silico* approaches. In support of EPA's ToxCast™ predictive toxicology effort (Dix *et al.*, 2007), we have created the Toxicity Reference Database (ToxRefDB) for capturing information from *in vivo* toxicity studies. ToxRefDB includes endpoints from multiple study types, including chronic rat and mouse carcinogenicity two-year bioassays that have been previously reported and made publicly available (Martin *et al.*, 2009; USEPA, 2009b). ToxRefDB is being used to build computational models linking whole animal toxicity, and specific tissue and cellular phenotypes, to specific chemical-biological interactions detected by cellular, genomic and biochemical *in vitro* assays. The *in vivo* toxicity data captured in ToxRefDB is facilitating a transition to the National Research Council's vision for "Toxicity Testing in the 21st Century: A Vision and a Strategy" (Collins *et al.*, 2008; Gibson, 2010), by linking toxicity endpoints from animal studies to molecular targets and pathways relevant to humans.

The multigeneration study data entered into ToxRefDB provides anchoring *in vivo* reproductive toxicity data for the EPA ToxCast™ research program (USEPA, 2009a). Within

the ToxCast program, bioactivity profiles for hundreds of environmental chemicals are being derived from hundreds of *in vitro* assays (Dix, et al., 2007; Houck and Kavlock, 2008). Phase I of ToxCast is focused on chemicals with known *in vivo* toxicity data, supporting the development of *in vitro* data signatures predictive of these *in vivo* outcomes(Judson *et al.*, 2010). It is worth noting that for environmental chemicals, unlike pharmaceuticals, quantitative *in vivo* toxicity data is essentially restricted to animal species. Nearly all of the ToxCast Phase I chemicals are food-use pesticide active ingredients that have undergone numerous mammalian toxicity tests, including guideline multigeneration studies. This highly standardized dataset provided in ToxRefDB facilitates profiling ToxCast Phase I chemical toxicity based on parental, offspring and reproductive effects.

Traditional toxicity testing for the risk assessment of environmental compounds or groups of compounds can cost millions of dollars and take years of effort. Since 1970, EPA has accumulated a vast store of high quality regulatory toxicity information on hundreds of compounds, most of which has been inaccessible to computational analyses. The curation and structuring of this chemical toxicity information into ToxRefDB has created a valuable resource for both retrospective and prospective toxicological studies (Martin, et al., 2009). In addition to the chronic/cancer rat and cancer mouse studies and multigeneration studies reported here, we are also extracting developmental toxicity studies in the rat and rabbit. The multigeneration reproductive toxicity dataset- studies used by EPA in the pesticide registration process to assess the performance and integrity of the male and female reproductive systems (USEPA, 1996) include assessment of gonadal function, the estrous cycle, mating behavior, conception, gestation, parturition, lactation, weaning, and on the growth and development of the offspring. The multigeneration study also provides

information about the effects of the test substance on neonatal morbidity, mortality, target organs in the offspring, and data on prenatal and postnatal developmental toxicity.

Two historical test guidelines have been used for the multigeneration studies in ToxRefDB. Multigeneration studies according to the 1982 Reproductive and Fertility Effects guideline (USEPA, 1982) on over 700 chemicals have been conducted and submitted to EPA. Multigeneration studies according to the newer 1998 guideline (USEPA, 1998) on over 90 chemicals have been conducted and submitted, including 40 studies extracted into ToxRefDB. Information on data submissions to EPA was drawn from the Office of Pesticide Programs (OPP) Information Network- the OPPIN database. The 1998 guideline was harmonized by EPA's Office of Pollution Prevention & Toxic Substances (OPPTS) to meet testing requirements of the EPA's Office of Pollution Prevention and Toxics (OPPT) and OPP, as well as international guidelines published by the Organization for Economic Cooperation and Development (OECD). Both of the guidelines call for a two-generation study in which continuously treated male and female rats are mated to produce first generation offspring, and in turn the adult offspring are mated to produce a second generation.

Methods

Data Characteristics

Reviews of registrant-submitted multigeneration reproductive toxicity studies, known as Data Evaluation Records (DER), were collected for roughly 300 chemicals from EPA's Office of Pesticide Programs (OPP). File types of DER include TIFF, Microsoft Word, Word Perfect and PDF formats, some of which are not directly text-readable. Approximately 500 multigeneration reproductive toxicity DER were reviewed, and based on data quality a subset

of 329 were selected for curation into ToxRefDB. The first portion of the DER outlines the test substance, purity, lot/batch numbers, MRID (Master Record Identification), study citation, OPPTS test guideline (USEPA, 1982; USEPA, 1998) and reviewers of the study. The executive summary captures all of the basic study design information, including species and strain, doses, number of animals per treatment group and any deficiencies in study protocol. All dose levels were stored in ToxRefDB as 'mg/kg/day' and, where possible, recorded or calculated from food consumption data as an average over the entire dosing period. The executive summary also describes treatment-related effects observed at various dose levels in the study. The body of the DER provides detailed test material and animal information, and full dose response data in text and tables for all measured and observed endpoints. All treatment-related effects were captured for each study in ToxRefDB.

Multigeneration study DER contain all the information necessary to infer Lowest Effect Level (LEL) values for all treatment-related effects that were statistically or biologically significant. Typically the DER also designated 'critical' effects for each study, and Lowest Observed Adverse Effect Level (LOAEL) and No Observed Adverse Effect Level (NOAEL) for each study. If provided by the DER, ToxRefDB captured these study-level NOAEL, LOAEL and critical effect data. However, it is important to note that the critical effects used to establish NOAEL, LOAEL and a reference dose (RfD) for a conventional chemical pesticide active, and to make regulatory risk assessment and management decisions, are based on a toxicological review of multiple studies across many study types.

Treatment-related effects were further identified as either a 'Parental', 'Offspring', or 'Reproductive' effect. Consistent with DER, 'Parental' endpoints were defined as systemic

toxicity observed in the male or female adult parents, and exclude effects directly related to reproduction (e.g., reproductive organ toxicity). 'Offspring' endpoints were defined as systemic toxicity observed in the pre-weaning and juvenile animals, and exclude birthing indices up to post-natal day (PND) four (e.g., litter size and live birth index). 'Reproductive' endpoints were defined as observed effects on the reproductive performance or capacity of the animals and included all reproductive organ toxicities, effects on estrous cyclicity, sperm parameters, fertility, and mating, and pre-natal and early post-natal viability.

A small number of ToxCast Phase I chemicals were not pesticide active chemicals, such as some perfluorinated compounds and phthalates. Though DER and pesticide registration studies were not available for these chemicals, there was often high quality, standardized reproductive toxicity studies available from the National Toxicology Program, peer-reviewed literature, or other sources. When data from such studies were available, it was curated into ToxRefDB consistent with information taken from DER.

Data Model and Quality Control

The relational data model for ToxRefDB was previously described (Martin, et al., 2009) in a diagram showing the data model and field-level. A Data Entry Tool was developed for database population, including a controlled vocabulary for reproductive and other test data (available for download at http://actor.epa.gov/toxrefdb/). Additional data entry and quality control procedures for ToxRefDB have been previously described (Martin, et al., 2009).

Full descriptions of the available data and conclusions as to the potential for the pesticides to cause harm to humans or the environment, risk mitigation measures, and the regulation of pesticides can be found at U.S. EPA Office of Pesticide Programs website

(http://www.epa.gov/pesticides). The study-level critical effects captured in ToxRefDB and taken from individual DER and studies cannot be correlated directly to regulatory determinations or RfDs without additional information and analysis.

Data Output and Analysis

The structured toxicity information stored within ToxRefDB can be extracted in various formats utilizing SQL™ queries. For the purpose of providing computable outputs, i.e. quantitative outputs amenable to statistical analysis, a consistent data output was used. The cross-tabulated data output consisted of rows of chemical information (e.g., Chemical Abstracts Service Registry Number, chemical name), by columns of toxicity endpoints with the value entered being the lowest dose at which the endpoint was observed (i.e., LEL) in 'mg/kg/day'. Even though NOAEL/LOAEL values for each study's 'Parental', 'Offspring', or 'Reproductive' effect can be queried from the database, the current analyses for ToxCast only utilized LEL. Log transformed potency values were derived using $-\log_2$ of LEL. A constant value of 12 was then added to zero-center the data allowing for zero to represent no observed effect. Therefore, a value of 1 would be equivalent to an effect at 2048 mg/kg/day and 18 would be equivalent to 0.015625 mg/kg/day. The log transformed values are predominantly used in the current analysis. However, millimolar concentrations (mmol/kg/day) were calculated for each endpoint using the molecular weight of tested chemical and the LEL in mg/kg/day. The resulting data formats are highly amenable to statistical data analysis, including descriptive and predictive data mining algorithms. These data tables are available on the ToxRefDB homepage: http://www.epa.gov/ncct/toxrefdb/.

Unsupervised two-way hierarchical clustering across all the chemicals, of all effects with incidence greater than five, was carried out based on log transformed potency values

and using Pearson's dissimilarity measure for both chemicals and effects. This analysis used Ward's method for linkage and the agglomerative clustering method and was implemented in R version 2.6.1 (Gentleman and Hornik, 2002). Clusters of chemicals were identified based on a distance height cutoff of five. The proportion of positive effects in each respective cluster to the proportion of effects out of the respective cluster was used as a measure of the weight of an endpoint in deriving the cluster classes.

Results

Summary Characterization of Multigeneration Study Results

This analysis focused on reproduction-related endpoints culled from 329 multigeneration rat studies on 316 unique chemicals entered into ToxRefDB (Table 2.1). The vast majority of studies (294 of 329) were performed using a two-generation protocol. There were seven one-generation studies, for which four were supplementary studies to longer-term two- or three-generation studies. Of the 28 three-generation studies, only first and second generation effects were used in subsequent analyses while third generation effects were excluded. In total, there were 11 chemicals with more than one study in this dataset. Four chemicals had an additional study run to satisfy study guideline requirements. Two chemicals had an additional study to test at additional dose levels. Five chemicals had two studies performed at similar dose levels and the conclusions between each pair of studies were similar.

Across all studies and treatment groups 12,230 treatment-related effects were observed, corresponding to 458 different, unique types of effects. Nearly twice as many total effects were observed in the F1 generation compared to the P1 or F2 generations. However, the F1 generation includes both juvenile (i.e., offspring) and adult effects, whereas P1 and F2

generations only represent adult or juvenile life-stages, respectively. In order to enable cross-generation comparisons, the count of chemicals and treatment-related effects across life-stage, endpoint category, and generation are presented in Table 2.2. Parental effects were associated with 275 of the 316 chemicals for both the P1 and F1 generation, whereas reproductive effects were associated with only 100 or 129 chemicals in the P1 and F1 generations, respectively. Besides more chemicals, there were 73% more adult effects within the reproductive endpoint category in the F1 generation, than in the P1. A similar number of chemicals and offspring category effects were observed in the F1 and F2 generation.

Hierarchical Clustering for Identification of Chemical-Phenotype Classes

Identification of chemical clusters with similar profiles of phenotypic activity was achieved by unsupervised two-way hierarchical clustering of 75 target-level effects (Figure 2.1). These 75 effects were selected based on an occurrence with greater than five of the 316 chemicals. These were defined as target-level effects because specific descriptive terms were aggregated to the target organ (i.e., liver) or measured index (e.g., lactation index), rather than all possible outcomes for each target (e.g., hypertrophy, hyperplasia, degeneration, etc.). Six chemical clusters were identified based on a distance height cutoff of five. Each cluster description in Figure 2.1 is derived from the mostly heavily weighted endpoints in forming the respective chemical cluster and does not mean that every chemical in the cluster causes the endpoint. Cluster 1 consists of the 14 chemicals with no observed toxicities across the 75 effects in this analysis. Cluster 2 contains 115 chemicals for which general systemic toxicities are driving the formation of the cluster. Interestingly, this cluster is also heavily weighted with endpoints relating to sperm counts and morphology, endocrine-related organ pathologies and weight changes, and delays in sexual maturation. Of the 115 chemicals, all

five phthalate compounds in ToxRefDB are found in this cluster. Cluster 3 contains 63 chemicals with limited toxicity for which parental and offspring body weight changes are driving the formation of the cluster. Cluster 4 formation is heavily weighted with cholinesterase inhibition effects and is comprised of 12 organophosphorus compounds. Clusters 5 and 6 contain 48 and 64 chemicals, respectively, and the formation of these clusters are heavily weighted with reproductive toxicity endpoints, including testicular and epididymal pathologies in cluster 5 and offspring viability in cluster 6.

The complete listing of chemical clusters and endpoint weights for each cluster is available for download from the ToxRefDB homepage (USEPA, 2009b). The unsupervised approach clearly segmented the chemicals into distinct classes based on their profile of phenotypic activity and these cluster assignments are available as endpoints for predictive modeling. This analysis also guides the manual endpoint selection process by highlighting groups of chemical-phenotype relationships. Many of these associative differences are expected, but others are not. For instance, reproductive performance, reproductive organ and offspring viability effects were segregated slightly from each other and to a greater extent from parental systemic effects and even delays in sexual maturation.

Comparative Analysis with Chronic and Subchronic Systemic Toxicity

Parental, reproductive and offspring potencies (i.e., inverse log transformed LEL) from the multigeneration studies were compared to potency values for systemic toxicity from two-year chronic and 90-day subchronic studies in the rat (Figure 2.2). For this comparison, data were available in ToxRefDB for 254 chemicals tested in both multigeneration and 2-year chronic studies, and 207 chemicals tested in both multigeneration and 90-day subchronic studies. The potency values compared rarely correspond to the same treatment-

related effect across study type. For the majority of chemicals, potency values between the multigeneration, chronic and subchronic studies were comparable, with a general linear relationship falling within ten-fold of each other. However, for four chemicals (bisphenol A, deltamethrin, flucycloxuron, flufenpyr-ethyl) that caused parental or reproductive effects in the multigeneration study, there was no systemic toxicity observed in either the chronic or subchronic studies. For another five chemicals (cyprodinil, diethyltoluamide, difenoconazole, ethametsulfuron methyl, thiamethoxam) potencies for the most sensitive multigeneration endpoints were more than 10-fold greater than for the most sensitive effects in chronic studies. Of these five chemicals only thiamethoxam was more potent based solely on reproductive endpoints, i.e., testicular atrophy. Decreasing the threshold from 10-fold to a two-fold increase in potency resulted in 37, 7 and 20 chemicals more potent for parental, reproductive, or offspring endpoints, respectively. Of the seven chemicals identified as two-fold more potent reproductive toxicants, no reproductive organ toxicity was observed in the rat chronic/cancer or subchronic studies for these chemicals- the multigeneration test detected reproductive toxicity that could have been missed in chronic or subchronic studies. Under the conditions of the 2-year chronic studies, the vast majority of chemicals observed effects at lower doses than in the multigeneration reproductive study. However, even in these cases, the multigeneration test often identified selective reproductive toxicants and endpoints not detected in the chronic study.

Comparative Analysis of Parental, Reproductive and Offspring Endpoints

Chemicals with increased potency in the second generation were identified by comparing P1 and F1, or F1 and F2 LEL across parental, reproductive and offspring endpoint categories for 316 chemicals (Figure 2.3). Specific second generation effects (i.e., F1

parental or reproductive, F2 offspring) not observed in the first generation (i.e., P1 parental or reproductive, F1 offspring), or sensitive effects occurring at a lower LEL in the second generation are provided for all 316 chemicals on the ToxRefDB homepage (USEPA, 2009b). For parental effects, 15 chemicals had specific effects in the F1 versus P1, and another 48 were more sensitive in the F1 versus P1 based upon at least a two-fold difference in LEL. For reproductive toxicity endpoints, 52 chemicals had specific effects in the F1 versus P1, and another 14 were more sensitive in the F1 versus P1 based upon at least a two-fold difference in LEL. For offspring toxicity endpoints, 14 chemicals had specific effects in the F2 versus F1, and another 28 were more sensitive in the F2 versus F1 based upon at least a two-fold difference in LEL. Across all the effect categories, this came to a total of 137 chemicals that displayed specificity or sensitivity for one of the endpoint categories (i.e., parental, reproductive, or offspring) in the second generation. However, the F1 reproductive or F2 offspring LEL was the most sensitive LEL across all endpoint categories for only 16 of these 137 chemicals. This analysis in ToxRefDB has identified a subset of chemicals for ToxCast predictive modeling that may be more specific or potent reproductive toxicants. It is important to note that these values are LEL for all treatment related effects, and are not necessarily critical effects being used for determination of NOAEL/LOAEL.

Selected Multigeneration Study Endpoints for Predictive Modeling

Figure 2.4 presents the incidence and distribution by generation of effects on reproductive performance, reproductive organs, offspring viability, and parental systemic toxicities selected for ToxCast predictive modeling. Toxicity profiles from multigeneration studies on 316 chemicals were based on a diverse set of 19 selected effects or effect aggregations distributed in various combinations across the P1, F1 and F2 generations. A

detailed table listing all 19 of these endpoints for the 316 chemicals, including endpoint descriptions and various transformations of LEL values, is available for download from the ToxRefDB homepage (USEPA, 2009b). Treatment-related changes to reproductive performance including fertility, mating, gestational interval, implantations, litter size, and live birth index demonstrated effects at different stages of the reproductive cycle. Besides effects of many chemicals on offspring viability at PND4 and PND21 (viability and lactation indices, respectively), pubertal delays were also recorded for some chemicals. Pubertal delays were not part of the ToxCast modeling dataset because only a small subset of chemicals and studies assessed these endpoints. Effects on reproductive performance and offspring viability were observed in 110 (35%) and 108 (34%) of the 316 tested chemicals, respectively. Effects on reproductive organs, both organ weight and pathology, were observed in 98 (31%) of the chemicals with roughly 50% of those chemicals causing the effect only in the second generation (F1 adult). Of the 98 chemicals, 31 caused both male and female reproductive organ effects, 43 male only, and 24 female only. Systemic target organ weight and pathology endpoints were also selected, including the liver, kidney and spleen, along with the endocrine related adrenal, pituitary and thyroid glands.

Discussion

The objective of the ToxCast research program is to develop a cost-effective rapid approach for screening and prioritizing a large number of chemicals for toxicological testing (Dix, et al., 2007). Using data from high throughput screening (HTS) bioassays developed in the pharmaceutical industry, ToxCast is building computational models to forecast the potential human toxicity of chemicals. These hazard predictions should provide EPA regulatory programs, including OPP, with science based information helpful in prioritizing

chemicals for more detailed toxicological evaluations, and therefore lead to using fewer animal tests. ToxCast is currently in the proof-of concept phase, wherein over 300 chemicals have been assayed in over 600 different HTS bioassays, creating bioactivity profiles being used to derive signatures predicting the known toxicity for these chemicals. The Phase I chemicals are primarily conventional pesticide actives that have been extensively evaluated using traditional mammalian toxicity testing, and hence have known properties representative of a number of toxicity outcomes (e.g., carcinogenicity; and developmental, reproductive and neural toxicity). Thus a critical component of ToxCast is ToxRefDB, which is being populated with data from OPP for pesticide active chemicals and being extracted from the evaluations on these studies conducted by OPP scientists. Comparable toxicity data from other toxicity sources (e.g., National Toxicology Program) are also being captured in ToxRefDB. A broader and more diverse set of complementary data on thousands of chemicals is being captured in EPA's Aggregated Computational Toxicology Resource (Judson *et al.*, 2008).

The underlying data represented in ToxRefDB has been evaluated by EPA in prior pesticide registration decisions, and the presence of effects in high-dose animal studies do not translate directly into significant human risk stemming from registered uses of the pesticide. It should be noted that the EPA uses animal toxicology studies, like those entered into ToxRefDB, as well as other sources of information such as effects on wildlife populations, mechanisms of action, use patterns, environmental fate and persistence, food residue levels, and human exposure potential in its determinations to register pesticides, and to establish acceptable levels of pesticide residues for uses in the United States. While pesticide toxicity

data currently predominates in ToxRefDB, the database is being expanded to a broader range of chemicals, both by category and use.

The toxicity data in ToxRefDB and the HTS data generated in ToxCast will be made publicly available through EPA websites. The first component of ToxRefDB was recently published (Martin, et al., 2009), presenting toxicity profiles from two-year rodent bioassays on 310 chemicals. Multigeneration reproduction study data for 316 chemicals was entered into ToxRefDB in order to uniformly assess the reproductive toxicity potential of the chemical set and to profile chemical toxicity across generation, life-stage, and different classes of endpoints. The unsupervised two-way hierarchical clustering by chemical and effect identified chemical-phenotype relationships for which to characterize the chemical set. In the current analysis six chemical clusters were identified along with the relative endpoint weights for each cluster providing an indication of the composition of each cluster with respect to chemicals and effects. For predictive modeling, the phenotypic profile of these chemicals and resulting cluster sets could be used to match up with HTS bioactivity profiles in much the same manner. In the meantime, the resulting cluster sets guided the organization of the manual analysis and corroborated the distinction between parental, offspring and reproductive effects in subsequent analyses.

All 12,230 effects in the multigeneration study dataset were placed into three major classes of effects; parental, reproductive and offspring. The LEL for each class or category of effects were used to identify sensitive or specific reproductive toxicants based on comparisons to chronic and subchronic study data and cross-generational comparisons within the multigeneration reproductive test. In general, chemical exposures under conditions of the multigeneration reproduction study were less sensitive than under the conditions of the two-

year chronic study and comparable to the 90-day subchronic study. The analysis did, however, identify a subset of chemicals with sensitive or specific reproductive or offspring toxicities when compared to systemic effects under longer continuous exposure periods. Similar insight can be gleamed from comparing endpoints occurring at a lower dose or only in the second generation, i.e., second generation sensitive or specific effects, respectively. Effects that occur in the first generation and are not corroborated in the second generation can be questioned as to its toxicological relevance. Conversely, effects with consistent increases in second generation sensitivity or specificity might reflect the need for reproductive or developmental exposure to occur. Comparisons across these broad classes of endpoints honed in on specific effects for which to characterize the chemical set. The primary set of effects selected as anchoring endpoints for ToxCast predictive modeling were reproductive indices, offspring viability, and male and female reproductive organ effects, along with a set of parental systemic organ toxicities.

The ToxRefDB multigeneration study dataset has also played a key role in retrospective analyses by providing uniformly captured legacy reproductive toxicity data and by expanding the scope of questions that can be asked of the data. Multigeneration reproductive toxicity studies traditionally involve assessment through two generations and the value of the second generation is now being assessed for its regulatory impact. The analysis may also influence study design changes in subsequent guideline studies (Janer *et al.*, 2007). The current study focused on providing endpoints for predictive modeling as part of the ToxCast research program (Dix, et al., 2007), but also began to address the importance of specific study design parameters, including differences across generation, life-stage and various classes of endpoints. Additional analysis will be performed on this dataset in

collaboration with OPP and other international chemical regulatory agencies to expound upon the role of these and other study design parameters with respect to chemical regulation and potential guideline study design changes. This includes the potential to assess the ability of the current and previous multigeneration study guidelines to identify reproductive effects related to endocrine disruption. Fifty-three of the 73 chemicals proposed for screening in the Endocrine Disruptor Screening Program (EDSP; http://www.epa.gov/endo/pubs/prioritysetting/draftlist.htm) have multigeneration studies entered into ToxRefDB and are part of the ToxCast Phase I chemical set. Where available, multigeneration study data for the remaining chemicals are now being entered into ToxRefDB. A focused analysis of the EDSP chemical set would be just one example of the utility of ToxRefDB and the toxicity data stored within it. The use of ToxRefDB to address many research and regulatory science questions regarding *in vivo* mammalian toxicity not only provides transparency, but also assists in guiding the next set of questions.

The diverse utility of ToxRefDB as a reference database for research applications such as ToxCast demonstrates the power of curating toxicity information into a relational database. In the current analysis on the multigeneration reproductive toxicity test, six chemical clusters were derived using unsupervised methods, and subsequently nineteen specific endpoints were identified for anchoring endpoints in predictive modeling. These endpoints are further defined by life-stage or generation, and characterize the 316 chemicals with respect to their reproductive toxicity potential. Capturing this reproductive toxicity data in ToxRefDB supports ongoing retrospective analyses, test guideline revisions, and computational toxicology research.

Tables

Table 2.1. Summary statistics for 329 multigeneration reproduction studies on 316 chemicals within ToxRefDB

		Treatment Groups	Treatment Groups w/ Effects	Effects[a]	Unique Effects[b]
Total		7,869	3,239	12,230	458
Generation	P1	2069	902	3331	322
	F1	2,775	1306	6,329	400
	F2	2,717	945	2,396	144
	F3	308	86	174	31

(a) – Total No. of effects observed in any treatment group

(b) - No. of distinct effects observed in at least one treatment group

Table 2.2. Distribution of chemicals and effects across life-stage, endpoint category and generation for 316 chemicals in ToxRefDB with a multigeneration reproductive study

Life-Stage		Adult	Adult	Juvenile
Endpoint Category		Parental[c]	Reproductive[d]	Offspring[e]
Generation	P1	275[a] (2935)[b]	100 (376)	
	F1	275 (3265)	129 (648)	255 (2274)
	F2			247 (1979)

[a] - No. of chemicals with at least one effect observed at specified life-stage, endpoint category and generation

[b] - No. of effects observed at specified life-stage, endpoint category and generation

[c] - Parental endpoints include adult body weight, mortality, clinical signs and target-organ weight and pathology effects

[d] - Reproductive endpoints include reproductive organ weight and pathology and reproductive indices (e.g., fertility, mating, live birth index, etc.)

[e] - Offspring endpoints include pup weight, offspring survival (e.g., viability and lactation index), and juvenile target-organ weight and pathology, and pubertal delay (e.g., PPS and VO) effects

Figures

Figure 2.1. Unsupervised two-way hierarchical clustering of 75 treatment-related effects from multigeneration reproduction tests on 316 chemicals in ToxRefDB. Six chemical clusters were identified based on a distance height cutoff of five. Each cluster description is derived from the mostly heavily weighted endpoints (see Results and http://www.epa.gov/ncct/toxrefdb/ for details) forming the respective chemical cluster and does not mean that every chemical in the cluster causes the endpoint.

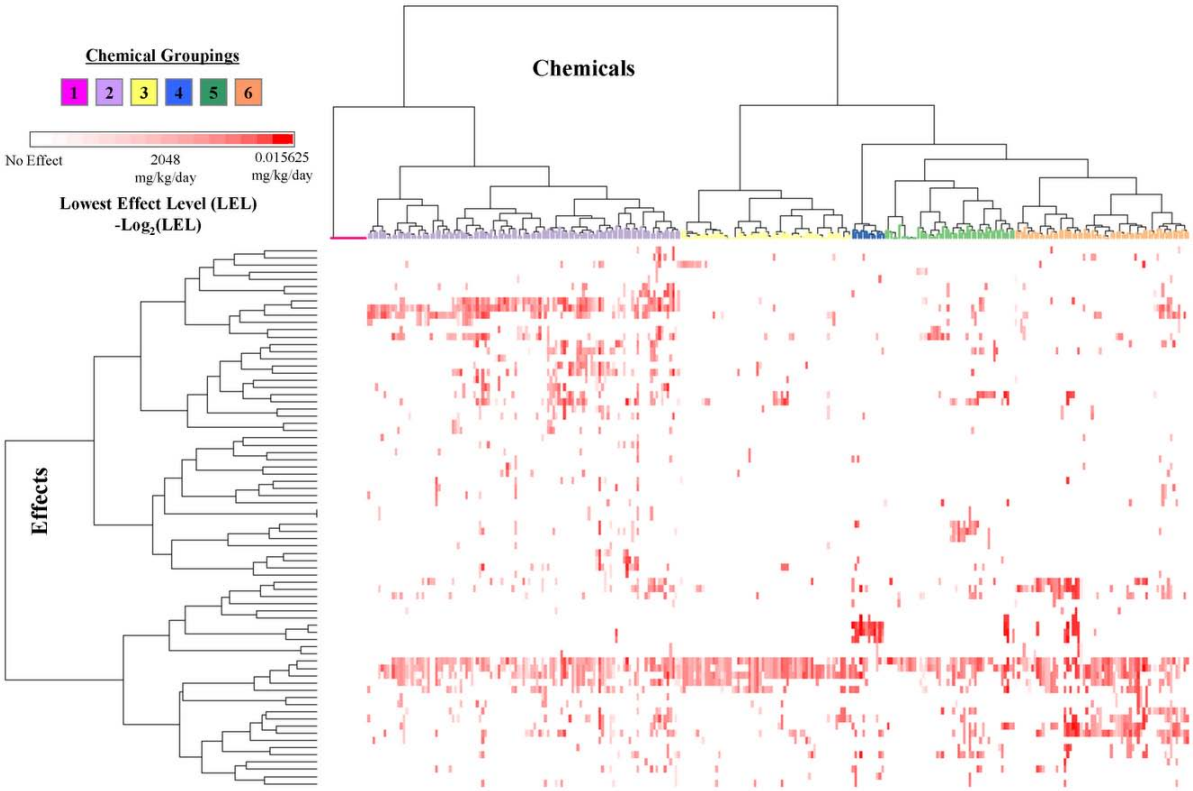


Figure 2.2. Parental, reproductive and offspring lowest effect levels (LEL; inverse log transformed) from multigeneration rat studies were compared to systemic LEL from chronic/cancer and subchronic rat studies for 254 and 207 chemicals, respectively. Points within gold lines indicate less than 2-fold difference between multigeneration and chronic studies. Points within orange lines indicate less than 10-fold difference between multigeneration and chronic studies. 'NE' stands for not established.

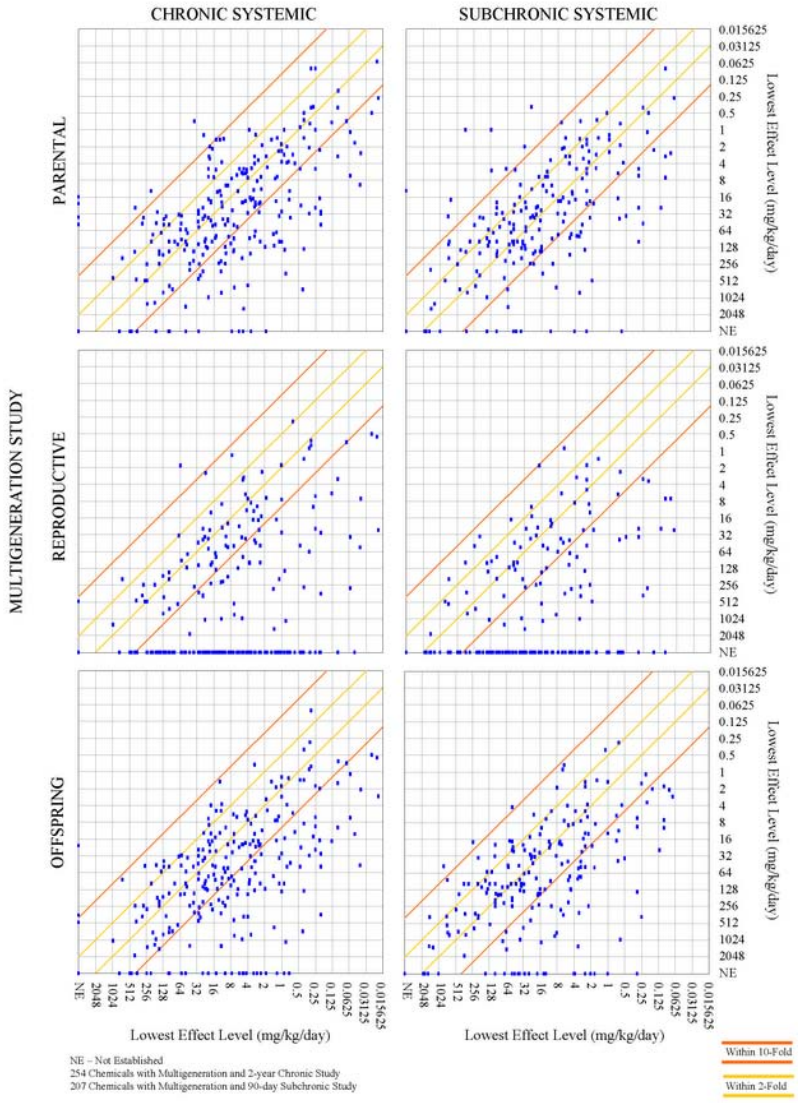


Figure 2.3. Comparing lowest effect levels (LEL) across generation and endpoint category. Points within dark orange lines indicate less than 2-fold difference between generations. Points within light orange lines indicate less than 10-fold difference between generations. 'NE' stands for not established.

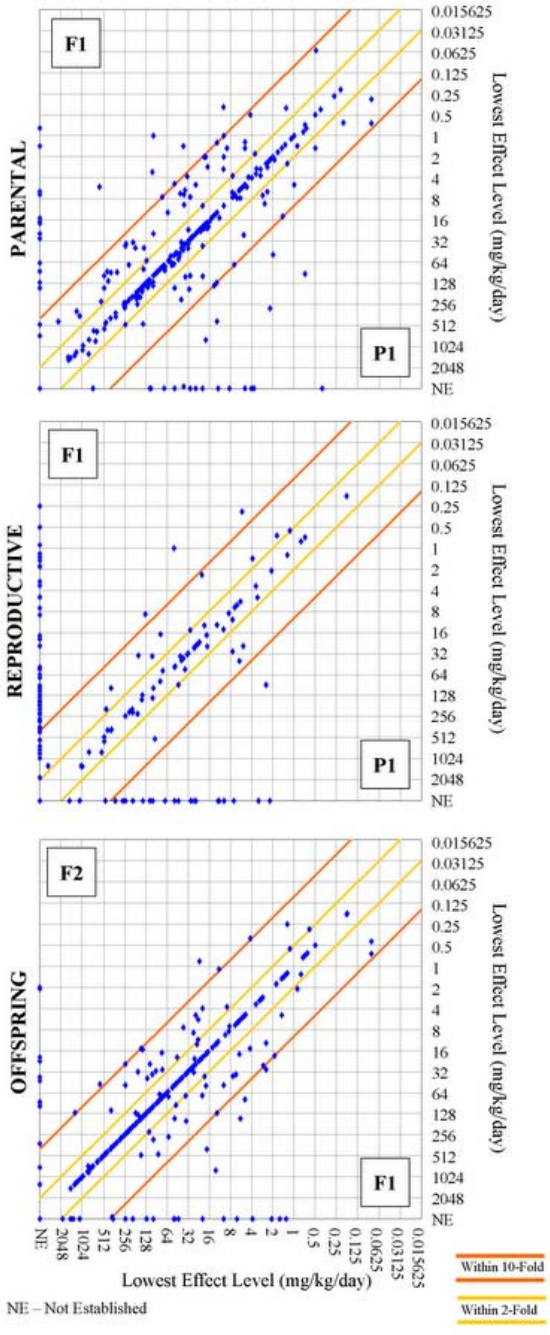
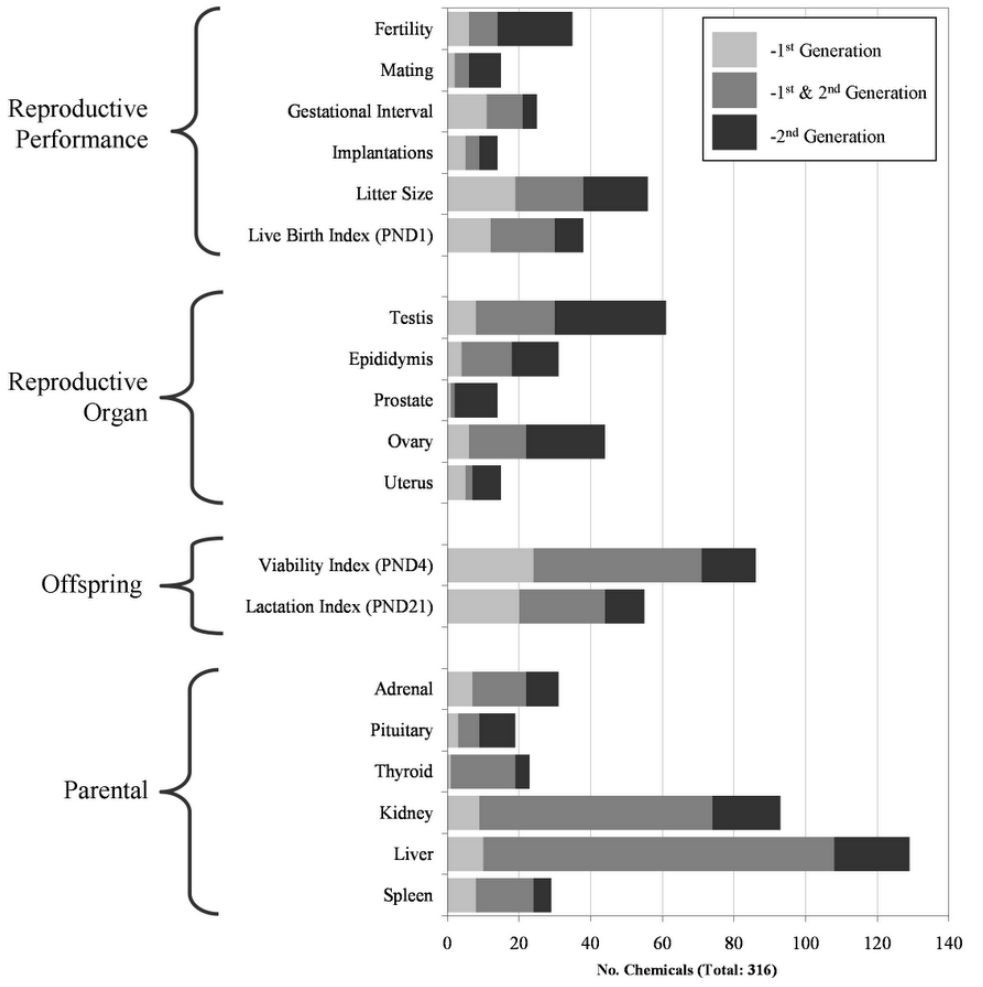


Figure 2.4. Incidence and distribution, by generation, of the 19 endpoints selected for predictive modeling, including reproductive, offspring and systemic toxicity endpoints from the rat multigeneration reproduction study (see Results and http://www.epa.gov/ncct/toxrefdb/ for details). The light gray bar indicates chemicals observing the endpoint only in the first generation, either P1 adult or F1 juvenile. The medium gray bar indicates chemicals observing the endpoint in both first and second generation treatment groups. The dark gray bar indicates chemicals observing the endpoint only in the second generation, either F1 adult or F2 juvenile.



References

Collins, F. S., Gray, G. M. and Bucher, J. R. (2008). Toxicology - Transforming environmental health protection. *Science* **319**, 906-907, DOI 10.1126/science.1154619.

Dix, D. J., Houck, K. A., Martin, M. T., Richard, A. M., Setzer, R. W. and Kavlock, R. J. (2007). The ToxCast program for prioritizing toxicity testing of environmental chemicals. *Toxicol Sci* **95**, 5-12, kfl103 [pii]
10.1093/toxsci/kfl103.

Gentleman, R. and Hornik, K. (2002). R 1.5 and the Bioconductor 1.0 releases. *Comput Stat Data An* **39**, 558-559.

Gibson, J. E. (2010). An Integrated Summary of Commentary on The National Academy of Sciences Report on "Toxicity Testing in the 21st Century: A Vision and a Strategy". *Hum Exp Toxicol* **29**, 33-35, Doi 10.1177/0960327109354659.

Houck, K. A. and Kavlock, R. J. (2008). Understanding mechanisms of toxicity: Insights from drug discovery research. *Toxicol Appl Pharm* **227**, 163-178, DOI 10.1016/j.taap.2007.10.022.

Janer, G., Hakkert, B. C., Slob, W., Vermeire, T. and Piersma, A. H. (2007). A retrospective analysis of the two-generation study: What is the added value of the second generation? *Reprod Toxicol* **24**, 97-102, DOI 10.1016/j.reprotox.2007.04.068.

Judson, R., Richard, A., Dix, D., Houck, K., Elloumi, F., Martin, M., Cathey, T., Transue, T. R., Spencer, R. and Wolf, M. (2008). ACTOR - Aggregated Computational Toxicology Resource. *Toxicol Appl Pharm* **233**, 7-13, DOI 10.1016/j.taap.2007.12.037.

Judson, R. S., Houck, K. A., Kavlock, R. J., Knudsen, T. B., Martin, M. T., Mortensen, H. M., Reif, D. M., Rotroff, D. M., Shah, I., Richard, A. M. and Dix, D. J. (2010). In Vitro Screening of Environmental Chemicals for Targeted Testing Prioritization: The ToxCast Project. *Environ Health Persp* **118**, 485-492, Doi 10.1289/Ehp.0901392.

Martin, M. T., Judson, R. S., Reif, D. M., Kavlock, R. J. and Dix, D. J. (2009). Profiling Chemicals Based on Chronic Toxicity Results from the US EPA ToxRef Database. *Environ Health Persp* **117**, 392-399, Doi 10.1289/Ehp.0800074.

USEPA (1982). OPP Guideline 83–4: Reproductive and Fertility Effects. Pesticide Assessment Guidelines, Subdivision F, Hazard Evaluation: Human and Domestic Animals. Office of Pesticides and Toxic Substances, Washington, DC.EPA–540/9–82–025

USEPA (1996). Reproductive Toxicity Risk Assessment Guidelines. *Fed. Regist* **61**, 56274–56322.

USEPA (1998). Health Effects Test Guidelines

OPPTS 870.3800 Reproduction and Fertility Effects. Office of Pesticides and Toxic Substances, Washington, DC.EPA 712–C–98–208

USEPA (2009a). ToxCast Homepage.http://www.epa.gov/ncct/toxcast

USEPA (2009b). ToxRefDB: Toxicity Reference Database.http://actor.epa.gov/toxrefdb

CHAPTER 3

IMPACT OF ENVIRONMENTAL CHEMICALS ON KEY TRANSCRIPTION REGULATORS AND CORRELATION TO TOXICITY ENDPOINTS WITHIN EPA'S TOXCAST™ PROGRAM[2]

Abstract

Exposure to environmental chemicals adds to the burden of disease in humans and wildlife to a degree that is difficult to estimate and, thus, mitigate. The ability to assess the impact of existing chemicals for which little to no toxicity data are available, or to foresee such effects during early stages of chemical development and use, and before potential exposure occurs, is a pressing need. However, the capacity of the current toxicity evaluation approaches to meet this demand is limited by low throughput and high costs. In the context of EPA's ToxCast™ project, we have evaluated a novel cellular biosensor system (Factorial™) that enables rapid, high-content assessment of a compound's impact on gene regulatory networks. The Factorial biosensors combined libraries of *cis*- and *trans*-regulated transcription factor reporter constructs with a highly homogeneous method of detection enabling simultaneous evaluation of multiplexed transcription factor activities. Here we demonstrate application of the technology towards determining bioactivity profiles by quantitatively evaluating the effects of 309 environmental chemicals on twenty-five nuclear receptors and forty-eight transcription factor response elements. We demonstrate coherent

[2] Previously published as 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**, 578-90, 10.1021/tx900325g.

transcription factor activity across nuclear receptors and their response elements and that Nrf2 activity, a marker of oxidative stress, is highly correlated to the overall promiscuity of a chemical. Additionally, as part of the ToxCast program, we identify molecular targets that associate with *in vivo* endpoints and represent modes of action that can serve as potential toxicity pathway biomarkers, and inputs for predictive modeling of *in vivo* toxicity.

Introduction

Estimating the toxicity of environmental chemicals is impeded by the sheer number of potential contaminants, the high costs of animal testing, and the poor prognostic power of traditional toxicity testing for assessing risks to humans. Relatively few predictive *in vitro* assays have been routinely used for screening, outside of those for genetic toxicology or specific molecular targets such as the steroid hormone receptors and the sodium channel hERG (Houck and Kavlock, 2008; Wilson et al., 2004). The paradigm of traditional toxicology has been challenged to shift towards using predictive or pathway-based toxicology approaches to more efficiently and systematically evaluate large numbers of chemicals for a diversity of toxicity endpoints (NRC, 2007). The U.S. Environmental Protection Agency (EPA) and National Institutes of Health (NIH) have responded to this challenge with major new computational toxicology testing and research initiatives. EPA's ToxCast^{TM} program (USEPA, 2009b), and the affiliated Tox21 program *(Collins et al., 2008)*, the latter a collaboration between three U.S. government entities – EPA, NIH's National Toxicology Program, and the NIH Chemical Genomics Center (NCGC) – are generating broad spectra of high-throughput/high-content biochemical and cell-based *in vitro* assays for a relatively large number of environmental chemicals (Collins, et al., 2008; Dix et al., 2007). These data are being used to derive biological and chemical profiles predictive of

in vivo biological endpoints and to serve as the basis for new toxicity screening and prioritization approaches.

One source of high-content, cellular data was obtained using a recently developed biological profiling technology, Factorial™ (Romanov et al., 2008), that enabled high-content, functional assessment of core components of cellular gene regulatory networks. Assessment was accomplished by measuring activity of transcription factors (TFs), i.e., the specialized classes of DNA-binding proteins that recognize regulatory elements in gene promoters and control transcription. Originally designed for multiplexed detection of specific cis-regulatory response element constructs (CIS), the technology has been further developed to provide assessment of trans-activating potential of multiple nuclear hormone receptors, a super-family of ligand-activated TFs (TRANS). The current technology was utilized in Phase I of EPA's ToxCast program to screen 320 environmental substances, the large majority of which are pesticide actives. We report here the results of that screening using forty-eight CIS and twenty-five TRANS assays chosen to place particular emphasis on factors and receptors that control toxicologically relevant cellular responses to xenobiotics, genotoxic stress, hypoxia, oxidative damage, immuno-modulation and endocrine disruption. Furthermore, we use these biological profiling results to identify perturbed gene regulatory networks and possible modes of action for the ToxCast 320 environmental substances. By introducing the plurality of CIS and TRANS assays into a human liver cell line, the biosensors enabled characterization of chemicals in the context of the cell's gene regulatory networks. We show that HepG2 cells responded to many chemical compounds through changes at the level of transcription factor and nuclear receptor activities. The changes elicited by the chemicals

reflect well-understood interactions in some instances, whereas in others they provide new insights into possible modes of action for chemical specific toxicities.

Methods

Chemical Library

The ToxCast 320 chemical library consists of 309 unique chemical structures meeting physicochemical property requirements for high-throughput screening *(8,9)*. Five substances were tested in duplicate (separately sourced) and three chemicals were tested in triplicate (sample replicates) for internal quality control purposes. Most of the compounds are pesticide active ingredients associated with extensive *in vivo* toxicity data generated in support of their registration process with the EPA. These data were extracted from documents, standardized, and compiled in the EPA ToxRefDB relational database (Knudsen et al., 2009; Martin et al., 2009a; Martin et al., 2009b). The derivation of the mammalian *in vivo* toxicity endpoints was focused on chronic/cancer, multigenerational reproductive, and prenatal developmental toxicity studies. For any given study type, roughly 250 of the 309 unique chemicals tested in this study have toxicity endpoint data in ToxRefDB. The full list of chemicals is available with quality reviewed structure-annotation from the EPA DSSTox website (USEPA, 2008). Chemicals samples were procured and plated by BioFocus DPI (San Diego, CA). Supplier-provided certificates of analysis indicated purity >97% for the large majority of chemicals (87%), and >90% purity for all but a few instances of technical grade or known mixtures. Follow-up analysis of an original solution plate by BioFocus DPI using LC and GC/MS (liquid and gas chromatography mass spectrometry), subsequent to assay screening, has confirmed mass identification, stability, and purity in excess of 90% for over 87% of the chemical library, with follow-up analysis underway for the remaining compounds. Summary

QC information mapped to chemical sample and solution IDs will be provided on the ToxCast website as an auxiliary chemical file (USEPA, 2009b). Compounds were dissolved in dimethyl sulfoxide (DMSO) to a final target concentration of 20 mM, in almost all cases. For testing in concentration-response format, serial dilutions were performed in DMSO followed by aqueous dilution in cell culture medium.

Assay Design and Implementation

Attagene Inc. (RTP, NC), under contract to the U.S. EPA (Contract Number EP-W-07-049), provided multiplexed reporter transcription unit (RTU) assays (Factorial™) consisting of forty-eight human transcription factor DNA binding sites transfected into the HepG2 human liver hepatoma cell line as previously described (Romanov, et al., 2008). In addition to the cis-acting reporter genes (CIS), a modification of the approach was used to generate a trans-system (TRANS) with a mammalian one-hybrid assay consisting of an additional twenty-five RTU library reporting the activity of nuclear receptor (NR) super-family members (Martin et al., 2010). A schematic representation of both systems is illustrated (Figure 3.1). The human ligand-binding domain of each nuclear receptor was expressed as a chimera with the yeast GAL4 DNA-binding domain that activated in trans a 5X-UAS-TATA promoter, which regulated transcription of a reporter sequence unique to each NR RTU. To ensure specificity of detection, each individual trans-RTU system including both receptor and reporter gene was separately transfected into suspended cells followed by pooling and plating of the transfected cells prior to screening. A major difference between the CIS and TRANS system is that in CIS activities of endogenous TFs are measured, whereas the TRANS assay evaluates changes in activities of exogenous,

chimeric NR-Gal4 proteins. Since the HepG2 cell line does not express some nuclear receptors the CIS assay cannot be used to evaluate these targets.

Concentration Selection

An initial cytotoxicity screen was performed to establish the maximum tolerated concentration (MTC) of the chemical library. The chemicals were tested for cytotoxicity against HepG2 cells in the MTT (3-(4,5-Dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide) tetrazolium assay (Mosmann, 1983) following 24 hour chemical exposure to five concentrations with an upper concentration of 50μM and 10-fold dilutions. All concentrations were run in triplicate. IC50 were determined based on fifty percent cell death, as measured by decreasing MTT conversion to formazan. The MTC was derived based on the one-third the calculated IC50 or, if no IC50 was determined, the MTC was set to 100μM. Results of cytotoxicity assessment are shown below.

Chemicals were then tested in the CIS and TRANS assays at seven concentrations starting at the MTC and followed by three-fold serial dilutions. Following exposure to chemical for a twenty-four hour period, cells were harvested and RNA was isolated and reverse transcribed into cDNA using fluorescent-tagged primers. Digestion with HpaI resulted in unique reporter gene fragments that were resolved and quantitated by capillary electrophoresis as previously described (Romanov, et al., 2008).

Statistical Analysis

Changes in transcription factor activity were expressed as fold-change over DMSO control by dividing expression levels of treated by the average of control (DMSO) treatments from the same assay plate. Both induction and suppression can occur. However, only 16 out of >20,000 possible chemical-assay combinations had marked suppression and were

observed at no more than two chemicals per assay. Therefore, suppression values were filtered out and only induction levels were considered in subsequent analyses. An initial maximum fold-change value (Emax) was used to compare across assays. Hierarchal clustering was performed using log 2-transformed Emax and subjected to clustering by Pearson's Dissimilarity using Ward's Method. Analyses were carried out using R version 2.8.1 (Gentleman et al., 2002). Emax values were also used to determine an optimal global cutoff as one criterion in establishing a hit, based on the agreement between replicates and stabilization of the overall hit rate as the Emax filter increases (Figure 3.2). Specifically, overall and hit concordances were calculated as is done in the final replicate analysis shown below, except solely using an Emax filter to establish hit calls. For each pair in the triplicate sets (comparing A with B, B with C and A with C), we asked if the chemicals were both hits, both non-hits or if they disagreed. Total concordance was defined as all 3 were hits or all 3 were non-hits, and non-concordance as either 1 or 2 were hits and the remainder were non-hits. The total overall concordance is the number of comparisons where the pair was either hits or both non-hits, divided by the total number of comparisons. The hit concordance is the number of cases where each of the pair was a hit divided by all cases where one or each of the pair was a hit.

The pairwise correlation of Emax values between any two assays was represented by R-squared. The distribution of correlations across all pairwise combinations of assays, a total of 2628, determined the ninety-fifth and ninety-ninth percentile; R-squared of 0.27 and 0.51, respectively. These values help determine the relative significance of the correlation between any two assays in this dataset.

A Hill function was fit to all fold-change data and an AC50 was derived, the concentration in which 50% of the maximal response, based on the fitted curve, is achieved. All assay-chemical combinations that did not achieve a fit with an R-squared >0.5 and an Emax greater than two were considered negative. A hit is considered anything that achieved an AC50 and met the criteria described above. A univariate analysis was conducted in order to test for associations, by way of relative risk (RR) values, between *in vitro* assays and mammalian *in vivo* toxicity endpoints. The mammalian *in vivo* toxicity endpoints, from chronic/cancer rat and mouse, multigeneration rat, and prenatal developmental rat and rabbit toxicity studies, were previously compiled and aggregated for use in predictive toxicology (Knudsen, et al., 2009; Martin, et al., 2009a; Martin, et al., 2009b). Each chemical-endpoint combination assessed in a toxicity study and captured in ToxRefDB were represented as the lowest effect level (LEL: mg/kg/day) if the endpoint was observed, thus defining a positive for an *in vivo* endpoint. RR values were calculated as:

$$RR = [TP / (TP+FP)] / [FN / (TN+FN)]$$

TP and FP represent the numbers of true and false positive chemicals, respectively, and TN and FN represent the numbers of true and false negative chemicals, respectively. A permutation test was developed in order to identify significant RR values and was carried out for 10,000 permutations. If the RR fell within the upper 95th percentile ($p<0.05$) of the permuted data, then the corresponding association was regarded as statistically significant. The permuted RR percentile values for all *in vitro* assays by *in vivo* endpoints were subjected to clustering by Pearson's Dissimilarity using Ward's Method.

Results

Cytotoxicity Assessment

Prior to running the CIS and TRANS assays, cytotoxicity was assessed in an attempt to establish a MTC for each chemical for purposes of concentration selection. Of the 320 total tested chemicals, 246 (77%) were not frankly cytotoxic under conditions of the assay and therefore a MTC of 100μM was used for subsequent testing. Sixty-four chemicals (20%) showed cytotoxicity at micromolar concentrations, whereas another ten chemicals (3%) displayed cytotoxicity at sub-micromolar concentrations.

Summary Statistics

In total, there were 23,360 chemical-assay observations with 1,923 (8.2%) showing significant responses. Of these hits, 78% of the AC50 were greater than 10μM with 4% being under 1μM. The median number of chemical hits across all assays was six, with a comparable hit rate across both CIS and TRANS formats. The pregnane X receptor response element (PXRE) was the most active of the assays with 234 significant chemical interactions (225 unique chemicals), consistent with the promiscuous nature of the receptor which is known to be activated by a wide range of pharmaceuticals, steroids, xenobiotics and natural products (Ekins et al., 2002). A complete listing of the assays with the number of hits per assay along with the maximal responder across the ToxCast 320 chemical library, and the Emax of the positive control, where available, is presented in Table 3.1. In the TRANS system, individual members of the TF family can be distinguished, for example PPARα, PPARγ and PPARδ, whereas the CIS system evaluates the integral activity of the entire PPAR family through use of the direct repeat (DR)-1 PPAR response element (PPRE) which does not show receptor isotype specificity (Lemay and Hwang, 2006). In total, 10 NR families across 16 TRANS assays were represented with their corresponding response elements.

Replicate Analysis

Of the 320 chemicals tested for transcription factor modulation there are three triplicates and five duplicates imbedded into the blinded chemical set. The triplicates were from the same production lot and vendor, whereas the duplicates were sourced independently and therefore excluded from the replicate analysis. Pairwise assessments of the triplicates were performed for the 657 chemical-assay combinations with 58 total hits (Figure 3.3). Including both positive and negative combinations, the overall concordance was >99%. Of the 58 hits, there was 87% concordance between the triplicates with twenty-seven pairs agreeing and four disagreeing. This is equivalent to nine assay-triplicate combinations in complete agreement and two assay-triplicate combinations having only a single hit. It should be noted that any disagreement in the triplicates would be accounted for twice because every comparison is performed pairwise. Therefore, 87% hit concordance was considered high and permitted the aggregation of replicates for specific downstream analyses making the total number of chemicals 309. For each triplicate the average AC50 was used when all or the majority were in agreement, otherwise the results were considered negative. The overall concordance among the duplicates was >96%, but a conservative approach was used due to the independent sourcing of the chemicals requiring both to be a hit and the average AC50 was then used, otherwise the results were considered negative. The complete data set is being made publicly available (USEPA, 2009b).

In vitro Relationships

Using Emax values, unsupervised hierarchal clustering of the 48 CIS and 25 TRANS assays demonstrated consistency across the chemical library (Figure 3.4). This was evidenced by co-clustering of the independently tested CIS and TRANS assays

corresponding to the same transcription factor function, including pregnane X receptor (PXR) with the PXR response element (PXRE), estrogen receptor alpha (ERα) with the estrogen response element (ERE), and peroxisome proliferator-activated receptor alpha and gamma (PPARα/γ) with the PPAR response element (PPRE). Such a response strongly suggests that the assay system is capable of detecting and reporting chemical perturbation of the cell, which is reflected in specific gene regulatory network alterations. Additional coherence in the dataset can be observed by the chemical-induced clustering of highly related nuclear receptors across their respective subtypes, including the retinoic acid receptors (RARα, RARβ, RARγ) and liver X receptor (LXRα and LXRβ). The corresponding response elements in the CIS assays, DR5 for RAR and DR4 for LXR co-clustered with the TRANS assays as expected. However, of the 10 NR families tested in both CIS and TRANS formats, 5 showed little evidence of co-clustering primarily due to a lack of significant activity in one of the assay systems, including farnesoid X receptor (FXR), glucocorticoid receptor (GR), vitamin D receptor (VDR), constitutive androstane receptor (CAR), and RAR-related orphan receptor (ROR). A lack of activity was not unexpected for certain targets with few known ligands, including FXR, VDR, and GR. Whereas, CAR and ROR are constituitively active permitting the evaluation of antagonists, but not agonists. All 5 NR families with robust activity were specifically and significantly correlated across assays.

PXR/PXRE

The xenobiotic sensor, PXR, which regulates a diverse set of xenobiotic response genes including phase I, II and III metabolic enzymes *(Gillam, 2002)*, responded to a large number of the chemicals in the library in both the CIS and TRANS format (Figure 3.5). While the TRANS assay used an exogenously expressed GAL4-PXR for activation of the

reporter gene, the CIS system used a 628 base-pair fragment (-7836 to -7208) of a PXR-regulated gene, CYP3A4, previously identified as containing a PXRE (Goodwin et al., 1999) together with endogenous PXR. These assays showed a strong correlation with an R-squared of 0.67. As stated in the Methods section, the 95th and 99th R-squared percentile was 0.27 and 0.51, respectively, thus demonstrating the significant correlation between the independently tested PXR_TRANS and PXRE_CIS assays. However, less robust responses were observed in the TRANS version across chemicals when compared to the CIS assay. Responses averaged approximately eight-fold lower in the TRANS assays, Sensitivity was thus lower in the TRANS version, resulting in the derived AC50 values of the TRANS assay being a nearly complete subset of the hits in the CIS assay. In addition, ~90% of the AC50s for TRANS hits fell within three-fold of the PXRE_CIS AC50 (R-squared of 0.56). This is evident among the top five most efficacious and most potent PXR agonists within the chemical library (Figures 3.5b and 3.5c). The qualitative and quantitative similarities between the PXRE_CIS and PXR_TRANS concentration response curves provide redundant and corroborating evidence of PXR agonism. The positive control and known human PXR ligand, rifampicin, was only run in the CIS format and showed comparable efficacy and potency with a number of the tested chemicals representative of diverse set of chemical classes, demonstrating the promiscuous nature of PXR and the suitability of the assay for detecting such agonists.

RAR/DR5

The endogenous ligands for the retinoid acid receptors, retinoids, control key components of development, differentiation and homeostasis (Chambon, 1996). The TRANS version of reporter for the retinoid receptor assays demonstrated increased sensitivity in

comparison to its response element, DR5, as shown with RARα in Figure 3.6a. While the reason for this is not apparent, it may be due to insufficient endogenous RAR levels in the HepG2 cells. Similar differences in sensitivity across other nuclear receptor families from CIS and TRANS versions may also be due to insufficient endogenous receptor in HepG2 cells, which is required in the CIS assays. In general, RARα, RARβ and RARγ all showed activation by similar chemicals and were positively and significantly correlated to their response element, DR5, with R-squared values of 0.58, 0.36, 0.27, respectively. Trans-retinoic acid, an endogenous RAR ligand, is teratogenic when developing animals are exposed experimentally. Hence, it is important to understand both the potency and efficacy of the chemicals identified as RAR activators. The top five most efficacious and potent RAR agonists, based on average RARα and DR5 results, were substantially less responsive than retinoic acid with respect to both potency and maximum efficacy (Figures 3.6b and 3.6c). The retinoic acid used as a positive control, tested only in the DR5_CIS assay, generated an Emax of 7.5 and an AC50 of 0.047μM compared to the highest Emax of 3.2 and the lowest AC50 of 0.76μM across the chemical library.

ER/ERE

Direct activation of the estrogen receptor (ER) is one mechanism for xenobiotic endocrine disruption through alteration of the physiological function of endogenous steroid hormone receptors (Witorsch, 2002). The TRANS ERα assay provides a measurement of receptor activation as does the ER response element (ERE) although the latter is less direct. The correlation between CIS and TRANS assays is above the 99th percentile with an R-squared of 0.53 (Figure 3.7). Among the most potent and/or efficacious chemicals, several are well known estrogen receptor agonists including bisphenol A, methoxychlor and its

metabolite, HPTE (Gaido et al., 2000; Gould et al., 1998; Gray et al., 1989). HPTE was the most potent in the chemical library, but showed low efficacy most likely due to limitations on the highest concentration tested coming out of the cytotoxicity pre-screen. Other chemicals showing significant activity in either the TRANS or CIS format have not been reported to be ERα agonists. Additionally, the ER TRANS assay generated 90 hits, higher than would be expected, demonstrating sensitivity but potentially lacking specificity. However, where there is good agreement across all assay types for novel findings, e.g. fludioxonil and flumetralin, strong consideration should be given to additional testing in other assay formats to confirm the activity of these chemicals as ER ligands. It should be pointed out that many of these chemicals primarily behave as partial agonists relative to a strong endogenous ligand such as 17-β-estradiol (ERE_CIS: AC50 of 0.035nM and Emax of 9.3; ERα_TRANS: AC50 of 0.87nM and Emax of 23) and this may impact their biological effects both quantitatively and qualitatively.

PPAR/PPRE

The PPRE binds the PPAR subfamily (α, β/δ, γ) to regulate genes associated with lipid metabolism (PPARα), fatty acid oxidation (PPARβ/δ) and adipocyte differentiation (PPARγ). In Figure 3.4, PPRE_CIS clusters with PPARα_TRANS and PPARγ_TRANS assays but not with PPARβ/δ_TRANS. The lack of significant activators of PPAR β/δ explains why it failed to cluster with the PPRE. In Figure 3.8a, the correlation of the PPRE and PPARγ is shown (R-squared = 0.53) with a subset of the chemicals (n=7) also displaying PPARα activity. These apparent PPARα/γ co-agonists and their involvement in rodent liver tumor formation is further investigated below. Rosiglitazone, a selective ligand of PPARγ used as the positive control, had roughly 30% greater efficacy and was five-fold

more potent than the rest of the chemicals, with the single exception of fentin in the PPRE_CIS assay (Figures8b and 8c). However, fentin demonstrated substantially depressed induction in PPARγ_TRANS and no activity in PPARα_TRANS. The PPRE_CIS response may be indicative of activation of the PPAR/RXR heterodimer signal-transduction pathway through RXR agonist activity (le Maire et al., 2009). This is supported by fentin's strong RXR responses in the RXRβ_TRANS assay with an AC50 of 83nM. This behavior, termed permissive heterodimeric activity (Aranda and Pascual, 2001; Bettoun et al., 2003), has been reported for other receptor pairs in the mammalian one-hybrid assay (Bettoun, et al., 2003). Rosiglitazone showed moderate induction of the PPARα_TRANS assay with an Emax of 3.6 and an AC50 of 6μM reflecting partial agonist behavior at much higher concentrations than for PPARγ, consistent with previously reported results (Reifel-Miller et al., 2005). Similar to the situation with ERα, the biological significance of the relatively lower efficacy and potency of these chemicals compared to the positive control may impact high-dose animal toxicity as well as create additional uncertainty for implications for human disease.

NRF2/ARE

The nuclear-factor-E2-related factor (NRF)-2, a member of the bZIP transcription factor family, regulates cytoprotective enzymes in response to oxidants and electrophilic compounds through binding to the antioxidant response element (ARE) (Nguyen et al., 2009). Regulated genes include γ-glutamylcysteine synthetase, NADPH:quinone reductase, and glutathione-S-transferase (Lee and Johnson, 2004). Many of the chemicals evaluated here produced significant induction of the NRF2/ARE reporter gene (Figure 3.9). Among the top ten activators are many pesticides known to cause oxidative damage in a variety of species. These include metolachlor (Stajner D., 2001), diquat dibromide (Wolfgang et al.,

1991), trichlorfon (Karademir et al., 2007), oxyfluorfen (Peixoto et al., 2006), alachlor (Burman et al., 2003), and dichlorvos (Yarsan and Cakir, 2006). Other significant activators of NRF2 were compounds capable of reaction with sulfhydryl groups, e.g. methyl isocyanate (Emax = 3.5), a breakdown product of metam sodium (Emax = 6.0), dazomet (Emax = 3.6) and metiram-zinc (Emax = 3.2). As this is an adaptive response, the toxicological relevance may be very dependent on specific *in vivo* exposure scenarios. However, using the NRF2_ARE_CIS assay results as an indicator of oxidative or electrophilic damage, we observed a relationship between NRF2 response and the non-specific and promiscuous behavior of various chemicals and assays. In comparing the NRF2 activity and the average activity across all other assays, a significant relationship emerged (R-squared = 0.5) (Figure 3.9). It should be noted that these values were log2-transformed in order to mitigate the impacts of averaging across assays with significant differences in dynamic range. This relationship cannot be attributed to assay conditions, because the positive control chemicals demonstrated very specific activity for their respective target with no evidence of NRF2 activity at comparable concentrations.

The phenomena of oxidative stress leading to promiscuous transcription factor activity may help explain the large number of active compounds in assays with fairly specific ligand binding domains, including ERα and PPARγ. For example, bisphenol A, a known ERα agonist, achieved an AC50 in 13 of the 73 assays, but only the AC50s for ERα_TRANS and ERE_CIS were significantly (>3-fold) more potent than the NRF2_ARE_CIS. Specifically, bisphenol A had AC50s of 1.1 and 0.64 μM in the ERα_TRANS and ERE_CIS assays, respectively, compared to an AC50 of 27 μM in the NRF2_ARE_CIS assay. An additional 26 chemicals hit the two ER assays and the Nrf2 assay, but their ERα_TRANS or ERE_CIS

AC50s were all comparable to their respective NRF2_ARE_CIS AC50. In general, these 26 chemicals showed promiscuous activity across both CIS and TRANS systems, averaging double the number of hits compared to the rest of the chemical library (11 versus 5.5). Thus, oxidative/electrophilic stress appears to elevate transcription factor activity non-selectively, and comparing potencies to NRF2 activity is therefore useful for putting compound activity into perspective and aiding interpretation of results.

In vitro to In vivo Associations

In comparing the 73 *in vitro* assay data to 77 *in vivo* endpoints from rodent chronic bioassays, rat multi-generational reproductive toxicity studies and rat and rabbit prenatal developmental toxicity studies, 133 significant univariate associations were established based on the methods described above. Various non-steroidal nuclear receptors involved in xenobiotic metabolism and/or lipid metabolism are specifically associated with target-organ pathologies and tumorigenesis, whereas steroidal and other nuclear receptors are specifically associated with endocrine-related organ pathologies, offspring survival and the developing system (Figure 3.10). Examples of these associations are detailed below.

PPAR & Rat Liver Tumors

Of the 309 chemicals, 256 have two-year rat cancer bioassay recorded in ToxRefDB. Of these, 101 were found to cause tumors at one or more target sites. Of those 101 chemicals, 23 were found to cause liver tumors, specifically hepatocellular adenomas or carcinomas. These 23 chemicals were evaluated for genotoxicity throughout the pesticide registration and re-registration process and generally found to be negative (USEPA, 2009a), which suggests that these chemicals predominantly act through non-genotoxic mechanisms of action. In comparing the *in vitro* assay data to *in vivo* endpoints, only PPARα and PPARγ were

significantly associated with rat liver tumors. The PPAR mode of action has been a focus of toxicological research with special emphasis on assessing relevance to humans (Lai, 2004). Additionally, the prevalence of rodent liver and other tumors in chronic toxicity studies of PPARα agonists and PPARα/γ co-agonists has resulted in the U.S. Food and Drug Administration (FDA) issuing guidance for development and use of PPAR agonists that requires special consideration be given to the carcinogenic potential of these agents (USFDA, 2008). In total, 143 chemicals activated PPARγ with seven of those chemicals also activating PPARα (Figure 3.11). Five out of these seven chemicals (diethylhexyl phthalate (DEHP), perfluorooctanoic acid (PFOA), imazalil, lactofen, and diclofop-methyl) are positive for rat liver tumorigenicity; while only bromoxynil and fenthion are not. PPARα assay results demonstrated the ability to independently classify chemicals as potential liver tumorigens. If a chemical was considered a hit for PPARα the relative risk for rat liver tumor induction is 9.9 ($p<0.01$) with very high specificity (99%) and lower sensitivity. Additionally, PPARγ agonist chemicals were significantly associated with rat liver tumorigens with a relative risk of 6.6 ($p<0.01$), but with high sensitivity (83%) and lower specificity. One could postulate that these chemicals act predominantly through a non-genotoxic peroxisome proliferation-mediated mode of action, and for a few chemicals there is data in the literature to support this, including PFOA (Lau et al., 2004), DEHP (Melnick, 2001), Lactofen (Butler et al., 1988; Lai, 2004), and Diclofop-methyl (Lai, 2004). It should be noted that oxadiazon and perfluorooctane sulfonic acid (PFOS) are purported to cause liver tumors via a PPARα mode-of-action (DeWitt et al., 2009; Richert et al., 1996; USEPA, 2003), but were not positive under the conditions of this assay, although both of these chemicals were PPARγ agonists. PFOS was tested at a top concentration of 14 μM and oxadiazon at 24.5 μM

because of cytotoxicity observed in the MTC determination. Two reports evaluating the *in vitro* activation of human PPARα by PFOS differed with one determining an AC50 of 13-15 μM (Shipley et al., 2004) and the other reporting no significant activity up to 250 μM (Takacs and Abbott, 2007); therefore, the interpretation of the present results are difficult. For oxadiazon, direct activation of PPARα has not been reported in the literature although *in vivo* peroxisome proliferation was seen in rodents treated with high doses (≥ 100 mg/kg/day for 28 days) and significant peroxisomal enzyme induction measured in primary rat hepatocytes cultures required 50-100 μM (Richert, et al., 1996).

Endocrine Pathology, Offspring Survival & the Developing System

In total, 251 and 261 chemicals have rat multigenerational reproductive and prenatal developmental toxicity studies recorded into ToxRefDB, respectively, with 225 chemicals having both studies in ToxRefDB. Two groups of assays showed significant associations with specific reproductive and developmental processes or targets (Table 3.2). As shown above, RARα and DR5 were correlated to each other and RAR has direct biological links to the developing system (Chambon, 1996). Chemicals positive for RARα/DR5 have a significantly greater potential for effects in various developmental systems and at different stages of development, including orofacial defects (i.e., cleft lip/palate) and urogenital malformations (i.e., renal and ureteric). In contrast, the estrogen receptor alpha (ERα) and its response element (ERE), a known endrocrine disrupting target, were associated with reproductive effects, including reproductive performance (i.e., decreased fertility and implantation loss). In addition to their specific associations, RARα/DR5 and ERα/ERE appear to have similar impacts on early offspring survival as indicated by shared associations to litter size and live birth outcomes. Interestingly, RARα/DR5 and ERα/ERE activity is not

correlated with an R-squared of roughly zero, meaning there shared associations are independent and may reflect varying mechanisms leading to the same adverse outcome. In general, the predictive power of probing these molecular targets for tissue-specific observations such as testicular or adrenal effects is limited, whereas data on hormone or cholesterol levels would most likely provide direct links from molecular to phenotypic alterations, but are not available for sufficient numbers of compounds in this chemical library.

Discussion

Cells respond to changes in their environment through a variety of recognition systems and signaling pathways that are integrated at the level of transcription factors, which coordinate appropriate responses. The multiplexed reporter gene technology described here attempts to probe this biological response by generating profiles of individual chemicals acting against a large panel of transcription factors in two different assay systems. These profiles provide information on potential molecular interactions of the chemicals, as well as more generalized activity such as induction of oxidative and electrophilic stress. Quantitation of the effects of chemicals on specific transcription factor responses provides information that can be used to determine potential for toxicity. Characterizing the responses to known toxicants in these assays allows the magnitude and potency of responses for new chemicals to be put into an appropriate context for evaluation. Testing chemicals in both the CIS and TRANS systems revealed specific limitations of each system, as well as complementarity of the data from each system. Valuable insights for interpreting large screening datasets came from comparing CIS and TRANS data, especially when limitations of one system, e.g., lack of endogenous activity for a particular target, were attenuated by data from the other. In cases

where both complementary assays demonstrated robust responses, greater confidence resulted from associating a given chemical with a specific molecular activity.

Chemicals can both directly and indirectly affect transcription factor activities. The nuclear receptor superfamily and the Ah receptor are examples of TFs that can bind xenobiotic chemicals directly through their ligand-binding domains, resulting in modulation of target gene activity. In some cases, receptors such as PXR appear to have evolved to recognize such xenobiotics and to control expression of a large series of Phase I, Phase II, and Phase III metabolizing enzymes and transporters (Kliewer et al., 2002). Whereas such interactions generally serve to help eliminate or detoxify a xenobiotic, they may also lead to undesirable consequences, resulting in a variety of toxicities when exposure is sufficiently high. For example, induction of the Phase I, II, and III activities can also cause effects on endogenous compounds, e.g., steroid hormone metabolism, altering normal physiology. In addition, such enzyme induction can cause interference with the expected metabolism of pharmaceuticals leading to side effects resulting from either too high or too low a dose of the drug (Wilkinson, 2005). Endocrine disruption is another example, whereby xenobiotics mimic natural ligands such as steroid hormones in binding to nuclear receptors, resulting in disruption of normal receptor signaling and potentially leading to reproductive toxicities (WHO, 2002). Outside the endocrine system, chemicals activating PPARα cause peroxisome proliferation in rodents and subsequent hepatocarcinogenesis, a cancer mode of action that does not seem to occur in humans (Lai, 2004). Thus screening environmental chemicals for the ability to affect ligand-activated transcription factors can provide a means to recognize the potential for the chemical to cause endocrine disruption and other toxicities.

Chemical effects on cells may also be detected though effects on non-ligand-activated TFs. Such TFs include those that respond to a wide variety of cellular stressors, including oxidative, electrophilic, hyperosmolarity, DNA damage, hypoxia, etc. These stress-response effects enable a cell to survive environmental changes by adapting through increasing levels of target gene products involved in reducing or buffering the stress, e.g. increasing glutathione levels to handle oxidative stress or heat-shock proteins to provide protection to newly synthesized proteins during heat or UV light stress (Lee and Johnson, 2004). When stress is beyond the tolerance of the adaptive systems, cell death occurs through necrosis or apoptosis, the latter allowing selective removal of irreparably damaged cells from the tissue. Examples of effects seen in the current study included many chemicals inducing oxidative/electrophilic stress (Nrf-2/ARE CIS) but few inducing DNA damage as measured by activation of the p53 CIS assay. The latter finding is consistent with the majority of the chemicals tested being food-use pesticides that would not be expected to have significant genotoxic effects. Because transcription factors of this general functional class are involved in adaptive responses, changes in their activation do not necessarily equate directly to toxicity. Rather, they suggest that the chemical inducing the response possesses a mode of action that may, under sufficient exposure conditions, lead to effects that can overwhelm the capacity of the adaptive response, resulting in toxicity. Evaluating the potential for toxicity of chemicals with such activity will require careful consideration of potency and exposure to properly inform estimations of risk.

In contrast to the receptors involved in nuclear receptor signaling and stress response, many of the other transcription factors screened in this study are known to be involved with specific signaling pathways controlling growth and differentiation. For example, E2F, MYC,

and SMAD regulate growth in virtually all cells (Grandori et al., 2000), whereas OCT, PAX, SOX, GLI, and TCF are among the factors involved in specific aspects of cellular differentiation (Maeda et al., 2007). In general, the reporter genes for these response elements showed only marginal response to the chemical library. The reasons for this are not clear. One possibility is that the signaling pathway controlling the specific reporter gene response is not functional under the conditions used for the assay. Not every reporter gene response was evaluated with suitable positive controls and, thus, activity could only be inferred from the fact that some basal level of reporter gene product was present in the cell. Use of a different cell line may be one way to expand functional pathway coverage. In addition, by prescreening for cytotoxicity, we ensured that frankly cytotoxic concentrations of chemicals would not be used in the screen. It is possible that interference with many of the transcription factors required for growth and differentiation results in cytotoxicity and, thus, the screening strategy employed to avoid this biased the results against finding effects on these pathways. Future screening efforts will take both of these points into consideration.

Relating effects observed with *in vitro* assays using transformed cell lines to *in vivo* toxicity remains a significant challenge. Here we evaluated statistical correlations between chemical effects on transcription factors and toxicity endpoints, the latter collated in a relational database derived from extensive animal testing in support of the U.S. pesticide registration process. The relatively few significant associations detected were plausible associations between peroxisome proliferator-activated receptors and liver toxicity, estrogen receptor activity correlating to reproductive toxicities and retinoic acid receptor toxicities associated with developmental defects. These results were confirmed by the co-clustering of specific targets across the CIS and TRANS systems, thus increasing confidence in the *in vivo*

associations. The application of *in vitro* screening to chemical prioritization is enhanced by multiple, complementary assays such as these for the same target. There are many possible explanations for the inability to discern more correlations. The associations between these selected groups of molecular targets and apical endpoints do not encompass all possible routes by which a chemical can induce an adverse effect or toxic outcome. Likewise, it may be that there are insufficient chemicals within the current library representing each possible route to enable detection of significant associations. It is also understood that there are significant uncertainties in drawing associations between high-dose laboratory animal toxicity studies and *in vitro* animal target activities, much less the further extrapolation from *in vitro* human molecular target activities, despite the greater relevance of the latter to potential human risk. Chemical *in vivo* effects in relation to *in vitro* effects also may be greatly affected by chemical ADME properties (adsorption, distribution, metabolism and elimination), with metabolic biotransformation, in particular, resulting in either much greater or reduced toxicity than might otherwise be predicted. Whereas the HepG2 cell line used here is derived from liver and maintains some metabolic capacity, it is much less than in the intact liver (Donato et al., 2008). Finally, many of the responses seen were statistically significant; however, the magnitude of the response was frequently much less than that seen with a positive control compound. Again, relating the level of efficacy observed in the *in vitro* assay to biological meaningful results *in vivo* poses significant challenges.

The multiplexed RTU assay results evaluated in the present study encompass only a fraction of the full range of biological targets and assays, both cell-free and cell-based, being run against the ToxCast Phase I chemical library *(*USEPA, 2009b*)*. However, the present assays are based on human transcription factors and, thus, are believed to have particular

potential relevance to humans and to informing *in vivo* extrapolations. The associations reported in this study, many of which are corroborated by results reported in the literature, demonstrate the biological relevance of several of the presently studied molecular targets for monitoring and probing toxicity pathways and for understanding mechanisms of action leading to specific chemical toxicities. Along with chemical properties and ADME considerations, combining these activity profiles with additional *in vitro* assay profiles having the potential to probe a broader range of biological responses in animals and humans, will offer greater opportunities to discern meaningful associations of *in vitro* profiles with *in vivo* effects. Ongoing efforts to enlarge the test chemical library will additionally enrich the chemical and biological information dimensions in relation to toxicity endpoints of regulatory interest and, in so doing, provide greater coverage of toxicity mechanisms and facilitate the development of robust methods suitable for prioritizing chemicals based on potential for toxicity.

Tables

Table 3.1. The 73 CIS and TRANS assays listed with the number of hits per assay, the maximal responder out of the ToxCast_320 chemical library, and positive control (where available). Maximal efficacy values (Emax) are provided in parenthesis for each listed test or positive control chemical.

Assay Name	No Hits (n=320)	ToxCast_320 Max Responder (Emax)	Positive Control (Emax)
CIS/TRANS Related Assays			
ERE_CIS	39	Pyridaben (8.7)	Estradiol (9.3)
ERa_TRANS	90	Pendimethalin (27)	Estradiol (23) 6-Fluor-Test (23)
IR1_CIS	3	Tebupirimfos (3.5)	
FXR_TRANS	1	Butachlor (2.3)	CDCA (6.5)
GRE_CIS	0	Spiroxamine (1.6)	Dexamethazone (5.0)
GR_TRANS	0	Butachlor (1.9)	
DR4_LXR_CIS	7	Tebupirimfos (3.8)	
LXRa_TRANS	23	Tebupirimfos (23)	T9 (55)

LXRb_TRANS	21	Tebupirimfos (13)	T9 (81)
PPRE_CIS	125	Resmethrin (5.7)	Rosiglitazone (6.7)
PPARa_TRANS	9	Lactofen (11)	GW7647 (11)
PPARd_TRANS	1	Flusilazole (2)	GW7647 (13)
PPARg_TRANS	146	Resmethrin (22)	GW7647 (39) Rosiglitazone (31)
PXRE_CIS	234	Flufenacet (31)	Rifampicin (29)
PXR_TRANS	102	Fipronil (7.9)	
DR5_CIS	27	Imazalil (3.2)	9-cis-Retinoic Acid (7.5)
RARa_TRANS	49	Lindane (5.9)	
RARb_TRANS	5	Oxadiazon (2.6)	
RARg_TRANS	4	Imazalil (2.2)	
RORE_CIS	35	Tetraconazole (3.3)	
RORb_TRANS	1	Tebufenozide (2.2)	
RORg_TRANS	1	Azoxystrobin (2.9)	
VDRE_CIS	134	Pyridaben (5.5)	
VDR_TRANS	0	Rotenone (1.9)	
PBREM_CIS	24	Prodiamine (2.8)	
CAR_TRANS	4	Phosalone (2.4)	
Remaining TRANS Assays			
ERRa_TRANS	3	Cyazofamid (2.1)	
ERRg_TRANS	13	Fenthion (2.5)	
RXRa_TRANS	0	Esfenvalerate (2.3)	
RXRb_TRANS	8	Fludioxonil (17)	
AR_TRANS	0	Butachlor (2.1)	6-Fluor-Test (6.3)
Hpa5_TRANS	0	Rotenone (1.9)	
HNF4a_TRANS	16	Trichlorfon (3)	
NURR1_TRANS	9	Pirimiphos-methyl (5.6)	
THRa1_TRANS	18	(Z,E)-Fenpyroximate (3.2)	T3 (11)
Assay Name	No Hits (n=320)	ToxCast_320 Max Responder (Emax)	Positive Control (Emax)
Remaining CIS Assays			
Ahr_CIS	54	Tetraconazole (72)	FICZ (100)
AP_1_CIS	52	Prochloraz (4.6)	
AP_2_CIS	0	Hexaconazole (1.7)	
NRF2_ARE_CIS	165	Dichlorvos (16)	
BRE_CIS	93	Fipronil (12)	
C_EBP_CIS	5	Cyazofamid (2.4)	
CMV_CIS	37	Tebupirimfos (3.4)	
CRE_CIS	52	Prallethrin (4.4)	Forskolin (2.5)
E2F_CIS	0	Triclosan (1.5)	
E_Box_CIS	1	Prallethrin (2.5)	
EGR_CIS	40	Prallethrin (8.1)	
Ets_CIS	0	Fluazinam (1.3)	
FoxA2_CIS	0	d-cis,trans-Allethrin (1.8)	
FoxO_CIS	0	Simazine (1.5)	
GATA_CIS	0	Rotenone (1.4)	
GLI_CIS	2	d-cis,trans-Allethrin (3)	
HIF1a_CIS	25	Tetraconazole (3.7)	
HNF6_CIS	0	Spiroxamine (1.5)	
HSE_CIS	22	Prallethrin (33)	Geldanamycin (5.7)
ISRE_CIS	0	Oxasulfuron (1.6)	
MRE_CIS	62	d-cis,trans-Allethrin (28)	
Myb_CIS	0	Hexaconazole (1.6)	
Myc_CIS	4	d-cis,trans-Allethrin (3.6)	
NFI_CIS	4	Prallethrin (2.6)	

NF_kB_CIS	10	Dichloran (3.4)
NRF1_CIS	0	Prohexadione-calcium (1.4)
Oct_MLP_CIS	86	Prallethrin (12)
p53_CIS	2	Dichlorvos (2.4)
Pax6_CIS	14	Prallethrin (5.2)
Sox_CIS	2	Cacodylic acid (2)
Sp1_CIS	8	d-cis,trans-Allethrin (2.9)
SREBP_CIS	6	d-cis,trans-Allethrin (2.5)
STAT3_CIS	0	Dichloran (1.8)
TA_CIS	4	Tribufos (2.5)
TAL_CIS	1	Prallethrin (2.3)
TCF_b_cat_CIS	0	Diphenylamine (1.7)
TGFb_CIS	8	Oxamyl (9.6)
Xbp1_CIS	12	Prallethrin (9.7)

Table 3.2. Two distinct molecular targets with known links to developmental and reproductive toxicities, retinoic acid receptor and estrogen receptor, demonstrate coherent associations to *in vivo* developmental and reproductive endpoints from ToxRefDB and are shown with their respective significant relative risk (RR) values ($p<0.05$).

	Prenatal Developmental			Multigeneration Reproductive						
	Malformations			Reproductive Performance			Offspring Survival			
Assay	Cleft Lip/Palate	Renal	Urogenital	Fertility	Gestational Interval	Implantations	Litter Size	Live Birth Index	Viability Index	Lactation Index
ERa_TRANS			5			9	2	2		
ERE_CIS				3						
RARa_TRANS	4	3			4	4	2	2	2	2
DR5_CIS	5	5	6							

Figures

Figure 3.1. Evaluation of ToxCast compounds by using libraries of cis- and trans-reporter transcription units. Schematic representation. (a.) Detection of cis-RTUs. A mix of 51 individual plasmids encoding cis-RTUs was co-transfected into suspension of HepG2 cells, cells were plated and stimulated with evaluated compound. At the end of incubation period, total RNA was isolated and detected as described previously *(*Romanov, *et al., 2008)*. (b.) Detection of trans-RTUs. Twenty four trans-RTU plasmids were separately transfected into HepG2 cells, the transfected cells pooled, plated, and stimulated with evaluated compound. At the end of incubation, total RNA was isolated and detected as described previously (Romanov, et al., 2008).

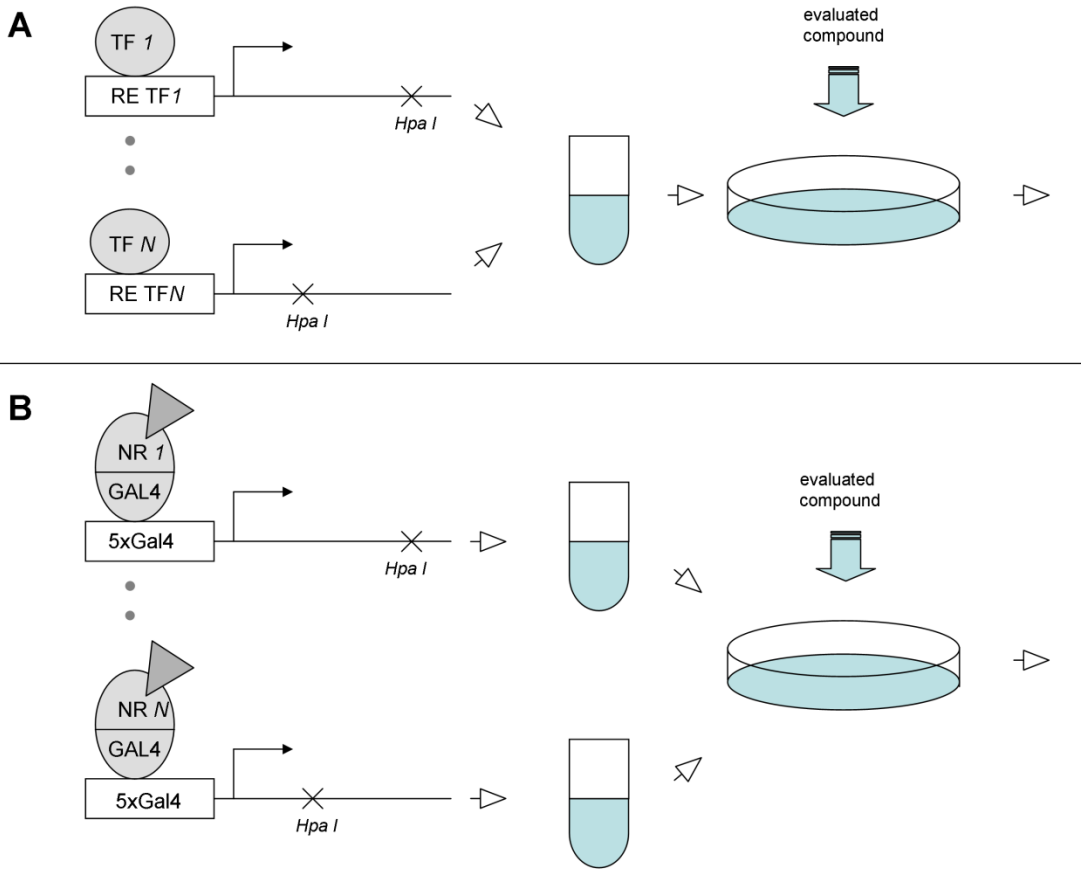


Figure 3.2. Overall and hit concordances across technical replicates determine optimal cutoff as a primary summary call criteria.

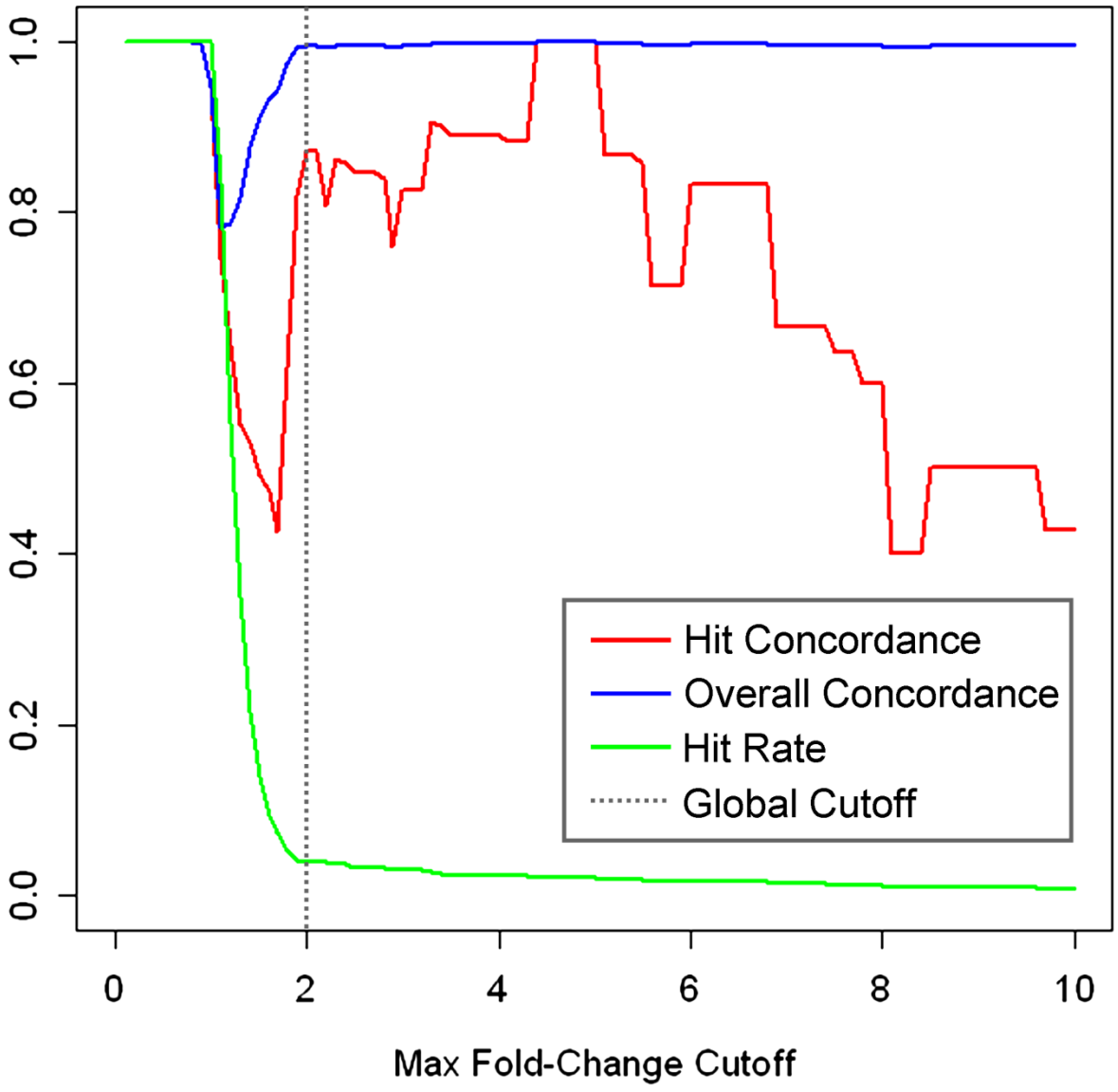


Figure 3.3. Replicate analysis performed on the technical replicates. For each pair in the triplicate sets (comparing A with B, B with C and A with C), we asked if the chemicals were both hits, both non-hits or if they disagreed. Total concordance was defined as all 3 were hits or all 3 were non-hits, and non-concordance as either 1 or 2 were hits and the remainder were non-hits. The total overall concordance is the number of comparisons where the pair was

either hits or both non-hits, divided by the total number of comparisons. The hit concordance is the number of cases where each of the pair was a hit divided by all cases where one or each of the pair was a hit.

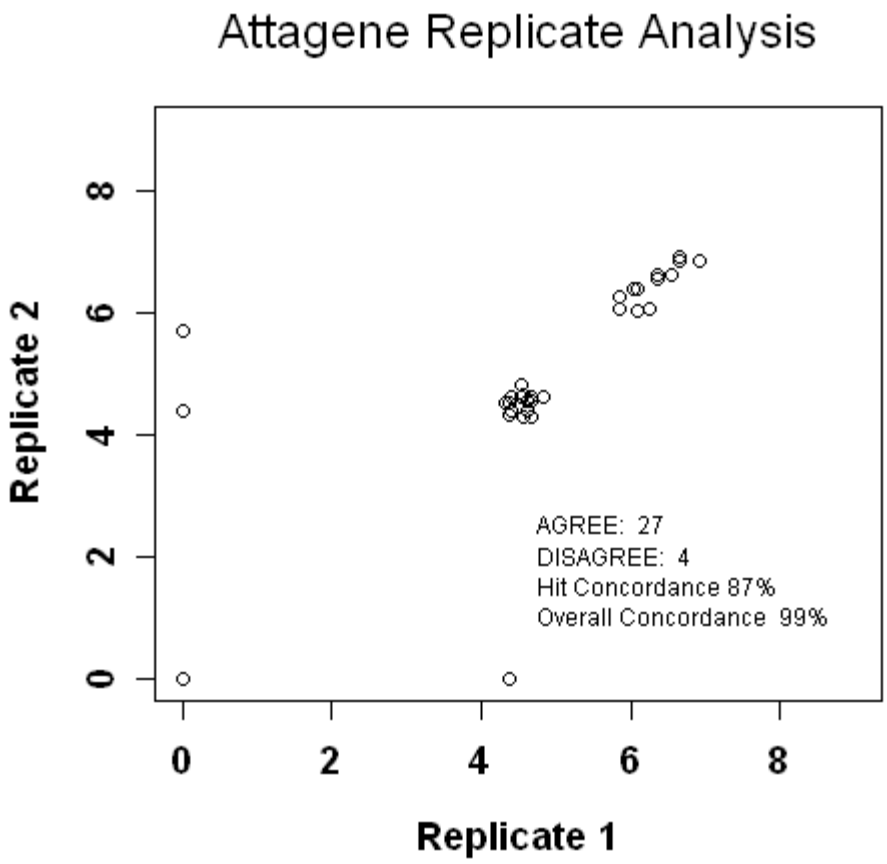


Figure 3.4. Two-way hierarchical clustering of the 48 CIS and 25 TRANS assays demonstrating the relationship between families of nuclear receptors and their response elements, independently tested in either the CIS or TRANS format.

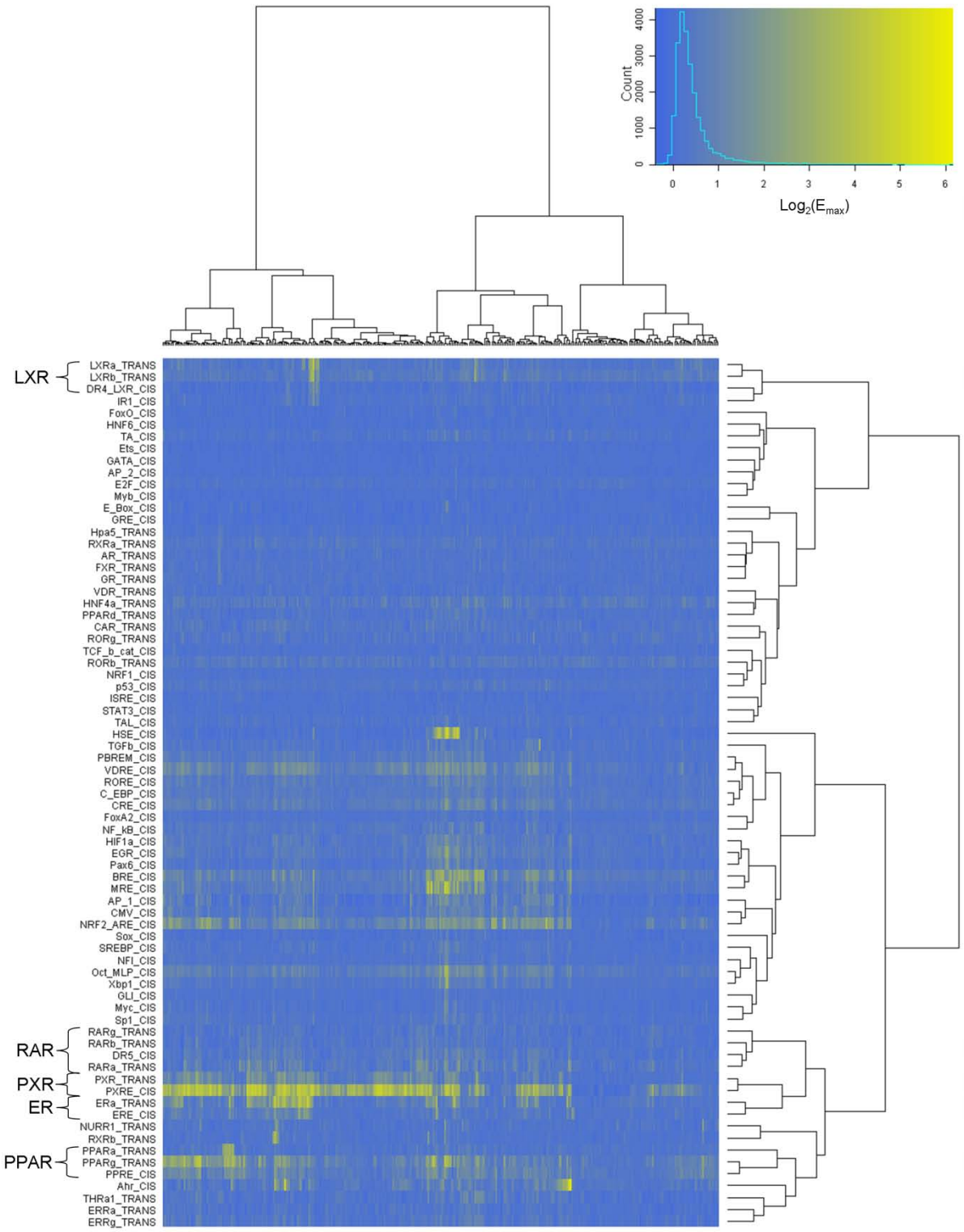

Count
0
1000
2000
3000
4000
0
1
2
3
4
5
6
Log2(Emax)
LXR
RAR
PXR
ER
PPAR
LXRa_TRANS
LXRb_TRANS
DR4_LXR_CIS
IR1_CIS
FoxO_CIS
HNF6_CIS
TA_CIS
Ets_CIS
GATA_CIS
AP_2_CIS
E2F_CIS
Myb_CIS
E_Box_CIS
GRE_CIS
Hpa5_TRANS
RXRa_TRANS
AR_TRANS
FXR_TRANS
GR_TRANS
VDR_TRANS
HNF4a_TRANS
PPARd_TRANS
CAR_TRANS
RORg_TRANS
TCF_b_cat_CIS
RORb_TRANS
NRF1_CIS
p53_CIS
ISRE_CIS
STAT3_CIS
TAL_CIS
HSE_CIS
TGFb_CIS
PBREM_CIS
VDRE_CIS
RORE_CIS
C_EBP_CIS
CRE_CIS
FoxA2_CIS
NF_kB_CIS
HIF1a_CIS
EGR_CIS
Pax6_CIS
BRE_CIS
MRE_CIS
AP_1_CIS
CMV_CIS
NRF2_ARE_CIS
Sox_CIS
SREBP_CIS
NFI_CIS
Oct_MLP_CIS
Xbp1_CIS
GLI_CIS
Myc_CIS
Sp1_CIS
RARg_TRANS
RARb_TRANS
DR5_CIS
RARa_TRANS
PXR_TRANS
PXRE_CIS
ERa_TRANS
ERE_CIS
NURR1_TRANS
RXRb_TRANS
PPARa_TRANS
PPARg_TRANS
PPRE_CIS
Ahr_CIS
THRa1_TRANS
ERRa_TRANS
ERRg_TRANS

Figure 3.5. Comparison of pregnane X receptor (PXR) activity in the CIS (PXRE) and TRANS format characterizes both relative efficacy and potency across the chemical library. Maximum efficacy (Emax) plotted for both the CIS and TRANS assays across the chemical library (a). The colored circles represent the chemicals positive (AC50 derived) or negative for the assays on the x- and y-axis. (b. and c.) The top 5 most efficacious and top 5 most potent chemicals averaged across both assays demonstrate the relative response compared to the positive control and the representative chemicals.

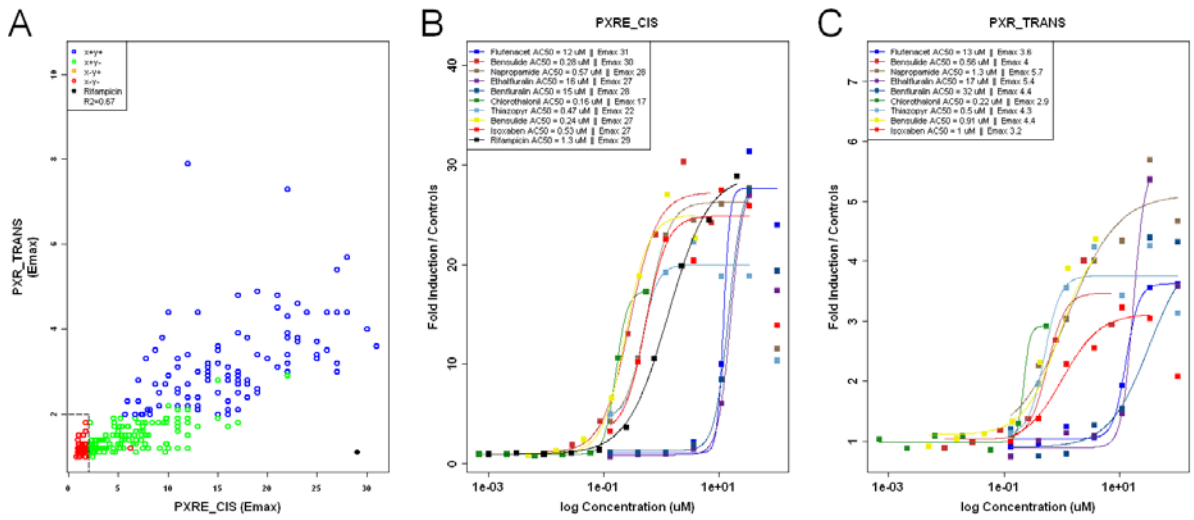


Figure 3.6. Comparison of retinoic acid receptor (RAR) activity in the CIS (DR5) and TRANS (RARα) format characterizes both relative efficacy and potency across the chemical library. Maximum efficacy (Emax) plotted for both the CIS and TRANS assays across the chemical library (a). The colored circles represent the chemicals positive (AC50 derived) or negative for the assays on the x- and y-axis. (b. and c.) The top 5 most efficacious and top 5 most potent averaged across both assays demonstrate the relative response compared to the positive control and the representative chemicals.

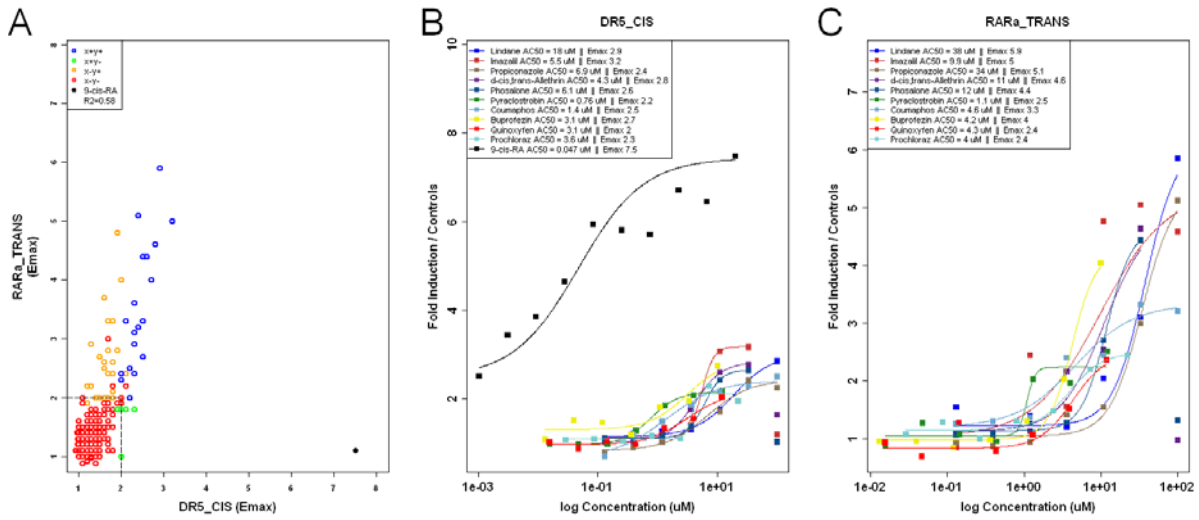


Figure 3.7. Comparison of estrogen receptor (ER) activity in the CIS (ERE) and TRANS (ERα) format characterizes both relative efficacy and potency across the chemical library. Maximum efficacy (Emax) plotted for both the CIS and TRANS assays across the chemical library (a). The colored circles represent the chemicals positive (AC50 derived) or negative for the assays on the x- and y-axis. (b. and c.) The top 5 most efficacious and top 5 most potent averaged across both assays demonstrate the relative response compared to the positive control and the representative chemicals.

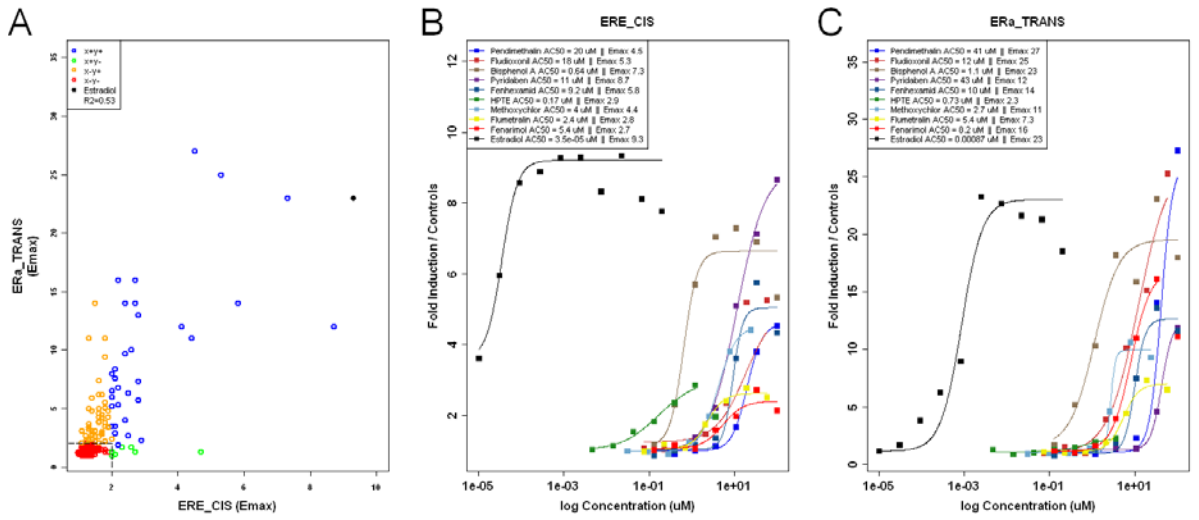


Figure 3.8. Comparison of peroxisome proliferator-activated receptor (PPAR) activity in the CIS (PPRE) and TRANS (PPARγ) format characterizes both relative efficacy and potency across the chemical library. Maximum efficacy (Emax) plotted for both the CIS and TRANS assays across the chemical library (a). The colored circles represent the chemicals positive (AC50 derived) or negative for the assays on the x- and y-axis. (b. and c.) The top 5 most efficacious and top 5 most potent averaged across both assays demonstrate the relative response compared to the positive control and the representative chemicals.

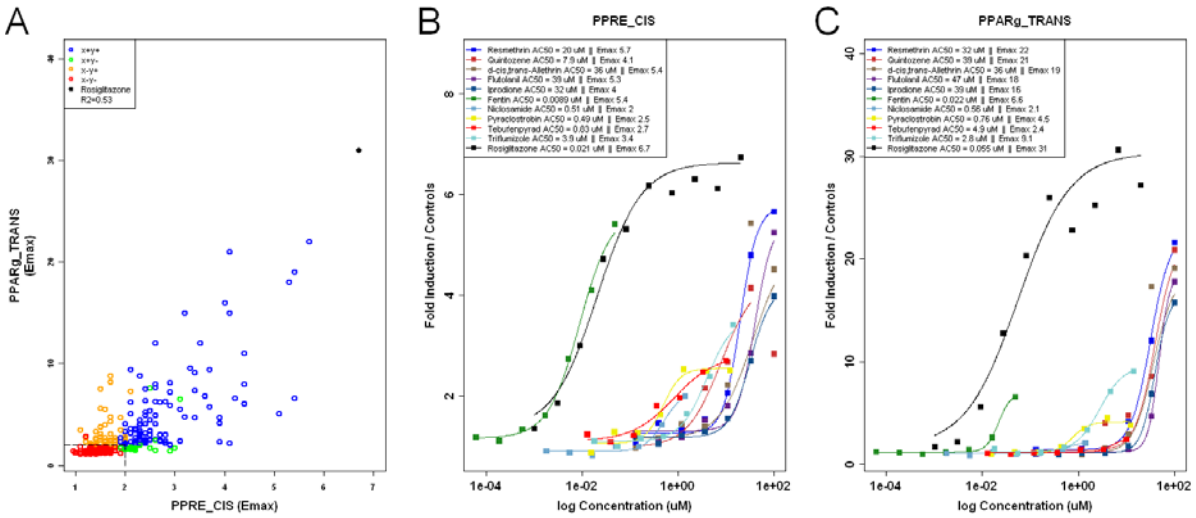


Figure 3.9. Oxidative stress is a plausible explanation for non-specific activity of some chemicals. (a.) Nrf2, a marker of oxidative stress, explains fifty percent of the variability in the overall activity across the remaining 72 targets probed in this study, both CIS and TRANS, based on the log-transformed maximal efficacy values (Emax). Emax values were log-transformed to minimize the effect of varying dynamic ranges across the assays. (b.) Representative dose response curves of the 10 most efficacious chemicals demonstrate the relative dose-response across selected chemicals.

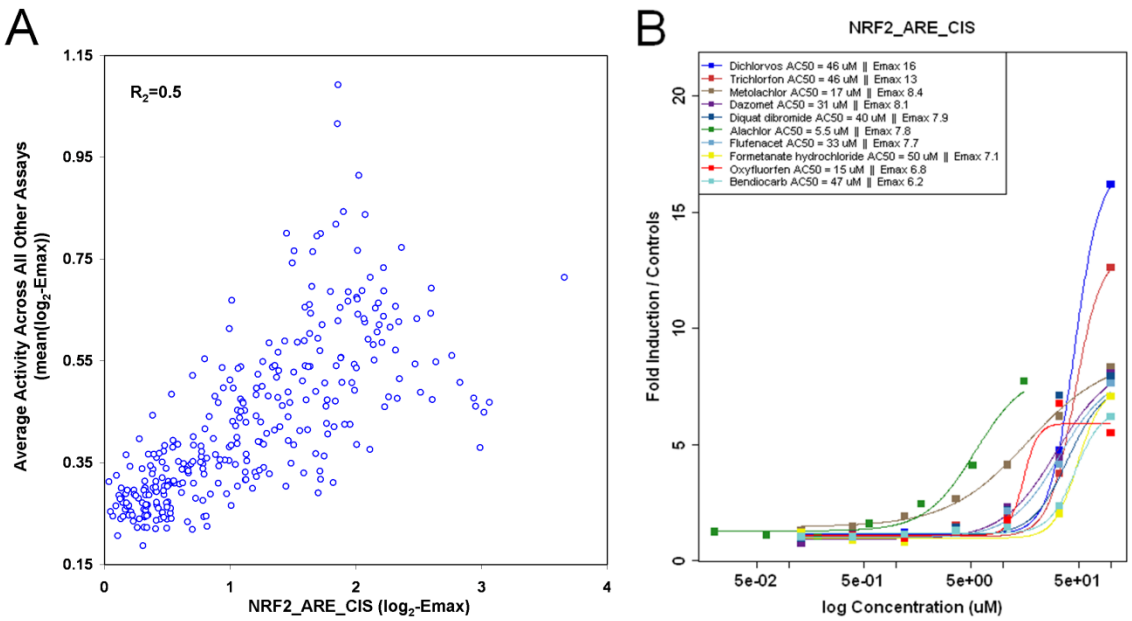


Figure 3.10. Two-way hierarchical clustering of the relative risk percentiles from the permutation test exhibits clustering of developmental and reproductive endpoints across species and study types and grouping of systemic endpoints across chronic and multigeneration studies. Of the 73 *in vitro* and 77 *in vivo* endpoints, 133 significant ($p<0.05$) assay/endpoint associations were established.

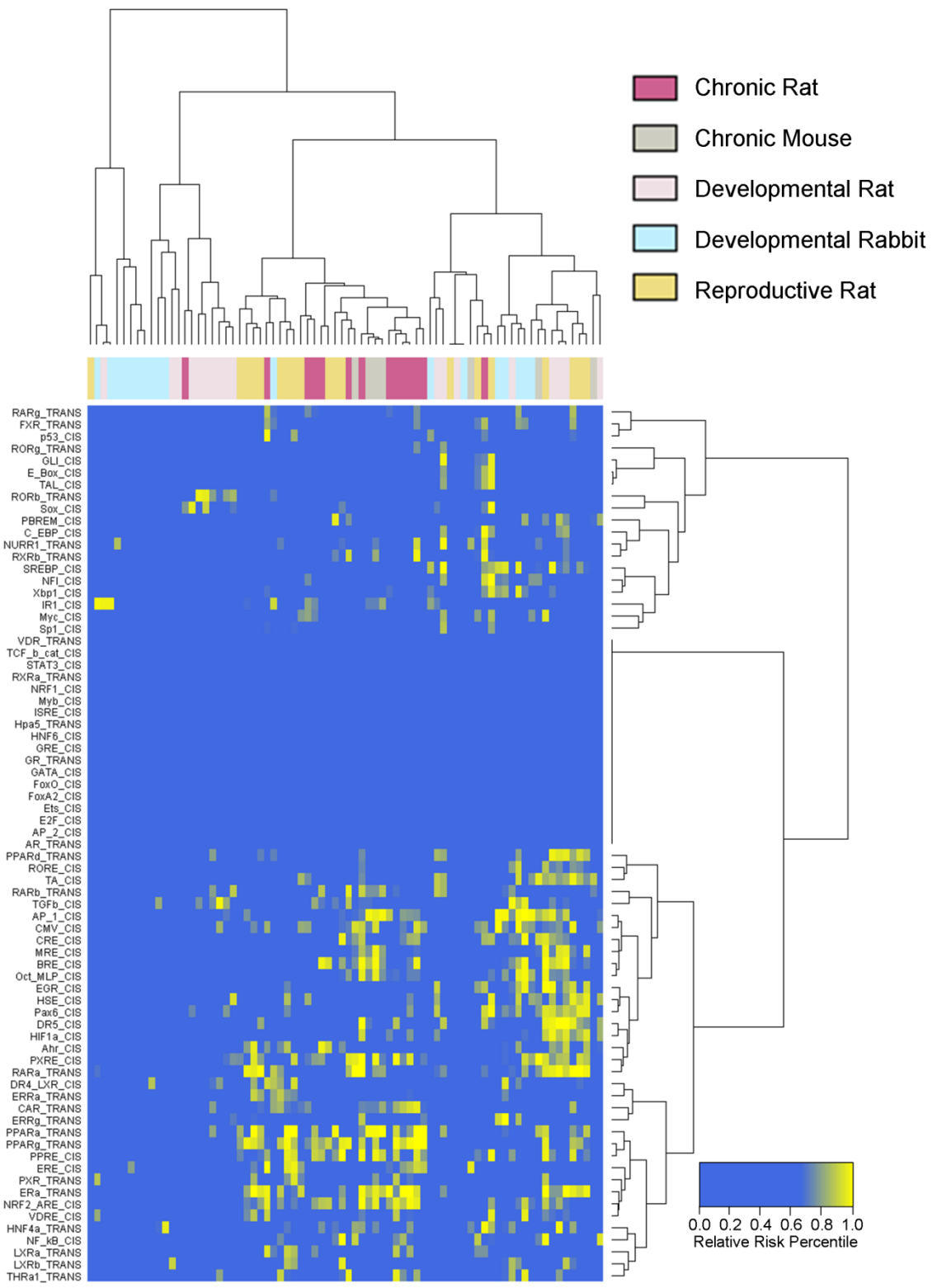
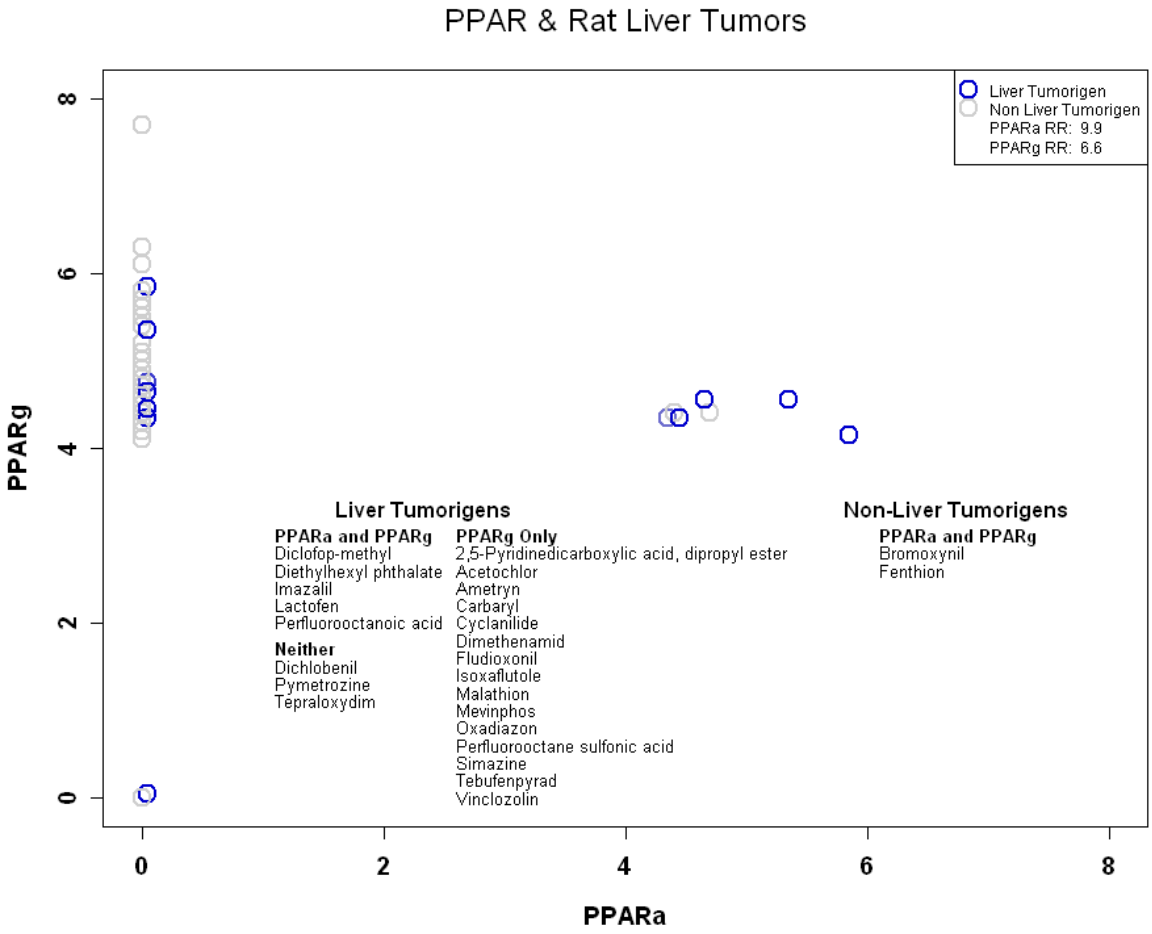


Figure 3.11. Associations were derived between PPAR activity, both alpha (a) and gamma (g), and rat liver tumorigenesis. In total, 5 of the 7 chemicals that significantly affected

PPARa and PPARg caused rat liver tumors in the chronic/cancer bioassay. Individual relative risks of 9.9 and 6.6 were established based on significant PPARa and PPARg activity, respectively.



References

Aranda, A. and Pascual, A. (2001). Nuclear hormone receptors and gene expression. *Physiol Rev* **81**, 1269-304.

Bettoun, D. J., Burris, T. P., Houck, K. A., Buck, D. W., 2nd, Stayrook, K. R., Khalifa, B., Lu, J., Chin, W. W. and Nagpal, S. (2003). Retinoid X receptor is a nonsilent major contributor to vitamin D receptor-mediated transcriptional activation. *Mol Endocrinol* **17**, 2320-8, 10.1210/me.2003-0148
me.2003-0148 [pii].

Burman, D. M., Shertzer, H. G., Senft, A. P., Dalton, T. P. and Genter, M. B. (2003). Antioxidant perturbations in the olfactory mucosa of alachlor-treated rats. *Biochem Pharmacol* **66**, 1707-1715, Doi 10.1016/S0006-2952(03)00475-1.

Butler, E. G., Tanaka, T., Ichida, T., Maruyama, H., Leber, A. P. and Williams, G. M. (1988). Induction of hepatic peroxisome proliferation in mice by lactofen, a diphenyl ether herbicide. *Toxicol Appl Pharmacol* **93**, 72-80, 0041-008X(88)90026-9 [pii].

Chambon, P. (1996). A decade of molecular biology of retinoic acid receptors. *FASEB J* **10**, 940-54.

Collins, F. S., Gray, G. M. and Bucher, J. R. (2008). Toxicology - Transforming environmental health protection. *Science* **319**, 906-907, DOI 10.1126/science.1154619.

DeWitt, J. C., Shnyra, A., Badr, M. Z., Loveless, S. E., Hoban, D., Frame, S. R., Cunard, R., Anderson, S. E., Meade, B. J., Peden-Adams, M. M., Luebke, R. W. and Luster, M. I. (2009). Immunotoxicity of perfluorooctanoic acid and perfluorooctane sulfonate and the role of peroxisome proliferator-activated receptor alpha. *Crit Rev Toxicol* **39**, 76-94, 902566214 [pii]
10.1080/10408440802209804.

Dix, D. J., Houck, K. A., Martin, M. T., Richard, A. M., Setzer, R. W. and Kavlock, R. J. (2007). The ToxCast program for prioritizing toxicity testing of environmental chemicals. *Toxicol Sci* **95**, 5-12, kfl103 [pii]
10.1093/toxsci/kfl103.

Donato, M. T., Lahoz, A., Castell, J. V. and Gomez-Lechon, M. J. (2008). Cell lines: A tool for in vitro drug metabolism studies. *Curr Drug Metab* **9**, 1-11.

Ekins, S., Mirny, L. and Schuetz, E. G. (2002). A ligand-based approach to understanding selectivity of nuclear hormone receptors PXR, CAR, FXR, LXRalpha, and LXRbeta. *Pharm Res* **19**, 1788-800.

Gaido, K. W., Maness, S. C., McDonnell, D. P., Dehal, S. S., Kupfer, D. and Safe, S. (2000). Interaction of methoxychlor and related compounds with estrogen receptor alpha and beta, and androgen receptor: structure-activity studies. *Mol Pharmacol* **58**, 852-8.

Gentleman, R., Hornik, K. and Leisch, F. (2002). R 1.5 and the Bioconductor 1.0 releases. *Comput Stat Data An* **39**, 557-558.

Gillam, E. M. (2002). Coordinating clearance - the many phases of PXR! *Trends Pharmacol Sci* **23**, 548-9, S0165614702021582 [pii].

Goodwin, B., Hodgson, E. and Liddle, C. (1999). The orphan human pregnane X receptor mediates the transcriptional activation of CYP3A4 by rifampicin through a distal enhancer module. *Mol Pharmacol* **56**, 1329-39.

Gould, J. C., Leonard, L. S., Maness, S. C., Wagner, B. L., Conner, K., Zacharewski, T., Safe, S., McDonnell, D. P. and Gaido, K. W. (1998). Bisphenol A interacts with the estrogen receptor alpha in a distinct manner from estradiol. *Mol Cell Endocrinol* **142**, 203-14.

Grandori, C., Cowley, S. M., James, L. P. and Eisenman, R. N. (2000). The Myc/Max/Mad network and the transcriptional control of cell behavior. *Annu Rev Cell Dev Bi* **16**, 653-699.

Gray, L. E., Jr., Ostby, J., Ferrell, J., Rehnberg, G., Linder, R., Cooper, R., Goldman, J., Slott, V. and Laskey, J. (1989). A dose-response analysis of methoxychlor-induced alterations of reproductive development and function in the rat. *Fundam Appl Toxicol* **12**, 92-108.

Houck, K. A. and Kavlock, R. J. (2008). Understanding mechanisms of toxicity: Insights from drug discovery research. *Toxicol Appl Pharm* **227**, 163-178, DOI 10.1016/j.taap.2007.10.022.

Karademir, A., Durmusoglu, E. and Bakoglu, M. (2007). Health risk assessment of background PCDD/F exposure levels in Kocaeli, Turkey. *J Environ Sci Health A Tox Hazard Subst Environ Eng* **42**, 729-39, 778061313 [pii] 10.1080/10934520701304393.

Kliewer, S. A., Goodwin, B. and Willson, T. M. (2002). The nuclear pregnane X receptor: a key regulator of xenobiotic metabolism. *Endocr Rev* **23**, 687-702.

Knudsen, T. B., Martin, M. T., Kavlock, R. J., Judson, R. S., Dix, D. J. and Singh, A. V. (2009). Profiling the activity of environmental chemicals in prenatal developmental toxicity studies using the US EPA's ToxRefDB. *Reprod Toxicol* **28**, 209-219, DOI 10.1016/j.reprotox.2009.03.016.

Lai, D. Y. (2004). Rodent carcinogenicity of peroxisome proliferators and issues on human relevance. *J Environ Sci Heal C* **22**, 37-55, Doi 10.1081/Gnc-120038005.

Lau, C., Butenhoff, J. L. and Rogers, J. M. (2004). The developmental toxicity of perfluoroalkyl acids and their derivatives. *Toxicol Appl Pharmacol* **198**, 231-41, 10.1016/j.taap.2003.11.031
S0041008X04000468 [pii].

le Maire, A., Grimaldi, M., Roecklin, D., Dagnino, S., Vivat-Hannah, V., Balaguer, P. and Bourguet, W. (2009). Activation of RXR-PPAR heterodimers by organotin environmental endocrine disruptors. *EMBO Rep* **10**, 367-73, embor20098 [pii]
10.1038/embor.2009.8.

Lee, J. M. and Johnson, J. A. (2004). An important role of Nrf2-ARE pathway in the cellular defense mechanism. *J Biochem Mol Biol* **37**, 139-43.

Lemay, D. G. and Hwang, D. H. (2006). Genome-wide identification of peroxisome proliferator response elements using integrated computational genomics. *J Lipid Res* **47**, 1583-7, M500504-JLR200 [pii]
10.1194/jlr.M500504-JLR200.

Maeda, Y., Dave, V. and Whitsett, J. A. (2007). Transcriptional control of lung morphogenesis. *Physiological Reviews* **87**, 219-244, DOI 10.1152/physrev.00028.2006.

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**, 578-90, 10.1021/tx900325g.

Martin, M. T., Judson, R. S., Reif, D. M., Kavlock, R. J. and Dix, D. J. (2009a). Profiling Chemicals Based on Chronic Toxicity Results from the US EPA ToxRef Database. *Environ Health Persp* **117**, 392-399, Doi 10.1289/Ehp.0800074.

Martin, M. T., Mendez, E., Corum, D. G., Judson, R. S., Kavlock, R. J., Rotroff, D. M. and Dix, D. J. (2009b). Profiling the Reproductive Toxicity of Chemicals from Multigeneration Studies in the Toxicity Reference Database. *Toxicol Sci* **110**, 181-190, DOI 10.1093/toxsci/kfp080.

Melnick, R. L. (2001). Is peroxisome proliferation an obligatory precursor step in the carcinogenicity of di(2-ethylhexyl)phthalate (DEHP)? *Environ Health Perspect* **109**, 437-42, sc271_5_1835 [pii].

Mosmann, T. (1983). Rapid colorimetric assay for cellular growth and survival: application to proliferation and cytotoxicity assays. *J Immunol Methods* **65**, 55-63, 0022-1759(83)90303-4 [pii].

Nguyen, T., Nioi, P. and Pickett, C. B. (2009). The Nrf2-antioxidant response element signaling pathway and its activation by oxidative stress. *J Biol Chem* **284**, 13291-5, R900010200 [pii]
10.1074/jbc.R900010200.

NRC (2007). Toxicity Testing in the 21st Century: A Vision and a Strategy. National Academies Press, Washington D.C.

Peixoto, F., Alves-Fernandes, D., Santos, D. and Fontainhas-Fernandes, A. (2006). Toxicological effects of oxyfluorfen on oxidative stress enzymes in tilapia Oreochromis niloticus. *Pestic Biochem Phys* **85**, 91-96, DOI 10.1016/j.pestbp.2005.10.007.

Reifel-Miller, A., Otto, K., Hawkins, E., Barr, R., Bensch, W. R., Bull, C., Dana, S., Klausing, K., Martin, J. A., Rafaeloff-Phail, R., Rafizadeh-Montrose, C., Rhodes, G., Robey, R., Rojo, I., Rungta, D., Snyder, D., Wilbur, K., Zhang, T., Zink, R., Warshawsky, A. and Brozinick, J. T. (2005). A peroxisome proliferator-activated receptor alpha/gamma dual agonist with a unique in vitro profile and potent glucose and lipid effects in rodent models of type 2 diabetes and dyslipidemia. *Mol Endocrinol* **19**, 1593-605, me.2005-0015 [pii]
10.1210/me.2005-0015.

Richert, L., Price, S., Chesne, C., Maita, K. and Carmichael, N. (1996). Comparison of the induction of hepatic peroxisome proliferation by the herbicide oxadiazon in vivo in rats, mice, and dogs and in vitro in rat and human hepatocytes. *Toxicol Appl Pharmacol* **141**, 35-43, S0041-008X(96)90257-4 [pii]
10.1006/taap.1996.0257.

Romanov, S., Medvedev, A., Gambarian, M., Poltoratskaya, N., Moeser, M., Medvedeva, L., Diatchenko, L. and Makarov, S. (2008). Homogeneous reporter system enables quantitative functional assessment of multiple transcription factors. *Nat Methods* **5**, 253-60, nmeth.1186 [pii]
10.1038/nmeth.1186.

Shipley, J. M., Hurst, C. H., Tanaka, S. S., DeRoos, F. L., Butenhoff, J. L., Seacat, A. M. and Waxman, D. J. (2004). trans-activation of PPARalpha and induction of PPARalpha target genes by perfluorooctane-based chemicals. *Toxicol Sci* **80**, 151-60, 10.1093/toxsci/kfh130 kfh130 [pii].

Stajner D., M.-D. N., Milosevic M., and Zlokolica M. (2001). Herbicide stress and its effect on scavenging system of vegetable sorts. *Oxidation Communications* **24**, 602-608.

Takacs, M. L. and Abbott, B. D. (2007). Activation of mouse and human peroxisome proliferator-activated receptors (alpha, beta/delta, gamma) by perfluorooctanoic acid and perfluorooctane sulfonate. *Toxicol Sci* **95**, 108-17, kfl135 [pii]
10.1093/toxsci/kfl135.

USEPA (2003). Proposed OPPTS Science Policy: PPAR-Mediated Hepatocarcinogenesis in Rodents and Relevance to Human Health Risk Assessments.http://www.epa.gov/scipoly/SAP/meetings/2003/december9/peroxisomeproliferatorsciencepolicypaper.pdf

USEPA (2008). DSSTox EPA ToxCast™ High Throughput Screening Testing Chemicals Structure-Index File: SDF File and Documentation.http://www.epa.gov/NCCT/dsstox/sdf_toxcst.html

USEPA (2009a). Pesticide Reregistration Status.http://www.epa.gov/oppsrrd1/reregistration/status.htm

USEPA (2009b). ToxCast Homepage.http://www.epa.gov/ncct/toxcast

USFDA (2008). Guidance for Industry, Diabetes Mellitus: Developing Drugs and Therapeutic Biologics for Treatment and Prevention.http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm071624.pdf

WHO (2002). Global assessment of the state-of-the-science of endocrine disruptors. International Programme on Chemical Safety.http://www.who.int/ipcs/publications/en/ch1.pdf

Wilkinson, G. R. (2005). Drug therapy - Drug metabolism and variability among patients in drug response. *New Engl J Med* **352**, 2211-2221.

Wilson, V. S., Bobseine, K. and Gray, L. E., Jr. (2004). Development and characterization of a cell line that stably expresses an estrogen-responsive luciferase reporter for the detection of estrogen receptor agonist and antagonists. *Toxicol Sci* **81**, 69-77, 10.1093/toxsci/kfh180 kfh180 [pii].

Witorsch, R. J. (2002). Endocrine disruptors: Can biological effects and environmental risks be predicted? *Regul Toxicol Pharm* **36**, 118-130, DOI 10.1006/rtph.2002.1564.

Wolfgang, G. H., Jolly, R. A., Donarski, W. J. and Petry, T. W. (1991). Inhibition of diquat-induced lipid peroxidation and toxicity in precision-cut rat liver slices by novel antioxidants. *Toxicol Appl Pharmacol* **108**, 321-9, 0041-008X(91)90121-T [pii].

Yarsan, E. and Cakir, O. (2006). Effects of dichlorvos on lipid peroxidation in mice on subacute and subchronic periods. *Pestic Biochem Phys* **86**, 106-109, DOI 10.1016/j.pestbp.2006.02.002.

CHAPTER 4

PREDICTIVE MODEL OF RAT REPRODUCTIVE TOXICITY FROM TOXCAST HIGH THROUGHPUT SCREENING[3]

Abstract

The EPA ToxCast research program uses high throughput screening for bioactivity profiling and predicting the toxicity of large numbers of chemicals. ToxCast Phase I tested 309 well-characterized chemicals in over 500 assays for a wide range of molecular targets and cellular responses. Of the 309 environmental chemicals in Phase I of ToxCast, 256 were linked to high quality rat multigeneration reproductive toxicity studies in the relational Toxicity Reference Database. Reproductive toxicants were defined here as having achieved a reproductive lowest observed adverse effect level less than 500 milligram per kilogram of body weight per day. 86 chemicals were identified as reproductive toxicants in rat; 68 of those with sufficient *in vitro* bioactivity to model. Each assay was assessed for univariate association with the identified reproductive toxicants. Significantly associated assays were linked to gene sets and used for the subsequent predictive modeling. Using linear discriminant analysis and five-fold cross-validation, a robust and stable predictive model was produced capable of identifying rodent reproductive toxicants with 77±2% and 74±5% training and test cross-validation balanced accuracies, respectively. With a 21 chemical external validation set the model was 76% accurate, further indicating the model's potential

[3] Martin, M. T., Knudsen, T. B., Reif, D. M., Houck, K. A., Judson, R. S., Kavlock, R. J. and Dix, D. J. (2011). Predictive Model of Rat Reproductive Toxicity from ToxCast High Throughput Screening. *Biol Reprod*, biolreprod.111.090977 [pii] 10.1095/biolreprod.111.090977.

for prioritizing the many thousands of environmental chemicals with little to no hazard information. The biological features of the model include steroidal and non-steroidal nuclear receptors, cytochrome P450 enzyme inhibition, G protein-coupled receptors, and cell signaling pathway readouts- mechanistic information suggesting additional targeted, integrated testing strategies and potential applications of *in vitro* HTS to risk assessment.

Introduction

Current chemical evaluations in the U.S. range from either little to no evidence of safety for most industrial chemicals, or an expensive battery of animal tests for food-use pesticides that offers little mechanistic insights. No *in vivo* toxicology test uses more animals than the rat multigeneration reproductive test. It has been estimated that 70% of the total cost and 90% of the animal use for compliance with REACH will be due to reproductive toxicity testing (Hartung and Rovida, 2009). Addressing the existing chemical evaluation bottleneck can only be achieved through progressive changes to the current animal testing paradigm. A promising resource for addressing this bottleneck is computational toxicology, a field that integrates tools from computer science, bio- and chemi-informatics, molecular biology, and high throughput screening. Currently prescribed in *vivo* tests for chemical toxicity are resource-intensive, particularly for multigenerational reproductive and prenatal developmental assessment. Policy directives such as the European Union's (EU) Cosmetics Directive call for the elimination of animals for evaluating reproductive toxicity in 2013 for cosmetic products and development of alternative methods for safety evaluation. In the past, significantly less attention has been spent modeling or predicting chemical-induced reproductive toxicity, relative to efforts modeling cancer and other endpoints. Reasons for the meager effort in this area include a lack of reference animal toxicity data to model, as well as

the molecular and physiological complexity of maternal-fetal interactions, life-stages, and generational sensitivities (Cronin and Worth, 2008). Recent efforts capturing *in vivo* reproductive toxicity studies into databases, and *in vitro* bioactivity profiling have enabled the development of predictive, mechanistic and pathway-based models for these complex reproductive outcomes.

ToxRefDB, the Toxicity Reference Database, has been the primary tool for storing and accessing high quality toxicology studies and is available online for searching and download (USEPA, 2009). ToxRefDB has characterized thousands of studies using a standardized vocabulary, a uniform structure across study types, and a high level of internal and external quality control for the extraction of endpoints useful in developing predictive models (Martin *et al.*, 2009a). The primary study for assessing reproductive effects of chemicals is the multigenerational reproductive test (OPPTS 870.3800 and OECD 416), and is typically conducted under continuous exposure to male and female rats from 10-weeks pre-mating through lactation in the second generation. From multigeneration reproductive studies in ToxRefDB we have the capacity to identify individual or aggregated endpoints for predictive modeling across hundreds of chemicals and have made comparisons across generations to identify adverse impacts on developmentally sensitive reproductive endpoints, based on the prevalence of specific endpoints at later generations compared to the first generation (Martin *et al.*, 2009b). Generational comparisons using ToxRefDB have also been part of the OECD evaluation of the proposed Extended One-Generation Reproductive Toxicity Study (EOGRTS). ToxRefDB was the primary database used in the large scale retrospective analysis aimed at evaluating the impact of the second generation on risk assessments, and classification & labeling (C&L) in Europe (Piersma *et al.*, 2010). However,

the acceptance of the EOGRTS in lieu of the existing two generation test will not alleviate the chemical testing bottleneck for the many thousands of chemicals in commerce. One set of solutions to this testing bottleneck are alternative methods for chemical prioritization and intelligent, targeted testing decisions.

The use of alternative methods as part of an integrated reproductive and developmental toxicity testing strategy is currently being developed as a battery of *in silico*, *in vitro*, and *in vivo* tests (Hareng *et al.*, 2005; Spielmann, 2009). One component of this toolbox is the large-scale bioactivity profiling of chemicals in high throughput screening (HTS) and high-content assays. EPA's ToxCast™ research project has produced a substantial amount of HTS data on environmental chemicals for developing predictive models of toxicity (Dix *et al.*, 2007). Phase I of ToxCast profiled 309 toxicologically well-characterized chemicals in over 500 assays using nine technologies, including cell-free HTS assays and cell-based assays. ToxCast HTS data and multigenerational reproductive toxicity data from ToxRefDB provides an effective dataset for developing predictive toxicology models. In this study, we present a robust and stable predictive model of chemically-induced reproductive toxicity that demonstrates external predictivity useful for targeted testing prioritizations and significantly advancing predictive and computational toxicology.

Methods

Chemical

Phase I of the EPA ToxCast™ program employs a chemical library containing 320 samples consisting of 309 unique structures, 5 duplicates that were differently sourced and 3 triplicates as technical repeats for internal quality control. The rationale for chemical selection was based on several criteria: extensive chronic, cancer, multigenerational

reproductive, and developmental assay data available (95% of compounds meet this criteria); soluble in DMSO (-1 < logP < 6, i.e., log of the octanol/water partition coefficient; 97.5% meet this criteria); molecular weight range 250-1000 (90% meet this criteria); and commercially available with purity >90% (98% meet this criteria). These criteria were largely satisfied with a diverse set of pesticide active ingredients that have had guideline *in vivo* toxicology studies conducted as part of their registration process with the EPA. Several other miscellaneous chemicals of environmental concern meeting these criteria were also included in the library. Despite its large representation of pesticidal actives, the Phase I chemical library spans a wide range of property values and is quite structurally diverse, representing over 40 chemical functional classes (e.g., pyrazole, sulfonamide, organochlorine, pyrethroid, etc.) and over 24 known pesticidal mode-of-action classes (e.g., phenylurea herbicides, organophosphate insecticides, dinitroaniline herbicides, etc.). A complete listing of the quality reviewed and structure-annotated chemical library is available for download as a Structure Data Format (SDF) file at the DSSTox website (USEPA, 2008).

Chemicals comprising the ToxCast™ Phase I library were commercially procured and plated by BioFocus DPI (South San Francisco, CA). Supplier-provided certificates of analysis indicated purity >97% for the large majority of chemicals (87%), and >90% purity for all but a few instances of technical grade or known mixtures. Follow-up analysis of an original solution plate by BioFocus DPI using LC/MS (liquid chromatography mass spectrometry), subsequent to assay screening, has confirmed mass identification, stability, and purity for over 83% of the chemical library. For the majority of the remaining chemicals, currently employed analysis methods are known or suspected to be inadequate for confirming sample purity and, for the remaining 8% of the chemicals, follow-up studies have provided

some evidence of sample decomposition in DMSO over time. A QC summary result mapped to chemical solution sample is provided on the ToxCast™ website in association with assay results (U.S. EPA, 2008). All chemicals were included in the analysis regardless of analytics results, but were accounted for throughout the analysis process.

In Vivo (Class Data)

Multigenerational reproductive toxicity testing study design and treatment group information along with all treatment-related effects were manually collected into EPA's Toxicity Reference Database (ToxRefDB). The database structure, standardized vocabulary and ontology, and quality control procedures have been described previously (Martin, et al., 2009a). To date, ToxRefDB has captured 393 acceptable reproductive studies across 353 chemicals equating to 14, 347, 32 one-, two- and three- generation studies, respectively. An acceptable study can be defined as any study that adequately followed the multigeneration testing guideline, primarily determined by regulatory toxicologists from EPA's Office of Pesticide Programs, and that the review of the study contains sufficient information for complete entry into ToxRefDB. Of the 309 ToxCast chemicals, 256 chemicals have been linked to an acceptable reproductive study entered in ToxRefDB with 242 exact structural matches, 4 close structural matches presumed to be toxicological equivalents (e.g., parent-to-salt, salt-to-parent, different isomeric forms) not already linked to a ToxCast chemical, 4 close structural matches already linked to a ToxCast chemical (e.g., Fluazifop-butyl and Fluazifop-P-butyl), and 6 parent-to-metabolite pairs (e.g., Diethylhexyl phthalate and Phthalic acid, mono-2-ethylhexyl ester). An additional 39 chemicals have unacceptable reproductive studies while 14 chemicals have no data available in ToxRefDB.

In ToxRefDB, 650 unique effects were observed across the entire multigenerational reproductive toxicity study dataset, ranging from body weight decreases to organ weight changes to litter survival to fertility decrements. Each unique effect was mapped to one of three multigeneration study categories; parental (e.g., body weight, liver weight, and other systemic toxicities), reproductive (e.g., primarily fertility and early offspring survival), and offspring (e.g., offspring weight, longer-term offspring survival, and other systemic offspring toxicities during their juvenile period). Specifically, 120 effects were directly related to reproductive outcomes and another 175 effects indicated adverse offspring outcomes, with the remainder being systemic parental effects (Martin, et al., 2009b). Based on the review of each study, primarily by regulatory toxicologists from EPA's Office of Pesticide Programs, parental, offspring, and reproductive lowest observed adverse effect levels (LOAEL) were established based on the weight of evidence and expert judgment of the reviewer. The reproductive LOAEL (rLOAEL) was used to delineate a positive and negative set for reproductive toxicity, based on a 500 mg/kg/day cutoff. This cutoff value approximates the testing limit of 1000 mg/kg/day in the reproductive test guideline and accounts for the large uncertainty around the dose intake measurements and standard conversions used in capturing the dosing information across hundreds of chemicals and over 30 years of toxicity testing. Any chemical with a rLOAEL less than or equal to the cutoff was considered a positive and any chemical with a rLOAEL greater than the cutoff or that was not assigned a rLOAEL by the study reviewer was considered to be negative for reproductive toxicity. Specific effects within this endpoint category include reproductive performance measures (e.g., fertility, mating, gestational interval), male and female reproductive tract effects (e.g., testis, epididymis, ovary, uterus pathology and weight along with sperm measures and

morphology), and sexual developmental landmarks (e.g., preputial separation, vaginal opening and anogenital distance). Teratogenic endpoints from prenatal toxicity testing were not included as part of the definition of a reproductive toxicant for the purposes of this modeling effort. Additional information regarding the treatment groups including the life-stage and generation of the animals and the administered dose were captured in ToxRefDB to provide additional context for each chemical's reproductive toxicity potential.

In Vitro (Features)

As part of the ToxCast research program, the chemical library was tested in over 500 assays across 9 technologies, including high throughput cell-free assays and cell-based assays in multiple human and rodent primary and derived cell lines. A complete overview of the assays, assay selection, analysis methods, quality measures, and assay annotation have been previously published (Judson *et al.*, 2010). In general, AC50 (concentration at half maximal efficacy) values or LEC (lowest effective concentrations) were derived for each assay and time-point, where applicable. The complete data set, including AC50/LEC values and corresponding concentration response data for all chemical-assay measurement pairs is available from the EPA ToxCast website (U.S. EPA, 2008). For the purpose of predictive modeling, assays form the input features and can be thought of as the right side of the equation where some linear combination of these assays or sets of assays is equal to the class data; the reference *in vivo* endpoint.

AC50/LEC values were -log3 transformed (-log3 [AC50/1000]), and a value of 0 was given to all negative assay results. A log3 transformation and setting negatives to 1000µM was used over a log10 transformation and setting negatives to one molar, as has been done in previous publications of ToxCast results(Judson, et al., 2010), to enhance the scoring range

between high and low potency active chemicals and to decrease the distance between active and inactive chemicals. Therefore, the "assay score" where the AC50 was 100µM would have a value of roughly 2 while a 100nM AC50 would have a value of roughly 8. A "gene score" or "gene-set score" was derived based on the average assay score across a set of closely related assays, e.g., assays mapped to a single gene or gene family. Any chemical active in fewer than or equal to 10 assays (≤2% aggregate active) was removed from the initial model development due to the lack of information provided by the chemical's bioactivity fingerprint to discern active and inactive for any toxicity. The rationale for excluding the chemicals with little or no *in vitro* activity is based on the following logic. Specific chemicals may lack activity in *in vitro* assays for a number of reasons including chemical degradation, aqueous insolubility, lack of metabolic activation, or volatility. Such chemicals would be characterized by little to no activity across a broad range of *in vitro* assays. Since this behavior is, at least to some extent, relevant only to the *in vitro* systems, they are not good candidates for including in a model predicting *in vivo* activity. They were thus excluded from the training set and their exclusion is making no statement of a chemical's true reproductive toxicity potential.

Model (Class ~ Features)

The first step in the development of a predictive model was univariate feature selection. Each assay was compared to the training set of chemicals, positive and negative for reproductive toxicity, using continuous and dichotomous statistical methodology, including linear (Pearson's) correlation test, chi-square test, and t-test with the level of significance returned as p-values. Each assay with a p-value of less than 0.1 from any method passed the initial feature selection filter. The resulting assays were then grouped by gene or assay

family, as described above, to form the input features for subsequent modeling. In some instances, assays that were not statistically significantly associated but provided orthogonal or complimentary readouts for the same target were included in model development. This was performed for various nuclear receptor targets in which cell-based transcription factor assays were significantly associated with reproductive toxicity, whereas the more specific cell-free binding assays were not due to the low number of active chemicals. The highly specific assays provide increased evidence that a chemical interacts with a particular target. Significantly associated assays that were part of a large assay family or that were highly correlated to other higher prioritized assays, based on relative p-value and correlation, were excluded to minimize the total number of assays moving into the model development phase. For example, as part of ToxCast, 54 GPCR (G protein-coupled receptor) binding assays were evaluated, with 18 being significantly associated with reproductive toxicity. Of those 18 GPCR assays, 5 were selected based on having the greatest correlation collectively; adding further GPCR assays only lowered the overall association to reproductive toxicity.

Based on the selection of a small and balanced feature set, the prediction of reproductive toxicity potential was performed using linear discriminant analysis (LDA). Five-fold cross-validation was used to explore the stability of the resultant model, a process of developing the model using 80% of the chemical set and testing the model accuracy with the remaining 20% and repeating five times until all data has been used as both training and test datasets. The resulting cross-validation statistics are presented as the average and standard deviation of the training and test set balanced accuracies across the five runs. In addition, a subset of chemicals with positive findings in unacceptable studies within ToxRefDB and chemicals with clear literature evidence of reproductive toxicity or not were

used to assess the forward predictivity of the resultant model and to serve as an initial external validation set.

Results

The quality and forward predictivity of any model is limited by the quality of the feature and class data being used in the model development process. Therefore, strict and transparent methods were used for identifying the training set used in the initial modeling effort from both *in vivo* (i.e., class) data and *in vitro* (i.e., feature) data perspectives. Of the 256 chemicals linked to an acceptable reproductive study, 86 reported a reproductive LOAEL (rLOAEL) less than 500 mg/kg/day (Table 4.1). The additional 12 chemicals that reported an rLOAEL from an unacceptable reproductive study were not incorporated into the initial model development process, but were used for model assessment and external validation of the model. Six chemicals had rLOAEL above the 500 mg/kg/day cutoff and were considered negative for modeling purposes, including Fluoxastrobin (862 mg/kg/day), Trifloxysulfuron-sodium (631 mg/kg/day), Propoxycarbazone-sodium (1314 mg/kg/day), Oxasulfuron (1115 mg/kg/day), Isoxaben (1000 mg/kg/day), and Propamocarb hydrochloride (1000 mg/kg/day). The toxicity profile for these chemicals primarily consisted of high-dose systemic parental and offspring toxicities leading to confounding sexual developmental landmark findings and early offspring survival decrements. Of the total 98 chemicals identified as reproductive toxicants (i.e., 86 from acceptable and 12 from unacceptable studies), 49 chemicals observed treatment-related effects to the male and/or female reproductive tract, 51 chemicals caused decrements in reproductive performance, 67 chemicals affected early offspring survival, and 18 chemicals altered sexual development. A combined model of reproductive toxicity is presented, as opposed to individual models of

each endpoint class, due to the large overlap in chemicals across these endpoint classes, the lack of gender-specific phenotypes, and no mechanistic information in the guideline multigeneration reproductive studies.

A significant number of ToxCast chemicals had little to no *in vitro* activity across hundreds of assays. Aggregate activity for each chemical was calculated as the number of actives (i.e., achieving an AC50 and defined as a hit in the assay) divided by the total number of assays used in this analysis (n=512). A 2% activity cutoff was established based on the minimal impact of aggregate *in vitro* activity on the sensitivity and, to limited degree, specificity of resulting models. In total, 62 chemicals were identified as falling below the 2% cutoff and not used in the initial model development process. Table 4.2 summarizes the chemical counts for each chemical group based on *in vivo* reproductive study acceptability/availability and aggregate *in vitro* activity. The entire chemical library was split into these groups to identify a chemical set with the capacity to develop a stable and robust model without the negative impacts of low *in vivo* multigeneration study quality or potential limited amenability to *in vitro* screening. Thus, chemical group A was selected for the initial development of the predictive reproductive toxicity model, including internal cross-validation. Groups B, C, and D were used to evaluate the stability and to identify current weaknesses, limitations and gaps of the model. Groups E and F have also provided insight into the forward predictivity of the model, based on available open-literature reproductive toxicity studies. In conjunction with Table 4.2, a schematic of the full decision process, including chemical groupings, class definitions (i.e., positive or negative for reproductive toxicity), and final summary model statistics is provided as an overview and guidepost to the remaining more detailed results (Figure 4.1).

Of the 206 chemicals used in the initial development of the predictive reproductive toxicity model (i.e., chemical group A), 68 were identified as reproductive toxicants, roughly one third of the total. In relating the *in vitro* bioactivity to these reproductive toxicants, a set of assays and genes were identified as significant indicators of reproductive toxicity, based on their univariate association. In total, 36 out of over 500 assays were selected for model development and subsequently mapped to genes or gene sets (Table 4.3). The primary genes identified were nuclear receptors, both steroidal and non-steroidal, and included the androgen receptor (AR), estrogen receptor alpha (ERα; ESR1), peroxisome proliferator-activated receptors, alpha (PPARα) and gamma (PPARγ). These molecular targets have extensive literature detailing their role in normal reproductive function as well as reproductive and endocrine toxicity. A number of cytochrome P450 enzyme inhibition (CYP) assays, including aromatase (Cyp19a1), were also significantly associated with the reproductive toxicants. Interestingly, besides the human aromatase assay, rat CYP assays had increased association to the endpoint as compared to the human CYP assays. For the purposes of the model and based on the increase in overall statistical correlation, all associated rodent CYP assay scores, as well as aromatase, were averaged and used as a single feature, called 'CYP'. In addition to these genes and assay sets, individual assays representing cell-based markers of growth factor stimulation and cell signaling, including epidermal growth factor 1 (EGFR1) , transforming growth factor beta 1 (TGF-β1), vesicular monoamine transporter 2 (VMAT2), and nuclear factor kappa-light-chain-enhancer of activated B cells (NF-κB) were other positive indicators of reproductive toxicity potential. These assays were also averaged together as a miscellaneous set of assays and called 'OTHER'. As part of ToxCast, 54 GPCR (G protein-coupled receptor) binding assays were evaluated, with 18 being significantly

associated with reproductive toxicity. Of those 18 GPCR assays, 5 were selected based on having the greatest correlation collectively; adding further GPCR assays only lowered the overall association to reproductive toxicity. Assays targeting the pregnane X receptor (PXR) were negatively correlated with reproductive toxicity potential and used in the model development process with the expectation of providing some indication of the metabolic clearance of the chemical or representing general nuclear receptor promiscuity.

Using the combination of the selected gene/gene-set scores, a multivariate linear classifier was developed using linear discriminant analysis (LDA) and five-fold cross validation. The feature set included PPARα (average -log3(AC50/1000) across 3 assays), AR (average -log3(AC50/1000) across 3 assays), ERα (average -log3(AC50/1000) across 7 assays), PPARγ (average -log3(AC50/1000) across 4 assays), CYP (average -log3(AC50/1000) across 7 assays), GPCR (average -log3(AC50/1000) across 5 assays), OTHER (average -log3(AC50/1000) across 4 assays), and PXR (average -log3(AC50/1000) across 3 assays) for a total of eight features. Figure 4.2 demonstrates the relative impact on classification rates between individual assays, genes/gene-sets and the final model. In general, we find that aggregating multiple related assays into a single feature increased the classification rate and yielded a more balanced and stable model. Grouping the assays by gene and gene-sets also allows for assays with low hit prevalence that would otherwise not be included in the model to contribute to the overall assessment of whether or not a chemical interacts with a specific molecular target.

Using the eight gene and gene-set features a robust (i.e., high predictivity with high balanced accuracy; >70%) and stable (i.e., high test cross-validation and external validation accuracies; >70%) classifier or predictive model was generated as shown by resulting model

statistics (Table 4.4). The cross-validation balanced accuracy (BA, equal to the average of sensitivity and specificity) for the training and test sets, averaged across all 5 runs, were 77% and 74% for both training and test sets with a standard deviation of 2% and 5%, respectively. Conversely, using the single most significantly associated assay per gene or gene-set resulted in training and test balanced accuracies of 71% and 64%, respectively, illustrating the loss in predictivity and model stability when relying on a single assay to represent a molecular target or pathway. After demonstrating stability across the cross-validation runs, a model generated using all 206 Group A chemicals was optimized resulting in a balanced accuracy of 80% (p-value = 4.2E-17), indicating a highly predictive model for reproductive hazard.

Chemical group B was not included in the initial model development due to the lack of *in vitro* bioactivity across hundreds of assays. Interestingly, a comparable prevalence of reproductive toxicants was observed in chemical group B with 18 of the 50 chemicals characterized as actives (36% active versus 64% inactive). Only 20 chemicals in group B were active across any of the 33 assays or 7 input features that positively indicated reproductive toxicity. If the model is applied to chemical group B only, the balanced accuracy is 54% with a very low sensitivity of 11%. If the model is applied to chemical groups A and B, balanced accuracy and sensitivity drop to 75% and 66%, respectively. The diminished model performance, especially in terms of sensitivity when including low *in vitro* activity chemicals provides justification for considering these chemicals outside the domain of *in vitro* biological applicability, akin to the domain analysis performed in structure activity studies, and provide no evidence as to the safety or toxicity of the chemical. In real-world applications of this reproductive toxicity model, chemicals could be identified for follow-up analysis ranging from traditional animal toxicity testing, to additional *in vitro* screening

attempting to address confounding issues such as chemical decomposition, aqueous insolubility, or volatility, to the application of purely *in silico* models.

Chemical groups C and D are comprised of 39 chemicals that have been tested in guideline reproductive studies that were deemed unacceptable for a variety of reasons, including quality of the review, dose selection, and guideline adherence. It would not be expected that these studies were deemed unacceptable due to false positive findings; therefore, the 12 chemicals designated as reproductive toxicants were used to demonstrate external predictivity of the model. Examples of such chemicals include the putative anti-androgen Prochloraz (Laier *et al.*, 2006) and the possible endocrine disrupting chemical Fenitrothion (Okahashi *et al.*, 2005), both of which were predicted to be positive for reproductive toxicity. In total, 7 of the 12 reproductive toxicants in chemicals groups C or D were predicted to be positive. The same presumption for the positive findings cannot be extended to the negative findings across studies flagged as unacceptable. For example, the male reproductive toxicant boric acid (Chapin and Ku, 1994) caused only limited reproductive effects in the unacceptable guideline multigeneration reproductive study and showed little *in vitro* activity (chemical group D), possibly due to limited amenability to HTS.

Chemical groups E and F have no guideline-based multigeneration reproductive toxicity study entered into ToxRefDB and in most cases have never had such a study performed. However, of the 14 chemicals in groups E and F, 9 were linked to reproductive toxicity tests available in the open literature. Varying sources and degrees of evidence can be found for reproductive toxicity; Methoxychlor and its metabolite HPTE based on positive findings in numerous pubertal and other *in vivo* assays (Akingbemi *et al.*, 2000; Armenti *et*

al., 2008; Chapin *et al.*, 1997; Uzumcu *et al.*, 2006), Bromoxynil based on EU labeling as a reproductive toxicant (R62), and Methyl Cellusolve (2-Methoxyethanol) based on reproductive findings in multiple systemic repeat-dose and multigenerational studies (Canada, 1999), and Monocrotophos based on male and female reproductive toxicity across multiple studies (Rao and Kaliwal, 2002; Ratnasooriya *et al.*, 1996). Equivocal evidence of reproductive toxicity could be found for Alachlor (USEPA, 1998a), based on non-dose dependent effects on ovarian weight and pregnancy index effects, which did not result in a reproductive LOAEL being determined. Dimethyl phthalate (DMP) and its metabolite Methyl hydrogen phthalate (MHP) (Gray *et al.*, 2000; Kwack *et al.*, 2009) as well as Butralin (USEPA, 1998b) were considered to be negative for reproductive toxicity based on the available studies. The model correctly divided this subset of chemicals as reproductive toxicants or not with the exception of Monocrotophos, which was in the low *in vitro* activity group (chemical group F). Interestingly, Alachlor, which showed limited evidence of reproductive toxicity, was predicted to be positive and was just above the cutoff or model intercept, which could readily be interpreted as an equivocal prediction. In summary, 5 of 6 chemicals with literature evidence of reproductive or endocrine toxicity were accurately predicted while all 3 negative chemicals were accurately predicted.

The remaining 5 chemicals have no reproductive toxicity information available in the literature and were candidates for forward predictions. Based on the model, Symclosene and Phenoxyethanol were predicted to be negative, but it should be noted that the chemicals had low confidence in their purity from the analytical QC and/or low *in vitro* activity. Three chemicals with no reproductive toxicity data were predicted to be positives, including Diniconazole, Niclosamide, and Clorophene. Diniconazole, similar to many of the other

conazoles, demonstrated CYP inhibition, which was highly associated with decrements in early offspring survival. Niclosamide displayed fairly potent PPARγ agonist activity in multiple assays (top 5 of 309 chemicals for aggregate PPARγ activity), which was associated most with male and female reproductive tract effects. Androgen receptor binding was observed for Clorophene at similar potencies to CYP inhibition findings, which were both associated with delays in sexual development and decrements in reproductive performance. These results provide examples of how *in vitro* screening leading to targeted testing could be used to identify chemicals as potential reproductive toxicants based on model predictions. Additionally, the components of the predictive model have increased associations with specific endpoints and can help make recommendations about study design, including incorporating more sensitive or mechanistic endpoints into the study. A summary of the external validation (i.e., chemicals not used in training or testing the model and that have sufficient ToxRefDB or literature data to confidently classify the chemical as a reproductive toxicant or not) and forward validation (i.e., chemicals in which a prediction has been made but have no available evidence of whether or not the chemical is a reproductive toxicant or not) chemical sets demonstrates the forward predictivity of the model and provides examples of predictions made on chemicals with no reproductive toxicity information available (Table 4.5). Of the 21 external validation chemicals, 12 were accurately predicted as reproductive toxicants, 5 chemicals were incorrectly predicted as negative, and 4 were accurately predicted to be negative resulting in an external validation accuracy of 76% and a balanced accuracy of 85%.

In practice, the use of a predictive reproductive toxicity model can assist in prioritizing further targeted testing. Using chemical group A, we demonstrate the utility of

this model in decision making and how it could assist in alleviating the current chemical testing bottleneck. Depending on prioritization goals, increasing or decreasing the optimal balanced cutoff would alter the specificity, sensitivity and predictivity of the applied model (Figure 4.3). Using a high cutoff, testing the top 30 scoring chemicals would yield 26 reproductive toxicants. On the other hand, to identify the vast majority of reproductive toxicants (57 of the 66 total reproductive toxicants), one would have to test the top 136 of 206 scoring chemicals. If the prioritization task was to follow-up with an expensive and time-consuming multigeneration reproductive study in a short period of time, then a more specific approach (i.e., higher cutoff) may be more appropriate. If the prioritization task was to follow-up with a medium-throughput assay capable of testing many chemicals, then a more sensitive approach (i.e., lower cutoff) could be used, ensuring the testing strategy catches as many potential reproductive toxicants as possible. A maximum sensitivity of 86% and a maximum specificity of 97% are achieved dependent on the cutoff, which can be adjusted to the prioritization task.

Beyond the accurate prediction of reproductive toxicants identified solely from animal studies, we have compiled the available European Union (EU) C&L for reproductive toxicity (R60&62 for fertility and R61&63 for developmental toxicity) in Table 4.6. Of the 206 group A chemicals, 19 have been reviewed for EU classification with 7 being classified for fertility (R60&62), 8 classified for developmental toxicity (R61&63), and 4 classified for neither. In all, 14 of the 15 R60-63 classified chemicals were predicted by the current model to be positive. Only the metabolically activated Benomyl was a false negative using the predictive model (Lim and Miller, 1997). All 4 non-classified chemicals were predicted to be negative, but it should be noted that these chemicals could have been unclassified due to

insufficient data to assess C&L. As opposed to the risk assessment process where quantitative dose response information is needed, the C&L process evaluates the intrinsic hazard of a substance. The output of the predictive reproductive toxicity model appears well suited to C&L.

Discussion

The results of this analysis demonstrate that *in vitro* HTS data can be used to predict developmentally sensitive reproductive toxicity in the rat. The capacity to use ToxCast HTS data, costing roughly $20,000-$30,000 per chemical for over 500 assays, in predicting the reproductive toxicity of 100s to 1000s of chemicals could transform the way in which chemicals are prioritized and selected for targeted reproductive toxicity testing. Reproductive toxicity testing is animal intensive, time-consuming, and costly. Current testing requirements are expanding internationally beyond conventional pesticides to industrial chemicals and other chemical domains. Past, present and future multigenerational reproductive studies characterize reproductive toxicity through the integrated assessment of over 100 potential endpoints across varying life-stages and generations. Even with these large numbers of measured endpoints, the imprecise nature of many of the endpoints limits the ability to identify gender and life-stage specificity, let alone mechanisms of action. The complexity of the biology, physiology, and study design are primary reasons for using molecular and cellular markers to model reproductive toxicity, but these complexities are also the reasons previous modeling efforts have not shown dramatic success. Therefore, we have focused not only on the model development, but also on the detailed capture and uniform assessment of the reference *in vivo* reproductive toxicity information leading to a predictive and

biologically relevant model that can be applied not just to testing prioritization but also testing refinement or even replacement.

The overall accuracy and predictivity of the current model based on the cross-validation statistics and examples of forward predictivity demonstrates its potential for use in an integrated evaluation strategy for environmental chemicals. Additionally, the model has shown to be specific to reproductive toxicity and is not modeling general systemic toxicity as is evident with the lack of concordance with the systemic parental and offspring LOAEL. It should be noted that the ToxCast assay data is concurrently being used to develop independent predictive models of cancer, systemic and developmental toxicities. Once further model performance assessment has been performed on models developed using ToxCast data, the models could be combined into an integrated testing strategy. As a starting point in this process, the current reproductive toxicity model underwent performance-based assessment demonstrating its strengths and limitations. For example, chemicals that require metabolic activation such as Benomyl or Molinate will not be predicted as a reproductive toxicant by this model, at least not until HTS data using metabolically competent systems are available (Jewell *et al.*, 1998; Lim and Miller, 1997). Additionally, chemicals such as Boric Acid that likely causes its male reproductive toxicity through non-molecular interactions demonstrate limitations of the current model (Chapin and Ku, 1994; Jewell, et al., 1998) and points to the larger issue of chemical domain of applicability. The ToxCast Phase I chemical set contains a large number of conventional pesticides. The ToxCast Phase II chemical library contains about 700 chemicals with more diverse structural and use characteristics, including on-the-market and failed pharmaceuticals, food additives, antimicrobials, and other industrial chemicals. ToxCast Phase II will provide a robust external validation set testing the

forward predictivity of the current model, and evaluating the model's chemical domain of applicability. An advantage of developing predictive models using quantative HTS data linked to genes, proteins and pathways is the ability to identify gaps in the mechanisms covered by the model. There are also instances of chemicals predicted to be reproductive toxicants that caused minimal reproductive toxicity in the multigenerational study but have been shown to cause reproductive related effects in either chronic, developmental or other types of studies. Examples of reproductive related effects for Triclosan and Bensulide (Foran *et al.*, 2000; Zorrilla *et al.*, 2009) from other study types demonstrate the difficulty in definitively calling training set chemicals 'positive' or 'negative' for reproductive toxicity.

Among the chemicals selected for external validation, the model provided accurate predictions for 16 of the 21 chemicals. The five chemicals with inaccurate predictions provide valuable insight into potential limitations or gaps of the model. Interestingly, the 5 chemicals had a common phenotypic profile with respect to reproductive toxicity. Tribufos, Spiroxamine, Tefluthris, Disulfoton, and Esfenvalerate all caused reduced early offspring survival, particularly litter size decrease with little to no accompanying effects on reproductive performance or reproductive tract pathology. The rLOAEL for all 5 chemicals was set at the high dose tested based on the early offspring survival effects and the parental and offspring LOAEL were set at the lower dose levels. Based on the inclusive definition used for defining a positive for reproductive toxicity for model development all 5 were considered positive, but lack evidence of specific fertility-related or developmentally sensitive reproductive outcomes. Nonetheless, a gap in model predictivity has been identified and could potentially be filled using additional assay technologies, physical chemical properties and structural descriptors, or acute or short-term *in vivo* studies.

The model development process identified biologically plausible features and pathways from over 500 assays mapped to 100s of genes and spanning many reproductive relevant modes-of-action. PPARα activity was clearly associated with reproductive toxicity, with all 10 PPARα agonists in the training set (chemical group A) causing reproductive toxicity. Putative PPARα agonists (Lactofen (Butler *et al.*, 1988), Imazalil (Takeuchi *et al.*, 2006), Diclofop-methyl (Takeuchi, et al., 2006), DEHP (Klaunig *et al.*, 2003), MEHP (Klaunig, et al., 2003), and PFOA (Klaunig, et al., 2003)) and environmental chemicals identified as potential PPARα agonists through the ToxCast research program (Fluazinam, Emamectin benzoate, Vinclozolin, and Fenthion) span many chemical classes yet share a relatively common reproductive toxicity profile; a decrease in reproductive performance (i.e., decreased fertility) in 8 out of the 10 chemicals. Although a mechanistic link between PPAR activity and fertility or other reproductive impairments remains unclear (Peraza *et al.*, 2006), the role of PPAR in steroid metabolism and its activity in reproductive tissues infers that it is a plausible target for disruption of endocrine signaling and altered gametogenesis.

AR and ERα activity was also associated with reproductive toxicity. The ToxCast receptor profiling identified most if not all the known anti-androgenic and estrogenic chemicals in the current dataset, including well studied chemicals such as Vinclozolin, Bisphenol A, Methoxychlor, HPTE and Clorophene. The role of potency in determining a chemical's relative reproductive toxicity potential needs to be explored further, considering 5 of the top 7 scoring ERα activators (i.e., active across multiple ERα assays and at relatively low concentrations) did not cause substantial reproductive toxicity *in vivo*, including Flumetralin, Fenhexamid, Fludioxonil, Pyridaben, and Endosulfan. Additionally, the impact

of weak or partial nuclear receptor agonists and antagonists on reproductive toxicity potential and other toxicities needs to be explored further.

CYP enzyme inhibition, as compared to gene induction, was significantly more associated with reproductive toxicity. Alterations in steroid metabolism through CYP induction have been previously associated with reproductive impairment (Goetz *et al.*, 2007), however the non-specific inhibition of CYPs may be a surrogate for a chemical's capacity to disturb steroid metabolism including inhibition of key CYPs involved in steroidogenesis (e.g., Cyp19 and Cyp17). Related to CYP activity, PXR interestingly displayed a negative correlation/association with reproductive toxicity. In general, PXR lowered the false positive rate of the model by lowering the model score of chemicals with non-specific and low potency nuclear receptor activity. Robust PXR activity is an indication of potent xeno-sensing and potentially rapid metabolism.

The pyrethroid class of pesticides has shown limited reproductive toxicity in guideline toxicity studies, although there is limited evidence linking pyrethroid exposure to decreased human sperm quality (Meeker *et al.*, 2008). Of the 10 pyrethroids in chemical group A, only Resmethrin was considered a reproductive toxicant based on the criteria described in this manuscript. All 10 pyrethroids displayed low potency activity across one or more of the selected features, including AR, ERα, and PPARγ, but not CYP. Without the down-weighting based on each of their PXR activities the pyrethroids would have all been predicted to be reproductive toxicants.

A major component of the model not directly related to nuclear receptor biology and xenobiotic/steroid metabolism was GPCR binding. Numerous GPCR binding assays were significantly associated with reproductive toxicity. Those chosen to represent the GPCR

family were selected for statistical, and not biological, reasons as there is limited literature information on their role in reproduction in contrast to their well characterized role in nervous system function.

Platforms measuring EGFR, TGF-β1 and NF-κB activity were also associated with reproductive toxicity and make up the "OTHER" feature. All three gene products have been shown to modulate the relative sensitivity of developmental toxicants, especially aryl-hydrocarbon receptor signaling (Abbott *et al.*, 2003; Tian *et al.*, 1999) and may be indicative of altered xenobiotic metabolism, cellular proliferation, cell-cell signaling or potential epigenetic effects (Tian, 2009; Tian *et al.*, 2002). Overall the key targets in the model identify plausible modes of action leading to reproductive toxicity covering anti-androgenic, estrogenic, cholesterol/steroid metabolism, limited coverage of disruption of steroidogenesis, and altered xenobiotic metabolism modes of action.

There have been limited efforts toward the development of models predictive of reproductive toxicity, due in part to the lack of reference data in which to model. One resource for predictive models have come from structure based methods (i.e., Quantitative Structure Activity Relationship (QSAR) models) and the accuracy and predictivity of the resultant models has been limited. A comprehensive effort toward the prediction of reproductive and developmental toxicity was undertaken by the FDA (Matthews *et al.*, 2007). The resultant QSAR models were developed for endpoints such as sperm effects, female reproductive toxicity and male reproductive toxicity and were generally highly specific models with an average specificity across all generated models being 88%. However, the average balanced accuracy across all models was 58%, with the maximum balanced accuracy for any single model being 68% for predictive trans-species female reproductive

toxicity. It is difficult to assess the true accuracy and forward predictivity of the models based solely on the summary statistics, but the balanced accuracy values provide the most direct and unbiased comparison to the current model. Most likely, the limitation lies in the physiological complexity of reproductive toxicity and structural diversity of reproductive toxicants. The current predictive model has improved accuracy over any published QSAR model of reproductive toxicity and provides additional mechanistic information and indications of specific reproductive effects. The model also can be extended to include new data either covering the gene targets in the current model or new gene targets of other potential reproductive toxicity modes-of-action. Additional international efforts are underway with the goal of using alternative testing approaches in the detection of reproductive toxicants and on limited chemical sets have shown promise (Schenk *et al.*, 2010). However, the current ToxCast-based approach utilizes hundreds of diverse biological-chemical activities associated with many potential modes of action leading to reproductive toxicity. The output of the current model provides a binary classification. Applications beyond hazard identification and testing prioritization may require dose response and even mechanistic information. To accomplish this, research is underway incorporating toxicokinetic information into the modeling process using primary rodent and human hepatocytes, plasma protein binding, and pharmacokinetic modeling intended to reverse engineer expected oral dose required to achieve a particular *in vitro* bioactivity level (Rotroff *et al.*, 2010). Experimentally and computationally deriving dosimetry relevant to *in vivo exposures* has the potential to provide quantitative dose response information that can be incorporated into the modeling process. For example, the *in vitro* constitutive androstane receptor (CAR) and PXR activity on a set of conazole fungicides in ToxCast Phase I

demonstrated the dose response relationship between the equivalent *in vivo* levels required to observe the CAR/PXR activity and the known dose levels causing rodent liver toxicity (Judson *et al.*, 2011). Examples such as this provide a path toward incorporating *in vitro* assay data into the risk assessment process, but also demonstrate the amount of prior knowledge currently required to perform such an analysis. The vast majority of environmental chemicals have little to no prior toxicity data and those that do commonly lack information on potential modes-of-action or human relevance. The reality is that among the thousands of environmental chemicals few will ever have a multigeneration reproductive study run. Over the past 30 years, only, 500 chemicals have been run in multigeneration reproductive studies due to the high animal and financial burden for such large scale animal testing (Hartung and Rovida, 2009). A practical solution and pressing need, especially with regards to reproductive toxicity testing, is for prioritization tools such as the current model, to make more informed reproductive toxicity testing decisions.

Cross-validation and external validation sets used to develop and assess the quality of the reproductive toxicity model helped identify strengths and weaknesses of the present model and will help focus future research. Using HTS assays as the input into the model provided mechanistic insights and helped further characterize the predicted chemicals beyond negative and positive prediction outcomes. However, a subset of chemicals were deemed outside the domain of applicability based on low in vitro activity due to physical chemical characteristics, biological gaps, chemical decomposition or volatility. Further research is needed to better characterize this chemical subset to expand the current model's domain of applicability. A limited number of chemicals selected from the ToxCast Phase I chemical set were used for external validation and provided supporting evidence of the model quality. A

large set of chemicals from ToxCast Phase II will have a full complement of in vitro bioactivity data and rodent reproductive toxicity studies will be used to further evaluate, validate and expand the predictive model. Additionally, ToxCast Phase II contains a library of failed pharmaceuticals with pre-clinical and clinical toxicity outcomes as well as reference chemicals with known mechanisms of reproductive toxicity. In addition to diversifying the current chemical library these chemicals will aid in the expansion of predictive reproductive toxicity model development toward mechanistic and human reproductive toxicity models useful in risk assessment applications.

The ability of this predictive reproductive toxicity model to externally predict numerous chemicals with biological and structural diversity demonstrates suitability for chemical testing prioritization. Although the model does not provide quantitative dose response information, it does provide accurate predictions of a chemical's reproductive toxicity potential. Since the model is based on HTS data, it is amenable to screening and prioritizing thousands of chemicals. Additionally, the biological features of the model provide mechanistic insights into modes of action useful in developing an integrated testing strategy for reproductive toxicity.

Tables

Table 4.1. Ninety-eight chemicals, linked to 86 acceptable and 12 unacceptable studies in ToxRefDB, achieved a reproductive LOAEL (rLOAEL $\leq$ 500 mg/kg/day) and used as the positive class set for the training and testing of the predictive model. In enforcing the 500/mg/day cutoff, rLOAEL were rounded to one significant figure due to the uncertainty of dose intake especially at high doses.

CASRN	*Chemical Name*	*rLOAEL (mg/kg /day)*	*Acceptable Study?*

71751-41-2	Abamectin	0.4	NO
30560-19-1	Acephate	25	YES
135410-20-7	Acetamiprid	51	YES
33089-61-1	Amitraz	12	YES
3337-71-1	Asulam	250	YES
35575-96-3	Azamethiphos	50	YES
1861-40-1	Benfluralin	401	YES
17804-35-2	Benomyl	234	YES
80-05-7	Bisphenol A	500	YES
134605-64-4	Butafenacil	23.8	YES
75-60-5	Cacodylic acid	17.9	YES
63-25-2	Carbaryl	92.4	YES
5234-68-4	Carboxin	20	YES
101-21-3	Chlorpropham	150	YES
64902-72-3	Chlorsulfuron	541	NO
210880-92-5	Clothianidin	31.2	YES
1134-23-2	Cycloate	50	YES
94-75-7	2,4-D	80	YES
94-82-6	2,4-DB	112	YES
1596-84-5	Daminozide	500	YES
117-81-7	DEHP	391	YES
333-41-5	Diazinon	35.2	YES
962-58-3	Diazoxon	35.2	YES
84-74-2	Dibutyl phthalate	531	YES
1918-00-9	Dicamba	419	YES
99-30-9	Dichloran	102	YES
120-36-5	Dichlorprop	220	YES
62-73-7	Dichlorvos	7.2	YES
51338-27-3	Diclofop-methyl	7.3	YES
115-32-2	Dicofol	2.4	YES
141-66-2	Dicrotophos	0.56	YES
60-51-5	Dimethoate	6.5	YES
122-39-4	Diphenylamine	399	YES
298-04-4	Disulfoton	0.12	NO
155569-91-8	Emamectin benzoate	1.8	YES
66230-04-4	Esfenvalerate	6.7	NO
60168-88-9	Fenarimol	1.2	YES
114369-43-6	Fenbuconazole	40	YES
122-14-5	Fenitrothion	0.68	NO
55-38-9	Fenthion	0.7	YES
76-87-9	Fentin	1.4	YES
120068-37-3	Fipronil	26.3	YES
69806-50-4	Fluazifop-butyl	5.8	YES
79241-46-6	Fluazifop-P-	5.8	YES

	butyl		
79622-59-6	Fluazinam	47.3	YES
103361-09-7	Flumioxazin	12.7	YES
85509-19-9	Flusilazole	17.5	YES
133-07-3	Folpet	180	YES
68157-60-8	Forchlorfenuron	544	YES
79983-71-4	Hexaconazole	50	YES
35554-44-0	Imazalil	80	YES
144550-36-7	Iodosulfuron-methyl-na	346	YES
55406-53-6	IPBC	37.5	YES
77501-63-4	Lactofen	26.2	YES
330-55-2	Linuron	54	YES
12427-38-2	Maneb	106	YES
94-74-6	MCPA	22.5	YES
4376-20-9	MEHP	391	YES
104206-82-8	Mesotrione	1.1	YES
950-37-8	Methidathion	1.25	YES
9006-42-2	Metiram-zinc	16	NO
7786-34-7	Mevinphos	0.5	YES
51596-11-3	Milbemectin	65.6	NO
2212-67-1	Molinate	0.8	YES
131-70-4	Monobutyl phthalate	531	YES
88671-89-0	Myclobutanil	50	YES
300-76-5	Naled	18	YES
27314-13-2	Norflurazon	103	YES
116714-46-6	Novaluron	298	YES
42874-03-3	Oxyfluorfen	146	YES
40487-42-1	Pendimethalin	215	YES
335-67-1	PFOA	30	YES
1763-23-1	PFOS	3.2	YES
2310-17-0	Phosalone	29.4	YES
86209-51-0	Primisulfuron-methyl	250	YES
67747-09-5	Prochloraz	31.3	NO
709-98-8	Propanil	53	YES
31218-83-4	Propetamphos	5.5	YES
60207-90-1	Propiconazole	238	YES
23950-58-5	Propyzamide	123	YES
10453-86-8	Resmethrin	70.8	YES
83-79-4	Rotenone	7	YES
148477-71-8	Spirodiclofen	178	YES
118134-30-8	Spiroxamine	44.8	NO
122836-35-5	Sulfentrazone	33	YES
119168-77-3	Tebufenpyrad	19.3	YES
96182-53-5	Tebupirimfos	1.25	NO

79538-32-2	Tefluthrin	12.5	NO
112281-77-3	Tetraconazole	6	YES
153719-23-4	Thiamethoxam	1.84	YES
43121-43-3	Triadimefon	90	YES
55219-65-3	Triadimenol	25	YES
2303-17-5	Tri-allate	30	YES
78-48-8	Tribufos	15	NO
52-68-6	Trichlorfon	175	YES
68694-11-1	Triflumizole	1.5	YES
131983-72-7	Triticonazole	250	YES
50471-44-8	Vinclozolin	4.9	YES

Table 4.2. Chemical groupings based on aggregate *in vitro* activity across the over 500 ToxCast assays, and *in vivo* reproductive study acceptability/availability within ToxRefDB.

	In vitro Activity	Little to No *In vitro* Activity (<2% Active)	Total *In vivo* Chemical Counts
Acceptable Reproductive Study	206 (A)	50 (B)	256
Unacceptable Reproductive Study	31 (C)	8 (D)	39
No Reproductive Study Available	10 (E)	4 (F)	14
Total *In vitro* Chemical Counts	247	62	309

Parenthesis: Each chemical group identified by letter

Table 4.3. Feature selection statistics based on univariate correlations and associations between individual assays or genes/gene-sets and reproductive toxicants in chemical group A, dichotomously represented (i.e., 1 for positive and 0 for negative).

Individual Assay	Correlation (p-value)	Gene / Gene Set	Correlation (p-value)
ATG_PPARα_TRANS	0.24 (6.6E-4)	PPARα	0.30 (9.6E-6)
NCGC_PPARα_Agonist	0.17 (1.7E-2)		
NVS_NR_hPPARα	0.17 (1.6E-2)		
NCGC_AR_Antagonist	0.18 (3.6E-3)	AR	0.31 (4.8E-6)
NVS_NR_hAR	0.28 (3.7E-5)		
NVS_NR_rAR	0.04 (NS)		
ATG_ERα_TRANS	0.17 (1.5E-2)	ERα	0.15 (2.3E-2)

ATG_ERE_CIS	0.04 (NS)		
NCGC_ERalpha_Agonist	0.05 (NS)		
NCGC_ERalpha_Antagonist	0.11 (9.3E-2)		
NVS_NR_hER	0.10 (NS)		
NVS_NR_mERα	0.10 (NS)		
NVS_NR_bER	0.10 (NS)		
ATG_PPRE_CIS	0.10 (NS)	PPARγ	0.14 (4.1E-2)
ATG_PPARγ_TRANS	0.09 (NS)		
NCGC_PPARγ_Agonist	0.06 (NS)		
NVS_NR_hPPARγ	0.14 (4.1E-2)		
NVS_ADME_rCYP2A2	0.30 (1.5E-5)	CYP	0.27 (1.1E-4)
NVS_ADME_rCYP2B1	0.23 (7.2E-4)		
NVS_ADME_rCYP2C12	0.14 (4.9E-2)		
NVS_ADME_rCYP2C11	0.17 (1.5E-2)		
NVS_ADME_rCYP2A1	0.21 (2.4E-3)		
NVS_ADME_rCYP2C13	0.21 (2.3E-3)		
NVS_ADME_hCYP19A1	0.17 (1.3E-2)		
NVS_GPCR_hOpiate_mu	0.26 (1.8E-4)	GPCR	0.34 (8.6E-7)
NVS_GPCR_h5HT6	0.22 (1.3E-3)		
NVS_GPCR_hAdra2C	0.21 (2.0E-3)		
NVS_GPCR_hPY2	0.20 (3.3E-3)		
NVS_GPCR_gOpiateK	0.19 (7.3E-3)		
BSK_hDFCGF_EGFR_up	0.14 (4.5E-2)	OTHER	0.28 (4.5E-5)
BSK_BE3C_TGFb1_up	0.08 (NS)		
NVS_TR_rVMAT2	0.21 (2.7E-3)		
ATG_NF_kB_CIS	0.14 (4.9E-2)		
ATG_PXR_TRANS	-0.14 (4.2E-2)	PXR	-0.14 (4.5E-2)
ATG_PXRE_CIS	-0.11 (NS)		
NCGC_PXR_Agonist_human	-0.09 (NS)		

(NS) – Not statistically significant (p-value>0.1)

Table 4.4. Performance metrics for the predictive model of reproductive toxicity, including cross-validation and optimized model statistics and weighting of model input features.

Cross-Validation Statistics		Full Model Statistics				Parameter Coefficients	
Learner	LDA	TP	55	F1	73%	PPARα	1.37
CV	5-fold	FP	28	RR	6.3	AR	0.98

No. F	8	FN	13	OR	17	ERα	0.45
Assays	36	TN	110	PPV	66%	PPARγ	0.23
BA Train	77%	SENS	81%	NPV	90%	CYP	0.28
SD Train	2%	SPEC	80%	Pred	78%	GPCR	0.5
BA Test	74%	BA	80%	P-Value	4.2E-17	OTHER	0.45
SD Test	5%	A	80%	Cutoff	0.6	PXR	-0.21

CV = Cross Validation; No. F = Number of selected features; Assays=Number of assays comprising the selected features; BA = Balanced accuracy (Average of sensitivity and specificity); SD= Standard deviation of the Balanced Accuracy for each Fold; TP= True Positive Count; FP= False Positive Count; FN=False Negative Count; TN= True negative count; SENS= Sensitivity; SPEC= Specificity; A = Accuracy; P-Value = Chi-Square P-Value; Cutoff = LDA Model Intercept; F1=F-measure (harmonic mean of precision and recall); OR= Odds ratio; PPV=Positive predictive value; NPV=Negative predictive value; Pred=Predictivity (Average of PPV and NPV)

Table 4.5. External validation chemical set used to test the forward predictivity of the model. Each chemical is associated with a chemical group based on reproductive study acceptability/availability and aggregate *in vitro* activity. Chemical group C showed evidence of reproductive toxicity based on positive findings in an unacceptable multigeneration study, while chemical groups E and F showed literature evidence of reproductive toxicity.

Chemical Group	CASRN	Chemical Name	Evidence of Reproductive Toxicity	Predicted Reproductive Toxicant	Model Score
E	2971-36-0	HPTE	Yes	Yes	11.9
C	122-14-5	Fenitrothion	Yes	Yes	5.5
C	67747-09-5	Prochloraz	Yes	Yes	3.4
E	1689-84-5	Bromoxynil	Yes	Yes	3.4
E	72-43-5	Methoxychlor	Yes	Yes	2.7
C	51596-11-3	Milbemectin	Yes	Yes	2.4
C	9006-42-2	Metiram-zinc	Yes	Yes	1.6
C	64902-72-3	Chlorsulfuron	Yes	Yes	0.8
F	109-86-4	Methyl cellusolve	Yes	Yes	0.8
C	71751-41-2	Abamectin	Yes	Yes	0.7
C	96182-53-5	Tebupirimfos	Yes	Yes	0.7
E	15972-60-8	Alachlor	Yes	Yes	0.7
C	78-48-8	Tribufos	Yes	No	0.0
C	118134-30-8	Spiroxamine	Yes	No	-0.1
C	79538-32-2	Tefluthrin	Yes	No	-0.1
C	298-04-4	Disulfoton	Yes	No	-0.2
C	66230-04-4	Esfenvalerate	Yes	No	-0.5
E	4376-18-5	Methyl hydrogen	No	No	0.0

		phthalate			
F	6923-22-4	Monocrotophos	No	No	0.0
F	131-11-3	Dimethyl phthalate	No	No	-0.2
E	33629-47-9	Butralin	No	No	-0.2
E	120-32-1	Clorophene	Unknown	Yes	4.4
E	83657-24-3	Diniconazole	Unknown	Yes	2.4
E	50-65-7	Niclosamide	Unknown	Yes	1.6
F	122-99-6	Phenoxyethanol	Unknown	No	0.0
E	87-90-1	Symclosene	Unknown	No	-0.2

Table 4.6. Comparison of predictive model results to classification and labeling for reproductive toxicity. Repro C&L is the European Union classification and labeling for reproductive toxicity, with R60 and 62 referring to fertility impairment, and R61 and 63 to developmental toxicity.

Chemical Name	Predicted Positive	Repro C&L	Model Score
Bisphenol A	Yes	R62	6.1
Vinclozolin	Yes	R60&61	4.7
Flusilazole	Yes	R61	4.6
Linuron	Yes	R62&61	2.9
Myclobutanil	Yes	R63	2.4
Fenarimol	Yes	R62	2.5
Fentin	Yes	R63	3.5
Fluazifop-P-butyl	Yes	R63	1.7
Flumioxazin	Yes	R61	0.9
Cyproconazole	Yes	R63	1.2
Diethylhexyl phthalate (DEHP)	Yes	R60&61	0.9
Isoxaflutole	Yes	R63	0.6
Fluazifop-butyl	Yes	R61	1.0
Dibutyl phthalate	Yes	R62&61	0.8
Benomyl	No	R60&61	0.0
Diuron	No	---	0.4
Lindane	No	---	0.0
Propazine	No	---	-0.3
Propargite	No	---	-0.5

Figures

Figure 4.1. Decision tree diagram representing the process by which the 309 ToxCast chemicals were grouped, based on *in vivo* study acceptability/availability and *in vitro* aggregate bioactivity, and subsequently defined as positive or negative for reproductive toxicity based on having achieved a reproductive lowest observed adverse effect level (rLOAEL) less than 500 mg/kg/day. Applying the model developed using chemical group A to all other chemical groups, individual and combined balanced accuracy values (average of sensitivity and specificity) are summarizing the results across the entire chemical library and provides context and summary information as each chemical group is discussed in greater detail throughout the results and discussion sections.

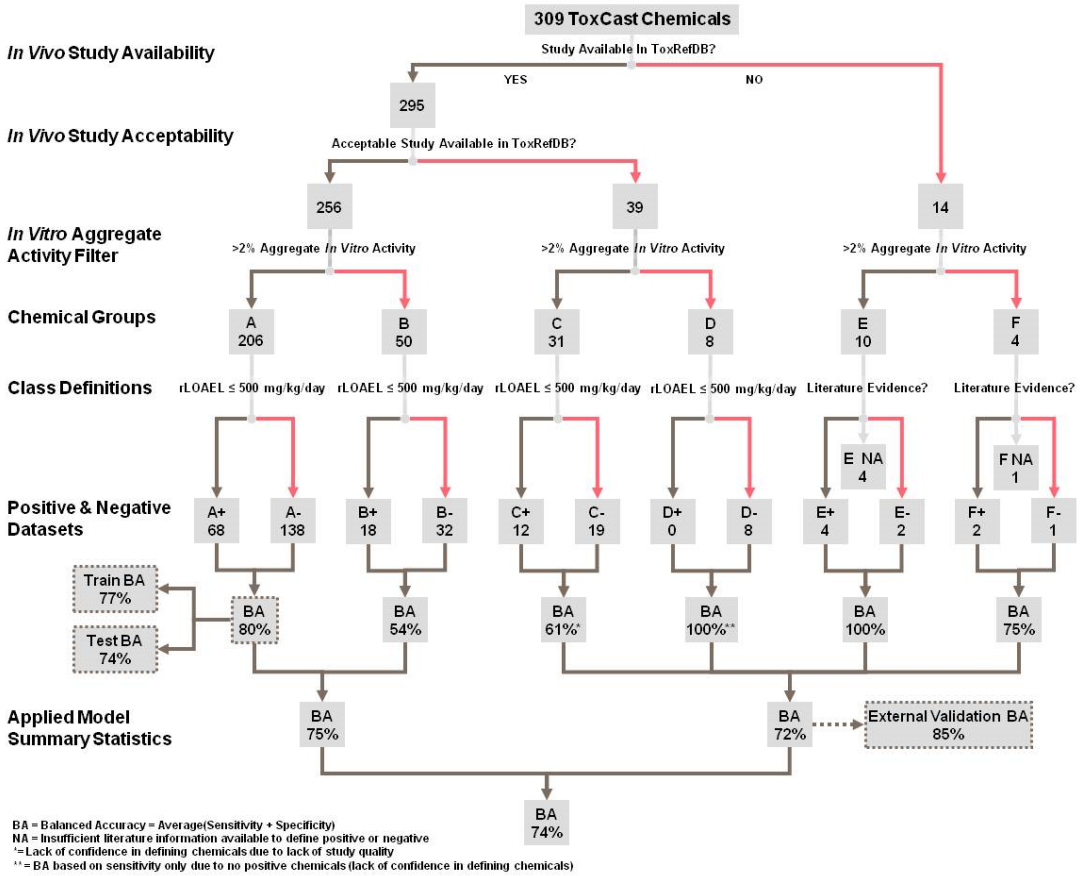


Figure 4.2. Classification or predictivity rates increase from individual HTS assays, to aggregated genes or gene-sets, while misclassification rates proportionately decrease, demonstrating the advantage of combining assays for same genes in model development. Classification rate (blue line) was calculated as the proportion of true positives (greater than the mean assay/gene/gene-set/model and positive for *in vivo* reproductive toxicity) over the total number positives (n=68). The misclassification rate (red line) was calculated as the proportion of false positives (greater than the mean assay/gene/gene-set/model score but negative for reproductive toxicity) over the total number of chemicals negative for reproductive toxicity (n=138).

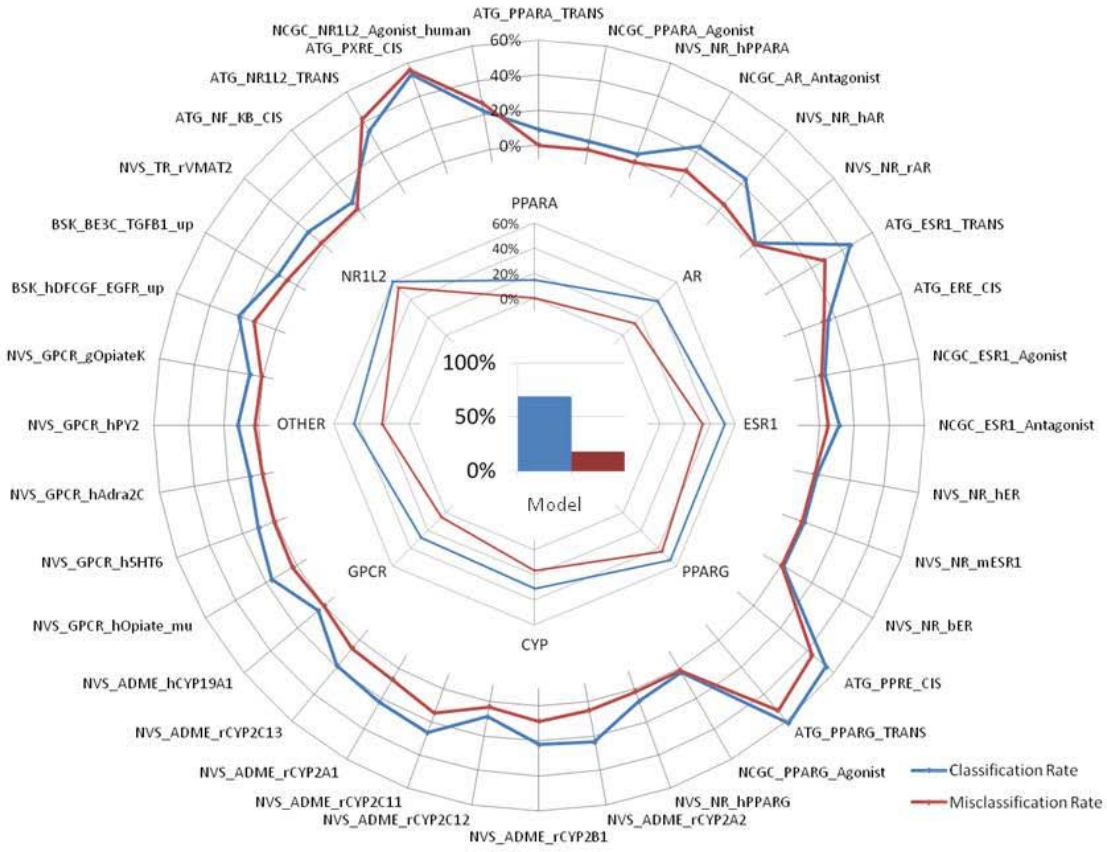
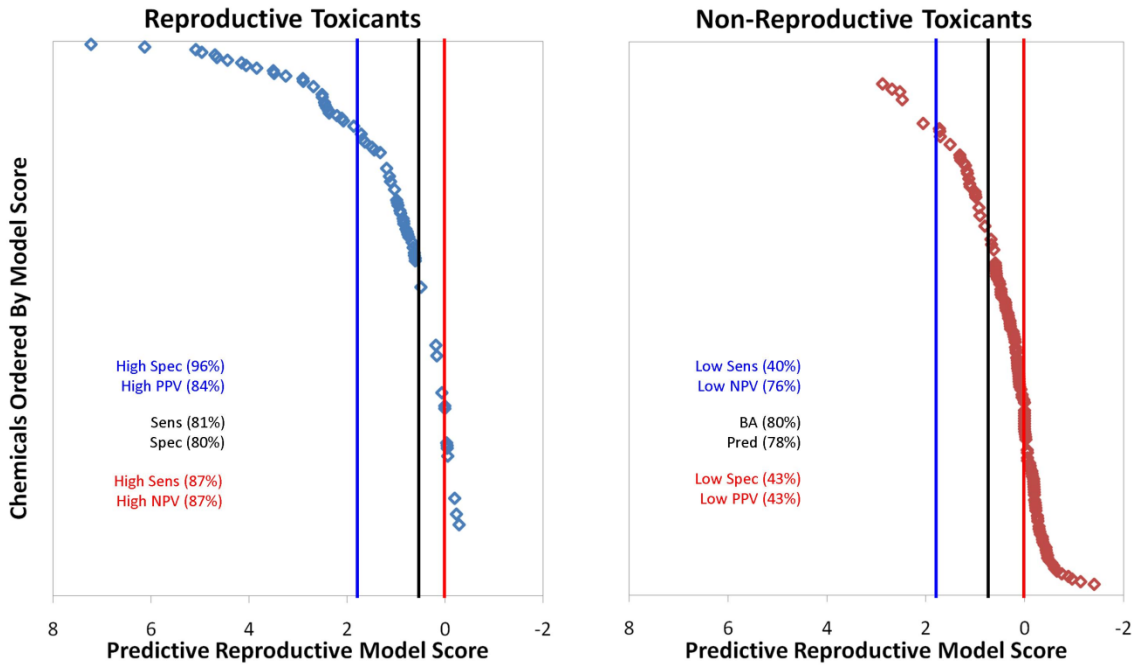


Figure 4.3. Chemicals ordered by their reproductive toxicity model score with the positive training set on the left and the negative training set on the right. The optimal cutoff was determined to be 0.6 (black line) and achieved a balanced accuracy of 80%. Depending on the prioritization goals, an increased or decreased cutoff could greatly alter your confidence in detecting a reproductive toxicant. Using an increased cutoff, one could test the top 32 scoring chemicals and expect to have 27 be reproductive toxicants (cutoff of 1.8 in blue). On the other hand, to accurately predict 59 of the 68 total reproductive toxicants one would have to test the top scoring 137 of 206 chemicals (cutoff of >0 in red).



References

Abbott, B. D., Buckalew, A. R., DeVito, M. J., Ross, D., Bryant, P. L. and Schmid, J. E. (2003). EGF and TGF-alpha expression influence the developmental toxicity of TCDD: Dose response and AhR phenotype in EGF, TGF-alpha, and EGF plus TGF-alpha knockout mice. *Toxicol Sci* **71**, 84-95.

Akingbemi, B. T., Ge, R. S., Klinefelter, G. R., Gunsalus, G. L. and Hardy, M. P. (2000). A metabolite of methoxychlor, 2,2-bis(p-hydroxyphenyl)-1,1,1-trichloroethane, reduces testosterone biosynthesis in rat leydig cells through suppression of steady-state messenger ribonucleic acid levels of the cholesterol side-chain cleavage enzyme. *Biol Reprod* **62**, 571-578.

Armenti, A. E., Zama, A. M., Passantino, L. and Uzumcu, M. (2008). Developmental methoxychlor exposure affects multiple reproductive parameters and ovarian folliculogenesis and gene expression in adult rats. *Toxicol Appl Pharm* **233**, 286-296, DOI 10.1016/j.taap.2008.09.010.

Butler, E. G., Tanaka, T., Ichida, T., Maruyama, H., Leber, A. P. and Williams, G. M. (1988). Induction of Hepatic Peroxisome Proliferation in Mice by Lactofen, a Diphenyl Ether Herbicide. *Toxicol Appl Pharm* **93**, 72-80.

Canada, H. (1999). Priority Substances List Assessment Report for 2-Methoxyethanol.http://www.hc-sc.gc.ca/ewh-semt/pubs/contaminants/psl2-lsp2/2_methoxyethanol/2_methoxyethanol_2-eng.php

Chapin, R. E., Harris, M. W., Davis, B. J., Ward, S. M., Wilson, R. E., Mauney, M. A., Lockhart, A. C., Smialowicz, R. J., Moser, V. C., Burka, L. T., Collins, B. J., Haskins, E. A., Allen, J. D., Judd, L., Purdie, W. A., Harris, H. L., Lee, C. A. and Corniffe, G. M. (1997). The effects of perinatal/juvenile methoxychlor exposure on adult rat nervous, immune, and reproductive system function. *Fund Appl Toxicol* **40**, 138-157.

Chapin, R. E. and Ku, W. W. (1994). The Reproductive Toxicity of Boric-Acid. *Environ Health Persp* **102**, 87-91.

Cronin, M. T. D. and Worth, A. P. (2008). (Q)SARs for predicting effects relating to reproductive toxicity. *Qsar Comb Sci* **27**, 91-100, DOI 10.1002/qsar.200710118.

Dix, D. J., Houck, K. A., Martin, M. T., Richard, A. M., Setzer, R. W. and Kavlock, R. J. (2007). The ToxCast program for prioritizing toxicity testing of environmental chemicals. *Toxicol Sci* **95**, 5-12, kfl103 [pii]
10.1093/toxsci/kfl103.

Foran, C. M., Bennett, E. R. and Benson, W. H. (2000). Developmental evaluation of a potential non-steroidal estrogen: triclosan. *Mar Environ Res* **50**, 153-156.

Goetz, A. K., Ren, H. Z., Schmid, J. E., Blystone, C. R., Thillainadarajah, I., Best, D. S., Nichols, H. P., Strader, L. F., Wolf, D. C., Narotsky, M. G., Rockett, J. C. and Dix, D. J. (2007). Disruption of testosterone homeostasis as a mode of action for the reproductive toxicity of triazole fungicides in the male rat. *Toxicol Sci* **95**, 227-239, DOI 10.1093/toxsci/kfl124.

Gray, L. E., Ostby, J., Furr, J., Price, M., Veeramachaneni, D. N. R. and Parks, L. (2000). Perinatal exposure to the phthalates DEHP, BBP, and DINP, but not DEP, DMP, or DOTP, alters sexual differentiation of the male rat. *Toxicol Sci* **58**, 350-365.

Hareng, L., Pellizzer, C., Bremer, S., Schwarz, M. and Hartung, T. (2005). The Integrated Project ReProTect: A novel approach in reproductive toxicity hazard assessment. *Reprod Toxicol* **20**, 441-452, DOI 10.1016/j.reprotox.2005.04.003.

Hartung, T. and Rovida, C. (2009). Chemical regulators have overreached. *Nature* **460**, 1080-1081, Doi 10.1038/4601080a.

Jewell, W. T., Hess, R. A. and Miller, M. G. (1998). Testicular toxicity of molinate in the rat: Metabolic activation via sulfoxidation. *Toxicol Appl Pharm* **149**, 159-166.

Judson, R. S., Houck, K. A., Kavlock, R. J., Knudsen, T. B., Martin, M. T., Mortensen, H. M., Reif, D. M., Rotroff, D. M., Shah, I., Richard, A. M. and Dix, D. J. (2010). In Vitro Screening of Environmental Chemicals for Targeted Testing Prioritization: The ToxCast Project. *Environ Health Perspect* **118**, 485-92, 10.1289/ehp.0901392.

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**, 451-62, 10.1021/tx100428e.

Klaunig, J. E., Babich, M. A., Baetcke, K. P., Cook, J. C., Corton, J. C., David, R. M., DeLuca, J. G., Lai, D. Y., McKee, R. H., Peters, J. M., Roberts, R. A. and Fenner-Crisp, P. A. (2003). PPAR alpha agonist-induced rodent tumors: Modes of action and human relevance. *Crit Rev Toxicol* **33**, 655-780, Doi 10.1080/10408440390250154.

Kwack, S. J., Kim, K. B., Kim, H. S. and Lee, B. M. (2009). Comparative Toxicological Evaluation of Phthalate Diesters and Metabolites in Sprague-Dawley Male Rats for Risk Assessment. *J Toxicol Env Heal A* **72**, 1446-1454, Doi 10.1080/15287390903212923.

Laier, P., Metzdorff, S. B., Borch, J., Hagen, M. L., Hass, U., Christiansen, S., Axelstad, M., Kledal, T., Dalgaard, M., McKinnell, C., Brokken, L. J. S. and Vinggaard, A. M. (2006). Mechanisms of action underlying the antiandrogenic effects of the fungicide prochloraz. *Toxicol Appl Pharm* **213**, 160-171, DOI 10.1016/j.taap.2005.10.013.

Lim, J. H. and Miller, M. G. (1997). The role of the benomyl metabolite carbendazim in benomyl-induced testicular toxicity. *Toxicol Appl Pharm* **142**, 401-410.

Martin, M. T., Judson, R. S., Reif, D. M., Kavlock, R. J. and Dix, D. J. (2009a). Profiling Chemicals Based on Chronic Toxicity Results from the US EPA ToxRef Database. *Environ Health Persp* **117**, 392-399, Doi 10.1289/Ehp.0800074.

Martin, M. T., Mendez, E., Corum, D. G., Judson, R. S., Kavlock, R. J., Rotroff, D. M. and Dix, D. J. (2009b). Profiling the Reproductive Toxicity of Chemicals from Multigeneration Studies in the Toxicity Reference Database. *Toxicol Sci* **110**, 181-190, DOI 10.1093/toxsci/kfp080.

Matthews, E. J., Kruhlak, N. L., Daniel Benz, R., Ivanov, J., Klopman, G. and Contrera, J. F. (2007). A comprehensive model for reproductive and developmental toxicity hazard identification: II. Construction of QSAR models to predict activities of untested chemicals. *Regul Toxicol Pharmacol* **47**, 136-55, S0273-2300(06)00174-7 [pii] 10.1016/j.yrtph.2006.10.001.

Meeker, J. D., Barr, D. B. and Hauser, R. (2008). Human semen quality and sperm DNA damage in relation to urinary metabolites of pyrethroid insecticides. *Hum Reprod* **23**, 1932-1940, DOI 10.1093/humrep/den242.

Okahashi, N., Sano, M., Miyata, K., Tamano, S., Higuchi, H., Kamita, Y. and Seki, T. (2005). Lack of evidence for endocrine disrupting effects in rats exposed to fenitrothion in utero and from weaning to maturation. *Toxicology* **206**, 17-31, DOI 10.1016/j.tox.2004.04.020.

Peraza, M. A., Cromey, D. W., Carolus, B., Carter, D. E. and Gandolfi, A. J. (2006). Morphological and functional alterations in human proximal tubular cell line induced by low level inorganic arsenic: evidence for targeting of mitochondria and initiated apoptosis. *J Appl Toxicol* **26**, 356-367, Doi 10.1002/Jat.1149.

Piersma, A. H., Rorije, E., Beekhuijzen, M. E., Cooper, R., Dix, D. J., Heinrich-Hirsch, B., Martin, M. T., Mendez, E., Muller, A., Paparella, M., Ramsingh, D., Reaves, E., Ridgway, P., Schenk, E., Stachiw, L., Ulbrich, B. and Hakkert, B. C. (2010). Combined retrospective analysis of 498 rat multi-generation reproductive toxicity studies: on the impact of parameters related to F1 mating and F2 offspring. *Reprod Toxicol*, S0890-6238(10)00351-5 [pii]
10.1016/j.reprotox.2010.11.013.

Rao, R. P. and Kaliwal, B. B. (2002). Monocrotophos induced dysfunction on estrous cycle and follicular development in mice. *Ind Health* **40**, 237-244.

Ratnasooriya, W. D., Jayatunga, Y. N. A. and Peiris, L. D. C. (1996). Monocrotophos impairs the fertility of male rats. *Med Sci Res* **24**, 403-406.

Rotroff, D. M., Wetmore, B. A., Dix, D. J., Ferguson, S. S., Clewell, H. J., Houck, K. A., LeCluyse, E. L., Andersen, M. E., Judson, R. S., Smith, C. M., Sochaski, M. A., Kavlock, R.

J., Boellmann, F., Martin, M. T., Reif, D. M., Wambaugh, J. F. and Thomas, R. S. (2010). Incorporating Human Dosimetry and Exposure into High-Throughput In Vitro Toxicity Screening. *Toxicol Sci* **117**, 348-358, DOI 10.1093/toxsci/kfq220.

Schenk, B., Weimer, M., Bremer, S., van der Burg, B., Cortvrindt, R., Freyberger, A., Lazzari, G., Pellizzer, C., Piersma, A., Schafer, W. R., Seiler, A., Witters, H. and Schwarz, M. (2010). The ReProTect Feasibility Study, a novel comprehensive in vitro approach to detect reproductive toxicants. *Reprod Toxicol* **30**, 200-218, DOI 10.1016/j.reprotox.2010.05.012.

Spielmann, H. (2009). The Way Forward in Reproductive/Developmental Toxicity Testing. *Atla-Altern Lab Anim* **37**, 641-656.

Takeuchi, S., Matsuda, T., Kobayashi, S., Takahashi, T. and Kojima, H. (2006). In vitro screening of 200 pesticides for agonistic activity via mouse peroxisome proliferator-activated receptor (PPAR)alpha and PPAR gamma and quantitative analysis of in vivo induction pathway. *Toxicol Appl Pharm* **217**, 235-244, DOI 10.1016/j.taap.2006.08.011.

Tian, Y., Ke, S., Denison, M. S., Rabson, A. B. and Gallo, M. A. (1999). Ah receptor and NF-kappa B interactions, a potential mechanism for dioxin toxicity. *J Biol Chem* **274**, 510-515.

Tian, Y. A. (2009). Ah receptor and NF-kappa B interplay on the stage of epigenome. *Biochem Pharmacol* **77**, 670-680, DOI 10.1016/j.bcp.2008.10.023.

Tian, Y. N., Rabson, A. B. and Gallo, M. A. (2002). Ah receptor and NF-kappa B interactions: mechanisms and physiological implications. *Chem-Biol Interact* **141**, 97-115, Pii S0009-2797(02)00068-6.

U.S. EPA (2008). EPA ToxCast.http://www.epa.gov/ncct/toxcast/

USEPA (1998a). Alachlor: Reregistration Eligibility Decision (RED)

USEPA (1998b). Butralin: Reregistration Eligibiity Decision (RED)

USEPA (2008). DSSTox EPA ToxCast™ High Throughput Screening Testing Chemicals Structure-Index File: SDF File and Documentation.http://www.epa.gov/NCCT/dsstox/sdf_toxcst.html

USEPA (2009). ToxRefDB: Toxicity Reference Database.http://actor.epa.gov/toxrefdb

Uzumcu, M., Kuhn, P. E., Marano, J. E., Armenti, A. E. and Passantino, L. (2006). Early postnatal methoxychlor exposure inhibits folliculogenesis and stimulates anti-Mullerian hormone production in the rat ovary. *J Endocrinol* **191**, 549-558, Doi 10.1677/Joe.1.06592.

Zorrilla, L. M., Gibson, E. K., Jeffay, S. C., Crofton, K. M., Setzer, W. R., Cooper, R. L. and Stoker, T. E. (2009). The Effects of Triclosan on Puberty and Thyroid Hormones in Male Wistar Rats. *Toxicol Sci* **107**, 56-64, DOI 10.1093/toxsci/kfn225.

CHAPTER 5

A PREDICTIVE MODEL OF REPRODUCTIVE TOXICITY: FORWARD VALIDATION AND APPLICATION

Abstract

A predictive model of reproductive toxicity, as observed in rat multigeneration reproductive (MGR) studies, was previously developed using high throughput screening (HTS) data from 36 assays mapped to 8 gene or gene-sets from Phase I of the USEPA ToxCast research program. The model was capable of predicting rodent reproductive toxicity with over 75% balanced accuracy. In a forward validation study, the predictive model was applied to 77 chemicals from ToxCast Phase II with the full battery of HTS and rodent reproductive toxicity data. The model sustained accuracy levels above 70% and demonstrated the ability to maintain predictivity in light of real-world constraints. These constraints included a more diverse chemical library, a new and independent round of HTS testing, increased prevalence of positives in the forward validation chemical set, single point versus concentration response data in a subset of assays, as well as assay attrition and replacement. A second study was performed on 381 chemicals from all phases of model development and validation to test the flexibility and applicability of the model as a tool in chemical testing decision-making. Three case studies were derived representing different statutory contexts that represent real-world regulatory actions, including waiving study requirements, requesting studies, and classifying chemicals. For each case study, the predictive model was optimized to maximally increase testing efficiency in terms of accuracy and cost. In all three

case studies, total cost savings were roughly 15% with false classification rates below 15%. For example, applying the model under case study 1 resulted in the waiving of the MGR study requirement for 67 chemicals based on negative predictions with 8 chemicals being positive, a misclassification rate of 12%. However, the costs savings of not testing 67of 381 chemicals when assessed purely from a testing efficiency viewpoint, i.e., cost of misclassification not accounted, would save close to $40 million dollars. The forward validated predictive model of reproductive toxicity will continue to evolve as new assays become available to fill recognized biological gaps and will be combined with other predictive models, particularly models of developmental toxicity, to form an initial tier to an overarching integrated testing strategy.

Introduction

Predicting toxicity using high-throughput screening and computational approaches will inevitably alter the current toxicity testing paradigms, but the degree and pace to which these technologies and approaches affect toxicity testing will play out over time. Current efforts by the U.S. Environmental Protection Agency's (EPA) National Center for Computational Toxicology in the ToxCast research program and the Tox21 interagency project between National Institutes of Health (NIH), Food and Drug Administration (FDA), and EPA are generating a wealth of biological pathway data on thousands of chemicals across hundreds of biological assays (Collins et al., 2008). To make incoming high throughput screening (HTS) data useful in terms of chemical toxicity evaluation, we have attempted to link *in vitro* screening data with increased risk of causing specific adverse outcomes (Judson et al., 2010). ToxCast Phase I produced and evaluated the largest publically available collection of HTS data on a chemical library. The Phase I chemical

library, of 309 chemicals, spanned many chemical classes but was over-enriched for food-use pesticides that were all designed to be bioactive. Moving into Phase II of ToxCast, the chemical library was ever increasing in diversity in terms of physical chemical property, structural, use, and class diversity. In order to fully evaluate predictive toxicity models from ToxCast Phase I, applying the model to chemicals with independently generated data is required.

Historically, predicting toxicity relied largely on structure and inherent chemical property-based toxicity alerts or on single assay readouts that attempted to mimic whole animal toxicities. Many lessons have been learned from these approaches concerning their limited success either in terms of predictivity or increasing testing efficiency. One such lesson is the need for appropriate modeling best practices, including external and forward validation (Tropsha and Golbraikh, 2007). Any model can be developed given enough data, but the ability to blindly predict a structurally diverse set of chemicals is difficult but necessary to validate any model (Golbraikh et al., 2003; Golbraikh and Tropsha, 2002; Shen et al., 2004; Tropsha and Golbraikh, 2007). Another lesson is that end-stage toxicity (e.g., cancer, malformations, infertility) is a complex manifestation of many events that cannot usually be modeled or predicted using a single property, feature, or assay. With the large amount of bioactivity data now available on hundreds of toxicologically well-characterized chemicals it is easy to ignore these lessons and throw hundreds of assays at a single endpoint and to expect robust models to emerge. This, not surprisingly, is not the case. Developing a predictive model of toxicity requires careful consideration in the methods used to analyze the screening data, annotate assays, identify training and test set chemicals, and model the data without over-fitting.

Developing a model for reproductive toxicity introduces additional considerations including life-stage and generational sensitivities, molecular versus physiological modes of action, and confounding systemic toxicities (Cronin and Worth, 2008; Perreault, 2008). Using a suite of HTS bioactivity assays we were able to produce a stable (i.e., test set accuracy commensurate with training set accuracy) and robust (i.e., able to accurately predict a limited external validation set) model of rat reproductive toxicity (Martin et al., 2011). The model accurately predicts effects on adult reproductive performance, male and female reproductive tract effects, early offspring survival, and sexual developmental landmarks. The model does not cover developmental malformations or systemic parental or offspring toxicities. Providing mechanistic insight may be one of the biggest advantages of HTS-derived models. In particular, this model of reproductive toxicity hones in on 8 genes or gene-sets that begin to characterize a chemical's potential mode of action. A chemical with strong estrogen or androgen receptor activity may be leading to reproductive impairment through endocrine disrupting pathways (Witorsch and Thomas, 2010), whereas a chemical with strong cytochrome P450 inhibition or peroxisome proliferator-activated receptor activity may cause reproductive impairment through xenobiotic or steroid metabolism alterations (Toda et al., 2003). Identifying mechanisms of action also enables an evaluation of which mechanisms were not covered and for which new assays need to be developed. For example, steroidogenesis is recognized as a mode of reproductive toxicity for which we have only a single assay, a CYP19A1 inhibition assay, so clearly other assays need to be added to test for activity against this pathway in a high throughput manner (Hecker and Giesy, 2008). In summary, the model features were PPARA, PPARG, ESR1, AR, GPCR (collection of GPCR binding assays), CYP (collection of cytochrome P450 inhibition assays), PXR and OTHER

(collection of significantly associated assays not with limited to no redundant gene coverage). The weight of each feature was determined in the modeling process using linear discriminant analysis (LDA) and a model cutoff was derived to separate the positive and negatives based on an optimal balanced accuracy.

The current study aims to forward validate the HTS-derived predictive model of reproductive toxicity by applying the model to a new set of chemicals. Independently, 297 chemicals were tested in the same or comparable HTS assays as the original 309 chemicals in Phase I of ToxCast (i.e., the chemical set used in model development). A subset of the 297 chemicals has high quality reproductive toxicity study data and served as the forward validation chemical set. The predictive model of reproductive toxicity was applied to the forward validation chemical set to assess the model's performance against this increasingly diverse set of chemicals, which included pharmaceuticals, industrial chemicals, antimicrobials, additional pesticides, and inert (i.e., other) ingredients. Case studies were then developed to assess the ability of the model to adapt to different regulatory conditions, e.g., waiving a testing requirement, requesting a study, or classifying chemicals in a data poor regulatory environment. Each case study was modeled after current regulations, but the model was assessed in isolation. Even though real-world testing decisions would be made using all available information, the goal of the case studies was to measure the impact of the current model on chemical testing decision-making. Testing the model across the various case studies ultimately demonstrated the acceptability and flexibility of the model as a tool for chemical testing decision-making.

Methods

Forward Validation Study

The EPA ToxCast™ program employs a chemical library containing 1011 samples consisting of 960 unique structures with 9 triplicates as technical repeats for internal quality control that are now incorporated into every phase of ToxCast. A summary of all *in vitro* and *in vivo* chemical libraries is graphically represented as well as the overlap between ToxCast, Tox21, and ToxRefDB, and the relationship between these testing efforts and the different phases of the predictive modeling effort (Figure 5.1). A large subset of the original Phase I library of 309 unique structures has now been re-procured for testing in new assays and currently consists of 293 unique structures, as 16 chemicals were removed from the initial library due to evidence of decomposition or sustained insolubility. The combined ToxCast chemical library of 960 chemicals contains 111 pharmaceuticals that failed in human clinical trials primarily due to toxicity, a large set of drug-like compounds, industrial chemicals, food additives, antimicrobials, additional ingredients in pesticide formulations, and pesticide active ingredients.

Additionally, a library of 1462 samples equating to 1421 unique structures was previously run at the NIH Chemical Genomics Center (NCGC) across numerous nuclear receptor and oxidative stress pathway assays in quantitative high-throughput screening format. Of the 1421 chemicals, 583 chemicals are included in the ToxCast library. In total, 297 chemicals have been tested as part of ToxCast Phase II, which were not tested in ToxCast Phase I (i.e., not included in the model development process), and tested at NCGC. Therefore, the 297chemical set has been tested across all assays used in the original predictive model of reproductive toxicity and can therefore be wholly evaluated by the model.

As the chemical library expanded beyond conventional pesticide actives the availability of guideline multigenerational reproductive toxicity (MGR) studies diminishes quickly. For the forward validation exercise, the available MGR studies were entered into EPA's Toxicity Reference Database (ToxRefDB) as described previously (Martin et al., 2009). Open literature studies were used to characterize the reproductive toxicity potential of drug or drug-like compounds and industrial chemicals. These studies included continuous breeding protocols and fertility and pre-/post-natal studies performed by the National Toxicology Program (NTP), academic and contract laboratories or pharmaceutical companies. In general, these studies were able to be successfully entered into ToxRefDB and a number were identified, for modeling purposes, as reproductive toxicants based on achieving a reproductive lowest observed adverse effect level (rLOAEL) less than or equal to 500 mg/kg/day as previously described (Martin, et al., 2011). In some instances, clear evidence of reproductive toxicity, or not, was available in the literature, but the study design was not conducive for entry into ToxRefDB. In total, 77 chemicals were characterized as being reproductive toxicants or not for purposes of applying the predictive model to the chemical set to test the forward predictivity of the model.

A complete overview of the ToxCast and NCGC assays including assay selection, analysis methods, quality measures, and assay annotation has been previously published (Huang et al., 2011; Judson, et al., 2010). Due to the prospective nature of the forward validation effort, data tested as part of ToxCast Phase II have been analyzed in a manner as similar as possible to ToxCast Phase I. The transcription factor activation assays generated by Attagene™ Inc. and indicated by "ATG" were analyzed across all ToxCast Phase II chemicals using the same procedure as previously described (Martin et al., 2010). The

ToxCast Phase II cell-free bioactivity profiling from Novascreen Biosciences was run in a single concentration screening mode in duplicate using the same protocol as ToxCast Phase I to select chemicals for concentration response testing. An analysis was performed using the single point screening and concentration response data from ToxCast Phase I to assess whether hit calling and active concentration estimations (e.g., AC50) could accurately be derived from single point data, which has been demonstrated before (Moody et al., 1999). Therefore, Novascreen assays, beginning with "NVS", had AC50 values estimated based on single concentration screening values run in duplicate. Nuclear receptor data generated at the NCGC and cell-based cell signaling assays from BioSeek LLC, referred to as "BSK", were generated and analyzed in the same manner as ToxCast Phase I.

AC50/LEC values were -log3 transformed (-log3 [AC50/1000]), and a value of 0 was given to all negative assay results. A log3 transformation and setting negatives to 1000µM was used over a log10 transformation and setting negatives to one molar, as has been done in previous publications of ToxCast results (Judson, et al., 2010), to enhance the scoring range between high and low potency active chemicals and to decrease the distance between active and inactive chemicals. Therefore, the "assay score" where the AC50 was 100µM would have a value of roughly 2 while a 100nM AC50 would have a value of roughly 8. A "gene score" or "gene-set score" was derived based on the average assay score across a set of closely related assays, e.g., assays mapped to a single gene or gene family. The forward validation chemical set includes the 77 chemicals tested in a high quality reproductive study and tested across all assays in the mode. As previously described (Martin, et al., 2011), an *in vitro* activity filter of 2% (i.e., active in less than 2 out of every 100 assays) was applied to filter out those chemicals potentially confounded by insolubility, degradation or other issues

leading to a lack of HTS testing amenability. Of these 77 chemicals, 62 chemicals had greater than 2% *in vitro* activity and were used as the primary forward validation chemical set. The model was then applied to the full 77 chemical forward validation set and to the *in vitro* activity filtered 62 chemical forward validation set resulting in model predictions based on the parameters and model cutoffs (i.e., linear discriminant analysis (LDA) intercept) previously established (Martin, et al., 2011). The predicted outcomes were then compared to the known outcomes and performance metrics were reported, including balanced accuracy and predictivity.

Model Application Case Studies

Using the totality of chemicals with available reproductive toxicity studies used in all phases of model development (N=381), the model was applied in three case studies that test the models ability to be used under different statutory contexts. The three case studies involve applying the model to chemicals with:

1. Statutory requirement to run MGR and ability to waive requirement
2. Statutory authority to request MGR
3. Limited to no statutory authority to request MGR

To customize model outputs, the predictive model was tuned to optimize specific summary statistics, including sensitivity, specificity, balanced accuracy, positive and negative predictivity, and overall predictivity. Optimization was done by adjusting the percent *in vitro* activity filter and the model cutoff. The original filter and cutoff were chosen based on optimizing the balanced accuracy alone, whereas the case studies may weight certain output statistics higher than others. To assess the impact that incorporating the model into testing and decision-making, cost savings estimates were also calculated based on the increased

evaluation or testing efficiency. In the case of waiving study requirements, cost savings per chemical (N=381 chemicals) were calculated as the cost of the HTS battery and the cost of the MGR study times the percent of un-waived studies minus the total cost of a MGR study. In the cases of requesting studies, the cost savings per chemical were calculated assuming current reproductive testing prioritization schemes are equal to random chance. For example, if 40% of chemicals are positive (i.e., prevalence) then on average it would take 10 studies to identify 4 positive chemicals chosen at random. Therefore, cost estimates per chemical were calculated as the difference between percent of studies requested when chosen randomly and chosen based on the model predictions multiplied by the cost of a MGR study.

Results

Forward Validation Study

The predictive model of reproductive toxicity was internally trained and tested using 206 chemicals, primarily pesticide active ingredients. The model was preliminarily externally validated using a set of 26 chemicals with well-characterized reproductive toxicity information, but generally lacked chemical diversity from the training dataset and the HTS tested was performed at the same time (Martin, et al., 2011). The sustained model accuracy with the internal test and external validation chemical sets successfully evaluated the model as not being over-fit and having the potential to be applied to a more diverse chemical library. To fully assess the external predictivity of the model and the reproducibility and stability of the model inputs, the model was applied to a forward validation chemical set. The forward validation chemical set was generally tested in the same HTS assays, but roughly a year later, and is a diverse set of chemicals spanning pesticides, industrial chemicals, drug-like compounds, antimicrobials, food-additives, and other formulation ingredients. To date,

297 chemicals out of 676 ToxCast Phase II chemicals have been tested in all assays required to apply the model. Of those 297 chemicals, reproductive toxicity information in ToxRefDB or from quality public (e.g. National Toxicology Program Continuous Breeding Studies) or open literature sources was obtained for 77 chemicals (Table 5.1). Based on observed reproductive LOAEL less than or equal to 500 mg/kg/day, 51 chemicals were considered positive and 26 negative. The high prevalence of positive chemicals in the forward validation set was expected due to the fact that the majority of the chemicals did not require a MGR study, but came from the open literature which may be biased towards positive findings. The 77 chemicals were further filtered to 62 chemicals, 43 positive and 19 negative, using the *in vitro* activity cutoff filter, which removed 15 chemicals that were active in less than 2% of all tested assays. The model was applied to the filtered and unfiltered forward validation chemical sets to assess the overall accuracy of the model with or without the *in vitro* activity filter applied and to assess the relative impact of filtering on interpretation of model prediction results.

Before the model could be applied, large-scale data analysis was performed on all available ToxCast HTS assays, to date. Highly comparable or identical methods were applied to the ATG, NCGC, and BSK datasets (Huang, et al., 2011; Judson, et al., 2010). The NVS dataset was only tested at a single concentration (10 micromolar for cytochrome P450 inhibition assays and 25 micromolar for all others) in duplicate. In Phase I of ToxCast a single concentration screen was performed and a subset of chemical-assay combinations was selected to be run in concentration response format. Using the single point data from NVS Phase I, AC50 values were estimated and compared to the published AC50 values derived from concentration-response analysis (Knudsen et al., 2011) to evaluate the accuracy of the

alternative approach as was previously published (Moody, et al., 1999). Using only the 16 NVS assays in the model and the 963 chemical-assay combinations run in concentration response format, the hit calls and negative-log10of the AC50 were compared (Figure 5.2). The sensitivity and specificity in comparing hit calls were 85% and 88% respectively, resulting in a balanced accuracy of 87%. The correlation of the 305 true positives was 76% with only 31 single-point AC50 estimations having greater than an order of magnitude difference. The high accuracy level and relatively good correlation between single point and concentration response AC50 values support the preliminary use of the NVS ToxCast Phase II single point data in lieu of the concentration response data. Additionally, the ToxCast Phase II chemical library was only run against the human NVS cytochrome P450 inhibition (CYP) assays as opposed to rat and human in ToxCast Phase I, which when comparing the ToxCast Phase I human CYP profile to the rat CYP profile was highly correlated at 80%. The final forward validation assay set had the same number of assays as the published model, 36 assays, covering the same 8 features.

The published predictive model of reproductive toxicity was applied to the forward validation chemical set, both *in vitro* activity filtered and not. The observed accuracy and balanced accuracy of the filtered forward validation chemical set was 77% and 72%, respectively, a value highly comparable with the published model accuracy values (Table 5.2). A 3% drop in balanced accuracy, to 69%, was observed when the unfiltered chemical set was tested, which was comparable to the drop in accuracy seen during the model development process. The model's ability to maintain predictivity in light real-world constraints, including a more diverse chemical library, HTS data generation independent of the original dataset, increased prevalence of positives in the forward validation chemical set,

single point versus concentration response data, and assay attrition and replacement, has made application of the model for testing decision-making a possibility.

Case Studies

Using the totality of chemicals with available reproductive toxicity studies used in all phases of model development (N=381), the model was applied in three case studies that test the models ability to be used under different statutory contexts. The three case studies involve applying the model to chemicals with:

1. Statutory requirement to run MGR and ability to waive requirement
2. Statutory authority to request MGR
3. Limited to no statutory authority to request MGR

Of the 381 chemicals, 151 chemicals are considered positive for the purposes of the following case studies and 230 are considered to be negative. Two model parameters, the *in vitro* activity filter and the model cutoff, were used to customize the model performance based on the requirements on each case study. The *in vitro* activity for each chemical was calculated as the number of active calls (i.e., AC50 determined) over the total number of assays. The model was not applied to the chemicals that fell below the *in vitro* activity filter. As the model was developed using linear discriminant analysis (LDA), a model cutoff or intercept was established that best separated the positive training set from the negative training set. Using the published model *in vitro* activity filter, 2%, and model cutoff, 0.6, the predictive model of reproductive toxicity performs with a balanced accuracy of 78% and applies to 305 chemicals, while 76 chemicals fell below the *in vitro* activity filter. By removing the *in vitro* activity filter, the model's balanced accuracy fell slightly to 74% but allowed the model to be applied to all 381 chemicals. It should be noted that the intention of

the case studies is not to provide a definitive mechanism for waiving or requesting studies but to demonstrate the customizability of the model as a tool in chemical testing decision-making. However, each case study was devised with specific regulations in mind, including Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA), Food Quality Protection Act (FQPA), and Toxics Substances Control Act (TSCA), to ensure the practicality and applicability of each case study.

Case study 1 involves applying the model to chemicals with the statutory requirement to run a MGR study but where the MGR testing requirement could be waived by the regulatory body to reduce the overall testing requirements for a given chemical. The overall goal of this case study is to maximally reduce the number of studies required by waiving MGR testing while maintaining the highest confidence that the chemicals waived would have little to no reproductive toxicity potential (Figure 5.3). Therefore, the applied model was tuned to have a high sensitivity and a high negative predictive rate. To reduce the false negative rate, an 8% *in vitro* activity cutoff filter was applied, meaning the chemical had to be positive in roughly 40 out of 500 assays. The model was then applied to the remaining 161 chemicals using the optimal model cutoff of 0.6, as adjusting the cutoff did not improve the model sensitivity. This resulted in a balanced accuracy of 79% with high sensitivity and negative predictivity of 89% and 88%, respectively. The summary statistics translate into 67 waivers being granted across all 381 chemicals, leading to a 17.6% reduction in required MGR studies. The 8 chemicals of the 67 total waived that showed any reproductive toxicity were Benfluralin, Pendimethalin, Tri-allate, Butafenacil, Dichlorvos, Dicofol, Spiroxamine and Tribufos, but all 8 had accompanying systemic parental toxicities at doses at or below the rLOAEL, which were predominantly driven by high dose litter size and reproductive organ

weight effects. A two-generation reproductive study in rats costs upwards of $750,000 (Scialli, 2008), while the full battery of HTS data costs roughly $30,000. The average cost savings under conditions of case study 1 would be roughly $102,000 per chemical or $39 million in total savings when 67 of 381 MGR studies were waived, a 14% reduction. Cost savings was calculated as the difference between the total cost of a MGR study and the cost of the HTS assay battery and the fractional cost of the MGR based on the percent of un-waived studies. The cost savings estimate assumes that the HTS bioactivity profiling is only being used for reproductive testing decision-making, which is unlikely, but cost sharing estimates cannot be made at this time. Additionally, the cost savings estimates solely demonstrate the savings of increased testing efficiency and does not quantify the cost of misclassification.

Case study 2 involves applying the model to chemicals with the statutory authority to request a MGR study, but for which justification needs to be provided. The overall goal of this case study is to balance the time, animal, and financial costs of requesting a MGR study with the potential impact that a MGR study could have on regulating the chemical. Therefore, the applied model was tuned to have a high specificity and a high positive predictive rate (Figure 5.4). No *in vitro* activity cutoff filter was applied as this would not increase either the specificity or positive predictivity. The model was therefore applied to all 381 chemicals using an adjusted model cutoff of 2.35 resulting in a specificity of 97% and positive predictivity of 88%. Under the conditions of case study 2, 52 chemicals would have MGR studies requested. Of the 52 chemicals with MGR study requests, 46 chemicals would cause reproductive toxicity in the MGR study and potentially impact the regulation of the chemical. For the cost savings estimation based solely on increased testing efficiency, we

calculated the number of studies that would need to be requested if selected by chance to identify 46 positive chemicals. The positive prevalence in the reference dataset was roughly 40%, which is expected to be high based on the positive reporting bias in the open literature but was used as a conservative estimate. Therefore, 113 studies would have to be conducted to identify 46 positive chemicals if selected randomly, as opposed to 52 studies when selected based on the model prediction. The total cost savings would be over $34 million in comparing 113 MGR study requests versus 52 MGR study requests and accounting for the cost of the full HTS assay battery. The average costs savings per chemical would be roughly $90,000 across the entire chemical set.

Case study 3 involves applying the model to chemicals with limited to no statutory authority to request a MGR study, but where hazard assessments still need to be performed. As most chemicals currently fall under this statutory context, assessment and testing prioritization is needed. An assessment, in theory, can be performed on those chemicals most and least likely to cause reproductive toxicity based on the applied model. The remaining chemicals can then be prioritized based on their likelihood of being a reproductive toxicant, based on their model score. To identify chemicals most likely to be reproductive toxicants, a strict cutoff of 3 was used resulting in the identification of 28 chemicals, 27 of which demonstrated clear reproductive toxicity (Figure 5.5). Using the same model filters from Case Study 1, 67 chemicals were identified as being negative for reproductive toxicity. In total, an initial assessment could be performed on 95 chemicals, a 25% reduction in the number of chemicals requiring an assessment. To produce similar reductions in the number of chemicals accurately assessed with chemicals selected randomly, 66 and 99 MGR studies would have to be performed to accurately identify 27 positive and 59 negative chemicals,

respectively. The average cost savings per chemical would be $80,000 if these chemicals were all being selected for follow-up MGR studies as calculated in case study 2. The need for safety assessment without long-term testing, the lack of statutory power to request studies, and the limited public resources to run MGR or similar studies would cost the $30,000 to run the entire battery of HTS assays and provide an initial assessment for 25% of the chemical library. The three case studies provide perspective on the capacity of the predictive model to assess a large library of chemicals and estimates on the increased cost efficiency of incorporating the predictive model into the chemical testing decision making process.

Discussion

We have developed a forward validated classification model of general reproductive toxicity capable of being optimized for a diverse set of applications regarding chemical testing decision-making. The published model of reproductive toxicity demonstrated a high level of accuracy and external predictivity, but was primarily trained and tested using conventional pesticide active ingredients (Martin, et al., 2011). The question remained whether a model trained and tested using such a narrow window of chemical space could be forward predictive across the broader chemical landscape. Using the available ToxCast Phase II chemical library and corresponding HTS data, a forward validation study was conducted using pharmaceuticals, other drug-like compounds, antimicrobials, food additives, pesticide formulation/inert ingredients, and additional pesticide active ingredients. The available *in vivo* reproductive toxicity study data quickly diminishes outside of the pesticide registration studies, yet 77 chemicals were characterized for the purposes of the forward validation study. The prevalence of positive chemicals within this set was inversely related to that of the original training and test dataset with roughly two-thirds being positive. We attribute this to

the number of chemicals for which only open literature studies were available, which has a recognized bias towards positive reporting. The published model's cutoff parameter was established based on optimized balanced accuracy with the prevalence being roughly one-third positive. Surprisingly, the cutoff of 0.6 remained optimal as did the *in vitro* activity filter of 2% or roughly 10 out of 500 assay actives. The *in vitro* activity filter removed 15 chemicals which is comparable to the roughly 20% removed in the training dataset. The 77% (72% balanced accuracy) and 70% (69% balanced accuracy) accuracy levels with the model applied to the filter and unfiltered forward validation chemical set, respectively, quickly demonstrated the forward predictivity and stability of the model.

The ability of the model to be forward predictive and to maintain predictivity in light of replacing NVS rat CYP inhibition assays with human CYP assays and the use of the single-point AC50 estimates as opposed to full concentration response data, characterizes the model as highly practical, extensible, and flexible. Over time the model inputs, the individual HTS assays, will inevitably be modified due to changes in availability and advancements in assay technology. In addition to replacing assays over time, new assays will be developed with improved ability to assess perturbations of pathways and gene targets currently in the model and to fill data gaps. The data gaps that were previously discussed include steroidogenesis and metabolic activation across the targets (Martin, et al., 2011). A further recognized data gap identified from the forward validation chemical set is germ cell mutagenicity or clastogenicity. For example, acrylamide and a related chemical, N,N'-methylenebisacrylamide, both cause dominant lethal effects on spermatids and other mutagenic/clastogenic effects to the spermatazoa (Sakamoto and Hashimoto, 1986; Sakamoto and Hashimoto, 1988) , but were false negatives in the forward validation study.

Chemicals acting through increased reactivity or mutagenicity such as the acrylamide compounds or benzene containing chemicals like nitrobenzene have a limited *in vitro* bioactivity profile as many of these chemicals are confounded by cytotoxicity and the current assay set does not sufficiently address chemical reactivity or mutagenicity. Currently, efforts are underway to develop methods to fill this data gap, including ReProComet, and could be combined with models measuring receptor- and enzyme-based mechanisms of action (Cordelli et al., 2007). Additionally, QSAR models have shown limited success in the area of genotoxicity and mutagenicity. Models relevant to germ-cell mutagenicity have been developed and could be applied as part of an integrated reproductive testing strategy (Jensen et al., 2008; Novic and Vracko, 2010).

In addition to testing the forward predictivity and reproducibility of the model, the forward validation study also identified features that may have been over predictive in the original model. The only two features not maintaining a significant association with reproductive toxicity was PPARA and PPARG. The features, AR, ESR1, CYP, GPCR, OTHER and PXR, showed similar or even increased correlation with the forward validation chemical set. The highly sensitive ATG_PPARa_TRANS and ATG_PPARg_TRANS drove the decrease in correlation as the PPARA and PPARG gene-sets with these assays removed remain significantly associated. Although a mechanistic link between PPAR activity and fertility or other reproductive impairments remains unclear (Peraza *et al.*, 2006), the role of PPAR in steroid metabolism and its activity in reproductive tissues infers that it is a plausible target for disruption of endocrine signaling and altered gametogenesis. The addition of pharmaceutical compounds targeting PPAR, e.g., glitizars and glitazones, could provide evidence concerning whether or not potent and efficacious activators of PPAR cause effects

on reproduction. However, very limited information is publically available on these compounds in terms of reproductive toxicity. One hypothesis to test in the future is the relationship between the relative potency between PPAR activity and specific off-target activities with the reproductive toxicity profile of those chemicals. In general, the interactions between the gene targets and pathways incorporated into the current forward validated model, as well as other pathways, will be explored in a systems biology context, which will lead to a better understanding of concentration, dose and time relationships.

Rather than diving into the systems biological perspective, practical case studies were derived to explore the flexibility of the predictive reproductive toxicity model. The large dataset spanning all phases of model development provided a substantial chemical library with which to test the different statutory scenarios. Each case study explored different statutory situations or contexts which exist today for the vast majority of environmental chemicals. However, the purpose of these case studies was not to provide the full statutory context or make definitive testing decisions as the real-world implementation of the existing and future regulations would take into account all available information, scientific and policy. For instance, case study 1 relates to chemicals with the statutory requirement to run a MGR study, but can be waived at the discretion of the regulatory authority. For environmental chemicals, this case study is most closely related to the regulation of pesticides active ingredients under FIFRA and FQPA, where MGR studies are generally required, yet study waiver requests are submitted by the registrants for various reasons. The application of the model would differ between food-use and non-food-use pesticides and would most likely take into account, at minimum, the available toxicity (e.g., sub-chronic or developmental studies) and exposure potential information. The application of the predictive

model may be just one piece of information used in the overall waiver or study request process. The evaluation of the case studies performed in this study focused on the optimization and customization of the model assuming the model was the only tool or information available.

The requesting of study waivers in Case study 1 aimed to test the models ability to identify likely negative compounds through optimizing the model for negative predictivity and sensitivity. The *in vitro* activity filter was raised to 8% (i.e., at minimum 8 out of every 100 assays were active for that particular chemical) increasing the probability that the chemical's bioactivity fingerprint was adequately captured and that the chemical is not acting through a mechanism not covered in the current assay battery or that the chemical is not amenable to HTS. The outputs of case study 1 are not static as regulators could tune the model to further optimize based on different criteria as shown in the mock-up of the decision dashboard display (Figure 5.3). Case study 1 captured the application of the model for waiver requests which could substantially lower the testing burden of chemical companies and also lower the study evaluation process that the regulatory bodies must perform. A better understanding of the model's uncertainties is needed to ensure studies are not waived that would have played a major role in the regulation of the chemical, i.e., identifying critical effects and dose levels of concern. The cost of even missing a single chemical that would have otherwise been identified through traditional testing is difficult to quantify. However, the cost of not increasing testing efficiency and maintaining the status quo inherently keeps thousands of chemicals on the market and in our environment with no understanding of their toxicity potential.

Case study 2 explored the situation where costly MGR studies or alternative studies could be requested given some prior relevant information. This case study relates to antimicrobials regulated under FIFRA and FQPA, loosely to new industrial chemicals under TSCA, and is akin to the decision processes in place for tier two testing in EDSP. However, the case study exclusively demonstrates the ability of the model to be customized for the purposes of efficiently requesting MGR studies based solely on model predictions. Ideally, the study request would yield detailed hazard information including mechanistic information as well as dose response characterization. This is as opposed to numerous negative findings that do not contribute to the risk assessment process. The increased efficiency over randomly selecting chemicals for further testing or for choosing chemicals based on use patterns or estimated exposure yields a greater return on the investment. In contrast to case study 1, case study 2 required optimizing the model to have high specificity and positive predictivity. The model cutoff was therefore increased from the original 0.6 level to 2.35 limiting the number of requests made to less than 15% of the total number of chemicals, which is a reasonable rate of request and would be expected to produce roughly 9 positive findings out of every 10 tested chemicals. Depending on the statutory authority a greater or lesser percentage of chemicals could be requested.

Case study 3 challenges the model in that there is little to no data or statutory authority to request data on the set of chemicals, and therefore any assessment or classification would have to be almost entirely based on the outputs of the model. Therefore, the model was optimized for nearly every performance metric. The in vitro activity filter and model cutoffs were constrained to the point that model predictions were made on only 25% of the chemicals. However, for these chemicals, one would currently have little to no

capacity to obtain reproductive toxicity assessments. The remaining 75% of chemicals could be prioritized for assessment based on their model scores. The 25% of chemicals the model was applied to has roughly 90% accuracy and predictivity approaching a level of confidence, especially when other characteristics are taken into consideration (e.g., use, exposure potential, and physical/chemical properties), that assessments or classifications could be based. The three cases studies presented cover the statutory context of the vast majority of chemicals, especially environmental chemicals. Application of the model to chemicals within each case study showed clear increases in testing efficiency, animal use efficiency, cost reduction, and the characterization of previously untested chemicals.

Further developing the classification model predictive of reproductive toxicity toward quantitative predictions of dose, life-stage, and mechanistic relevancy will require a broader spectrum of assays and integration of the information into a systems modeling context. Progression of the model toward quantitative predictions may enable the complete replacement of the multigeneration reproductive study. In the meantime, the forward validated predictive model of reproductive toxicity requires a minimal investment per chemical to produce a signature of bioactivity capable of accurately identifying candidates for further reproductive testing. The predictive tool can immediately impact chemical testing decision making and set a course for ultimate replacement of high dose animal testing.

Tables

Table 5.1. Forward validation chemical set with associated reproductive LOAEL (rLOAEL) values in mg/kg/day and ‘NE” for chemicals that no rLOAEL was established based up to the highest dose tested. Percent active shows the number of assay for a given chemical that were

considered active over the total number of assays. A two percent *in vitro* activity filter was used to subset the chemicals for forward validation.

CASRN	Chemical Name	rLOAEL (mg/kg/day)	rLOAEL ≤ 500 mg/kg/day	>2% *In Vitro* Active	% *In Vitro* Active
54965-24-1	Tamoxifen citrate	0.003	Yes	Yes	18.8
51-52-5	6-Propyl-2-thiouracil	0.1	Yes	Yes	4.0
57-74-9	Chlordane	0.125	Yes	Yes	11.0
50-28-2	17beta-Estradiol	0.17	Yes	Yes	7.6
1461-22-9	Tributyltin chloride	0.25	Yes	Yes	32.9
13311-84-7	Flutamide	0.4	Yes	Yes	12.5
79-94-7	3,3',5,5'-Tetrabromobisphenol A	0.5	Yes	Yes	9.1
100-00-5	1-Chloro-4-nitrobenzene	0.7	Yes	Yes	3.0
131-18-0	Dipentyl phthalate	0.76	Yes	Yes	7.0
143-50-0	Kepone	<1	Yes	Yes	12.5
50-55-5	Reserpine	<1	Yes	Yes	8.5
59-05-2	Methotrexate	<1	Yes	Yes	11.2
77-09-8	Phenolphthalein	<1	Yes	Yes	11.2
4151-50-2	Sulfluramid	1.34	Yes	Yes	25.6
88-85-7	Dinoseb	2.33	Yes	Yes	4.9
5915-41-3	Terbuthylazin	2.5	Yes	Yes	3.1
79-06-1	Acrylamide	3.19	Yes	Yes	3.0
2795-39-3	Potassium perfluorooctanesulfonate	3.2	Yes	Yes	10.1
732-11-6	Phosmet	6.1	Yes	Yes	4.9
119-61-9	Benzophenone	8.8	Yes	Yes	2.7
534-52-1	2-Methyl-4,6-dinitrophenol	10	Yes	Yes	11.6
84852-15-3	4-Nonylphenol, branched	10	Yes	Yes	22.3
69-09-0	Chlorpromazine hydrochloride	12.5	Yes	Yes	19.5
72178-02-0	Fomesafen	12.5	Yes	Yes	2.1
2921-88-2	Chlorpyrifos	15	Yes	Yes	5.2
59-87-0	5-Nitro-2-furaldehyde semicarbazone	15	Yes	Yes	2.1
298-46-4	Carbamazepine	20	Yes	Yes	2.7
3825-26-1	Ammonium perfluorooctanoate	30	Yes	Yes	3.4
10222-01-2	2,2-Dibromo-3-nitrilopropionamide	30	Yes	Yes	5.2
107534-96-3	Tebuconazole	30	Yes	Yes	8.2
120-83-2	2,4-Dichlorophenol	49.1	Yes	Yes	3.0
121-14-2	2,4-Dinitrotoluene	50	Yes	Yes	4.0

446-72-0	Genistein	50	Yes	Yes	9.5
87-86-5	Pentachlorophenol	60	Yes	Yes	12.2
2634-33-5	1,2-Benzisothiazolin-3-one	75.1	Yes	Yes	7.6
140-66-9	4-(1,1,3,3-Tetramethylbutyl)phenol	100	Yes	Yes	10.9
85-68-7	Benzyl-butyl-phthalate	100	Yes	Yes	5.2
119-36-8	Methyl salicylate	150	Yes	Yes	4.3
52-51-7	Bronopol	200	Yes	Yes	13.1
65277-42-1	Ketoconazole	200	Yes	Yes	8.2
577-11-7	Docusate sodium	250	Yes	Yes	16.8
1806-26-4	4-Octylphenol	400	Yes	Yes	5.5
99-66-1	Valproic Acid	500	Yes	Yes	15.2
97-54-1	Isoeugenol	700	No	Yes	4.9
102-06-7	1,3-Diphenylguanidine	NE	No	Yes	10.7
1024-57-3	Heptachlor epoxide	NE	No	Yes	3.7
126-73-8	Tributyl phosphate	NE	No	Yes	5.5
137-30-4	Ziram	NE	No	Yes	2.4
149-30-4	2-Mercaptobenzothiazole	NE	No	Yes	7.3
151-21-3	Sodium dodecyl sulfate	NE	No	Yes	22.9
2058-46-0	Oxytetracycline hydrochloride	NE	No	Yes	2.4
2783-94-0	FD&C Yellow 6	NE	No	Yes	5.8
29420-49-3	Perfluorobutane sulfonate, potassium	NE	No	Yes	6.4
298-02-2	Phorate	NE	No	Yes	3.4
307-24-4	Perfluorohexanoic acid	NE	No	Yes	4.9
319-85-7	beta-1,2,3,4,5,6-Hexachlorocyclohexane	NE	No	Yes	2.7
34590-94-8	Propanol, 1(or 2)-(2-methoxymethylethoxy)-	NE	No	Yes	9.1
50-29-3	p,p'-DDT	NE	No	Yes	4.3
59756-60-4	Fluridone	NE	No	Yes	2.4
6422-86-2	1,4-Benzenedicarboxylic acid, bis(2-ethylhexyl) ester	NE	No	Yes	2.4
88-06-2	2,4,6-Trichlorophenol	NE	No	Yes	4.0
91-53-2	Ethoxyquin	NE	No	Yes	13.4
98-95-3	Nitrobenzene	2	Yes	No	1.5
110-26-9	N,N'-methylenebisacrylamide	10	Yes	No	0.6
58-08-2	Caffeine	30	Yes	No	1.5
106-87-6	4-Vinyl-1-cyclohexene dioxide	80	Yes	No	0.6
10605-21-7	Carbendazim	100	Yes	No	1.5
84-66-2	Diethyl phthalate	100	Yes	No	0.6
108-95-2	Phenol	318	Yes	No	1.2

62-23-7	4-Nitrobenzoic acid	500	Yes	No	0.9
103-90-2	Acetaminophen	1430	No	No	0.3
117-84-0	Dioctyl phthalate	NE	No	No	1.2
1861-32-1	Chlorthal-dimethyl	NE	No	No	1.8
57-50-1	Sucrose	NE	No	No	0.9
80844-07-1	Etofenprox	NE	No	No	1.8
822-06-0	1,6-Diisocyanatohexane	NE	No	No	1.2
95-48-7	2-Methylphenol	NE	No	No	0.3

Table 5.2. Performance metrics for the predictive model of reproductive toxicity applied to the 77 or 62 chemical *in vitro* activity un-filtered or filtered forward validation set using the same model parameters as previously published (Martin, et al., 2011). The forward validation summary statistics were highly comparable to the internal and external validation summary statistics.

Previously Published Model Statistics		Previously Published Parameter Coefficients		**Forward Validation Model Statistics**			
				Activity Cutoff Filter Not Applied		**Activity Cutoff Filter Applied**	
TP	55	PPARa	1.37	**TP**	**37**	**TP**	**37**
FP	28	AR	0.98	**FP**	**9**	**FP**	**8**
FN	13	ERa	0.45	**FN**	**14**	**FN**	**6**
TN	110	PPARg	0.23	**TN**	**17**	**TN**	**11**
BA Train	77±2%	CYP	0.28	**SENS**	**73%**	**SENS**	**86%**
BA Test	74±5%	GPCR	0.5	**SPEC**	**65%**	**SPEC**	**58%**
O-BA	80%	OTHER	0.45	**BA**	**69%**	**BA**	**72%**
O-A	80%	PXR	-0.21	**A**	**70%**	**A**	**77%**
EV-A	76%	Cutoff	0.6	**OR**	**5**	**OR**	**8.5**

BA = Balanced accuracy (Average of sensitivity and specificity); Train=Five-fold cross-validation training set (average and standard deviation); Test=Five-fold cross-validation test set (average and standard deviation); O=Optimized or Full Model; EV-A=External Validation Accuracy; TP= True Positive Count; FP= False Positive Count; FN=False Negative Count; TN= True negative count; SENS= Sensitivity; SPEC= Specificity; A = Accuracy; Cutoff = LDA Model Intercept; OR= Odds ratio

Figures

Figure 5.1. Schematic representation of the chemical libraries, their overlap with each other, and their inclusion into or exclusion from the model development process is summarized. In

total, 304 of the 309 chemicals from Phase I of ToxCast were included in the model development or external validation process. Only a subset were used in the final model development or evaluation process due to filtering based on low quality of MGR or other reproductive study or low *in vitro* activity across all assays. Ultimately, 77 chemicals were tested in Phase II of ToxCast, not in Phase I, at the NCGC as part of Tox21, and that had quality reproductive study data. This chemical set was used as the primary forward validation chemical set.

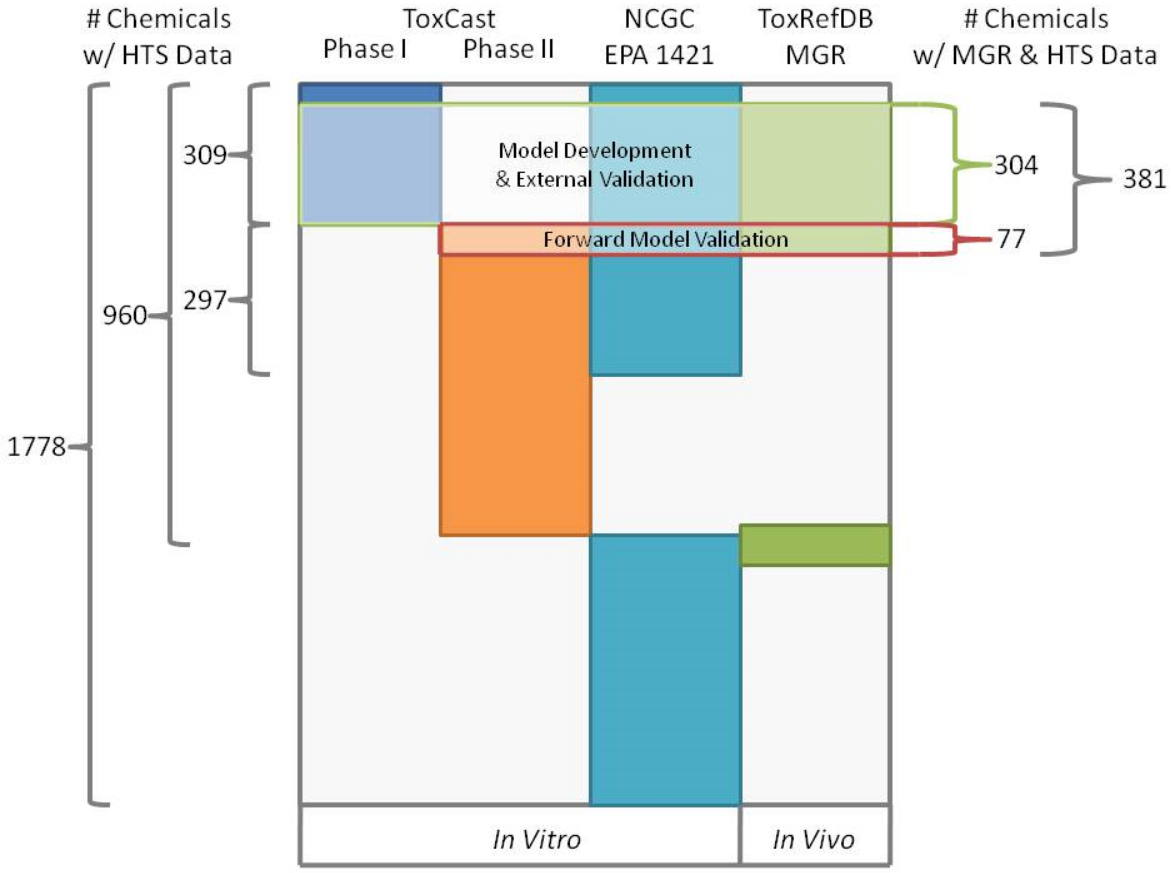


Figure 5.2. Estimation of AC50 values and determining active calls across the cell-free assays used in the initial predictive model compare to the final AC50 values and hit calls using the full concentration response data from ToxCast Phase I. Comparing the hit calls resulted in a 87% balanced accuracy, while the negative log10 AC50 among all 305 true

positives resulted in a percent correlation of 76%. Only 31of 305 chemical-assay combinations had estimated AC50 values that were greater than an order of magnitude from each other, demonstrating the ability to preliminarily use single concentration screening data from ToxCast Phase II in applying the predictive model of reproductive toxicity.

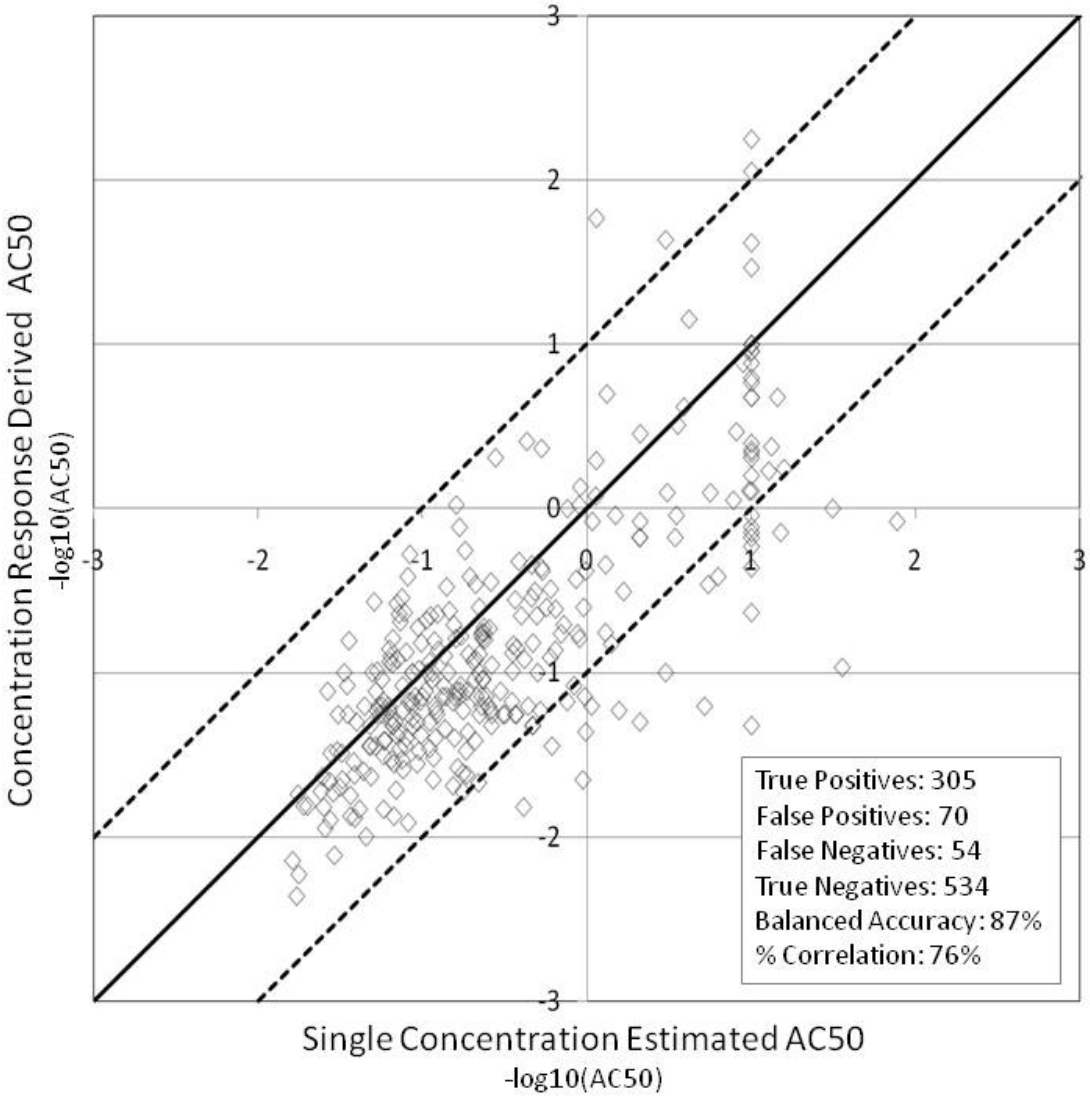


Figure 5.3. Decision dashboard for case study 1 represents an application of the predictive model of reproductive toxicity to chemicals with the statutory requirement to perform a multigeneration reproductive study (MGR). The left side of the dashboard displays specific

parameters or decisions that can be tuned or chosen and the right side represents the model outputs, including model performance. True positive (TP), false positive (FP), false negative (FN), and true negative counts are provided based on the model performance and the number of chemicals the model was applied to, which in turn computes sensitivity (Sens), specificity (Spec), positive predictivity (PPV), negative predictivity (NPV), balanced accuracy (BA) and predictivity (Pred).

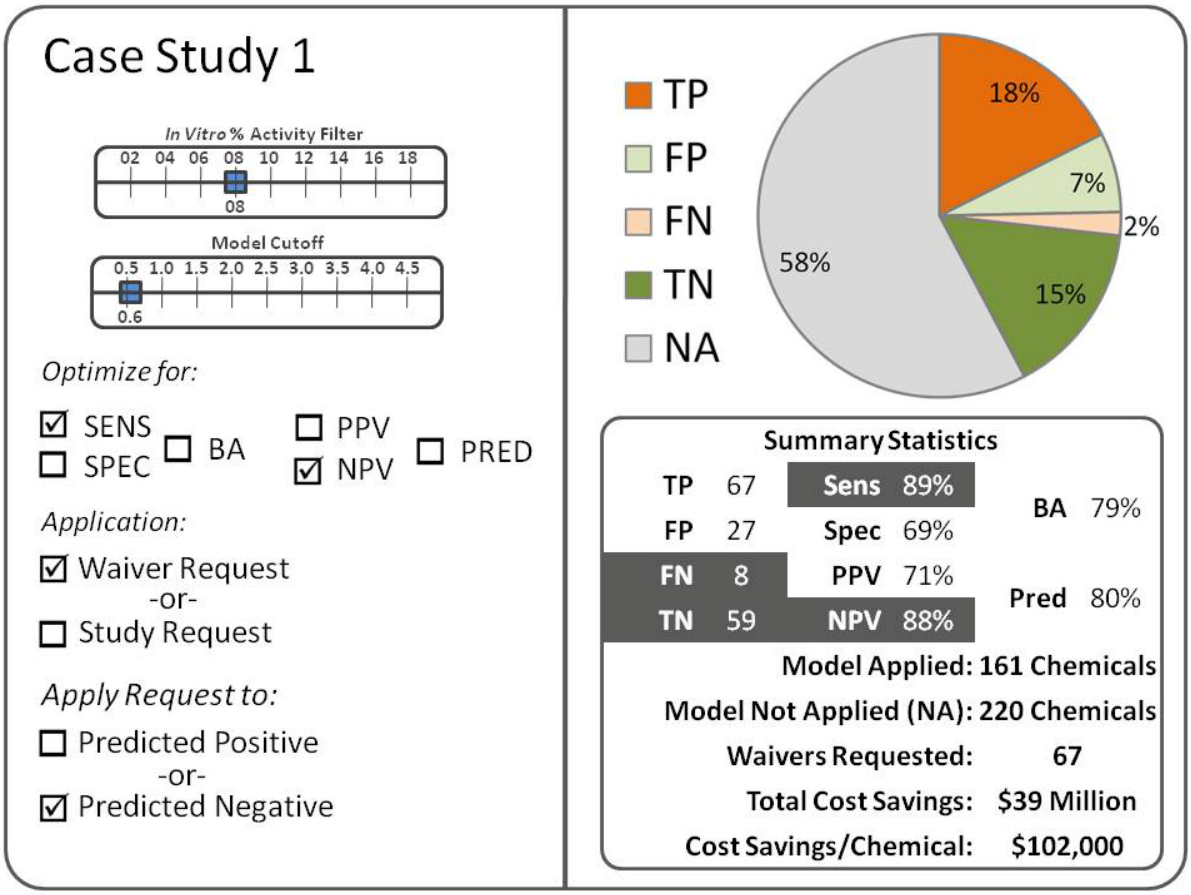


Figure 5.4. Decision dashboard for case study 2 represents an application of the predictive model of reproductive toxicity to chemicals with the statutory capacity to request a multigeneration reproductive study (MGR). The left side of the dashboard displays specific parameters or decisions that can be tuned or chosen and the right side represents the model outputs, including model performance. True positive (TP), false positive (FP), false negative

(FN), and true negative counts are provided based on the model performance and the number of chemicals the model was applied to, which in turn computes sensitivity (Sens), specificity (Spec), positive predictivity (PPV), negative predictivity (NPV), balanced accuracy (BA) and predictivity (Pred).

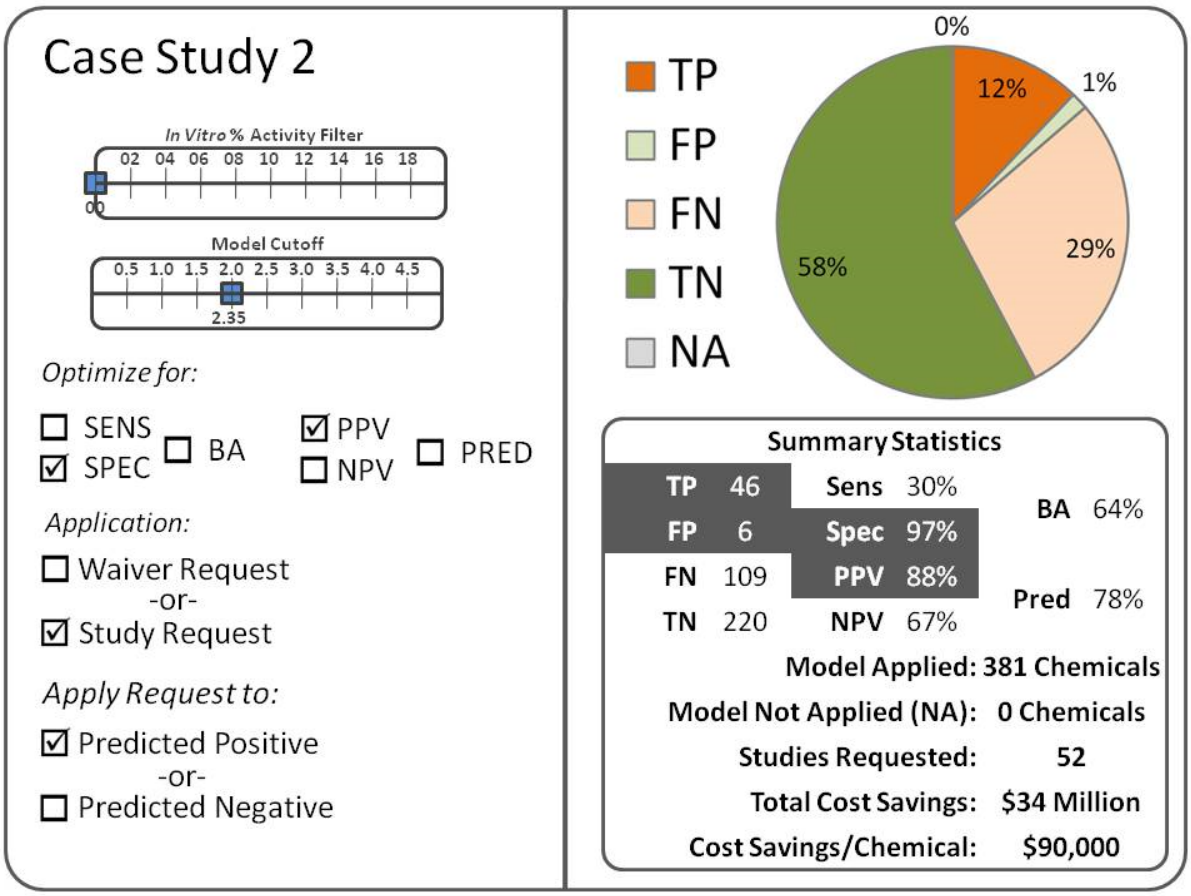
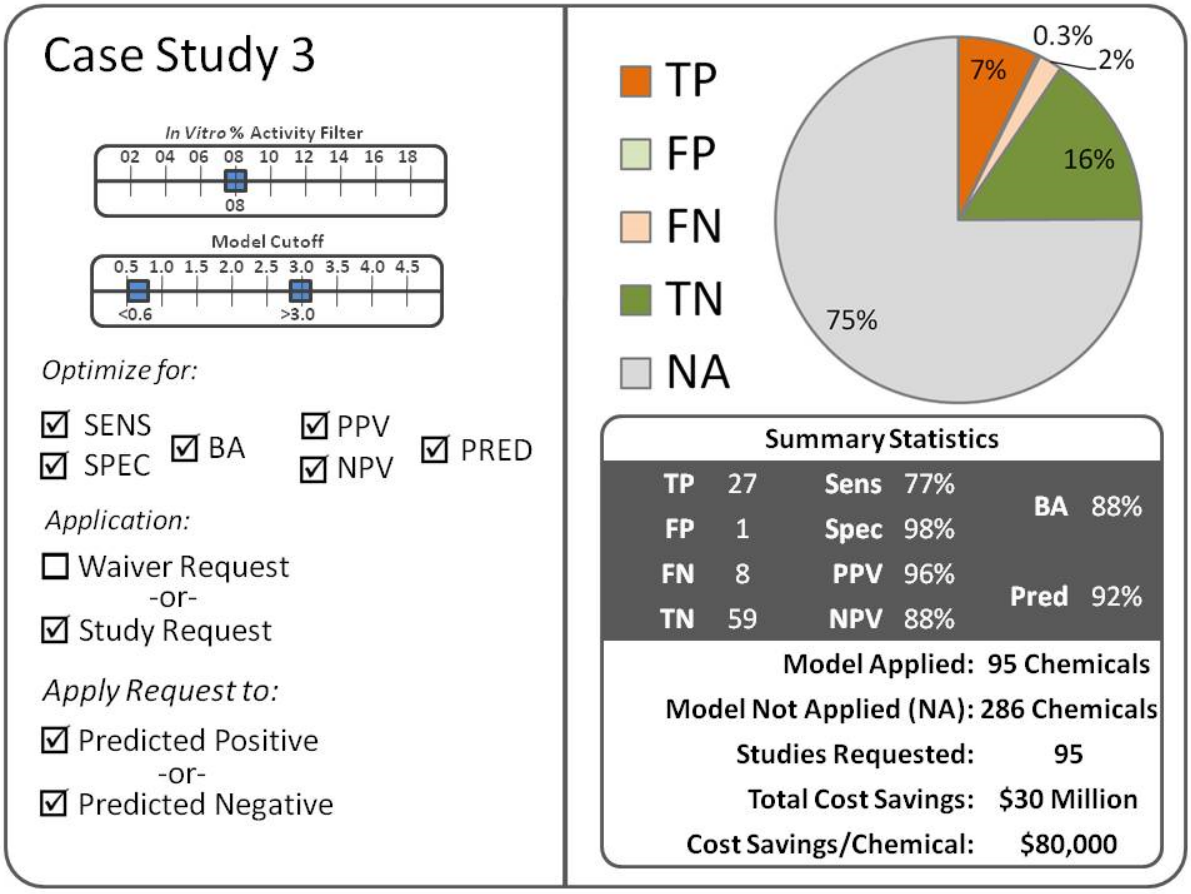


Figure 5.5. Decision dashboard for case study 3 represents an application of the predictive model of reproductive toxicity to chemicals with little to no statutory to request a multigeneration reproductive study (MGR). The left side of the dashboard displays specific parameters or decisions that can be tuned or chosen and the right side represents the model outputs, including model performance. True positive (TP), false positive (FP), false negative (FN), and true negative counts are provided based on the model performance and the number of chemicals the model was applied to, which in turn computes sensitivity (Sens), specificity

(Spec), positive predictivity (PPV), negative predictivity (NPV), balanced accuracy (BA) and predictivity (Pred).



References

Collins, F. S., Gray, G. M. and Bucher, J. R. (2008). Toxicology - Transforming environmental health protection. *Science* **319**, 906-907, DOI 10.1126/science.1154619.

Cordelli, E., Fresegna, A. M., D'Alessio, A., Eleuteri, P., Spano, M., Pacchierotti, F. and Villani, P. (2007). ReProComet: a new in vitro method to assess DNA damage in mammalian sperm. *Toxicol Sci* **99**, 545-52, kfm191 [pii]
10.1093/toxsci/kfm191.

Cronin, M. T. D. and Worth, A. P. (2008). (Q)SARs for predicting effects relating to reproductive toxicity. *Qsar Comb Sci* **27**, 91-100, DOI 10.1002/qsar.200710118.

Golbraikh, A., Shen, M., Xiao, Z., Xiao, Y. D., Lee, K. H. and Tropsha, A. (2003). Rational selection of training and test sets for the development of validated QSAR models. *J Comput Aided Mol Des* **17**, 241-53.

Golbraikh, A. and Tropsha, A. (2002). Predictive QSAR modeling based on diversity sampling of experimental datasets for the training and test set selection. *J Comput Aided Mol Des* **16**, 357-69.

Hecker, M. and Giesy, J. P. (2008). Novel trends in endocrine disruptor testing: the H295R Steroidogenesis Assay for identification of inducers and inhibitors of hormone production. *Anal Bioanal Chem* **390**, 287-91, 10.1007/s00216-007-1657-5.

Huang, R., Xia, M., Cho, M. H., Sakamuru, S., Shinn, P., Houck, K. A., Dix, D. J., Judson, R. S., Witt, K. L., Kavlock, R. J., Tice, R. R. and Austin, C. P. (2011). Chemical Genomics Profiling of Environmental Chemical Modulation of Human Nuclear Receptors. *Environ Health Perspect*, 10.1289/ehp.1002952.

Jensen, G. E., Niemela, J. R., Wedebye, E. B. and Nikolov, N. G. (2008). QSAR models for reproductive toxicity and endocrine disruption in regulatory use--a preliminary investigation. *SAR QSAR Environ Res* **19**, 631-41, 906424324 [pii]
10.1080/10629360802550473.

Judson, R. S., Houck, K. A., Kavlock, R. J., Knudsen, T. B., Martin, M. T., Mortensen, H. M., Reif, D. M., Rotroff, D. M., Shah, I., Richard, A. M. and Dix, D. J. (2010). In Vitro Screening of Environmental Chemicals for Targeted Testing Prioritization: The ToxCast Project. *Environ Health Perspect* **118**, 485-92, 10.1289/ehp.0901392.

Knudsen, T. B., Houck, K. A., Sipes, N. S., Singh, A. V., Judson, R. S., Martin, M. T., Weissman, A., Kleinstreuer, N. C., Mortensen, H. M., Reif, D. M., Rabinowitz, J. R., Setzer, R. W., Richard, A. M., Dix, D. J. and Kavlock, R. J. (2011). Activity profiles of 309 ToxCast chemicals evaluated across 292 biochemical targets. *Toxicology* **282**, 1-15, S0300-483X(11)00004-7 [pii]
10.1016/j.tox.2010.12.010.

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, Doi 10.1021/Tx900325g.

Martin, M. T., Judson, R. S., Reif, D. M., Kavlock, R. J. and Dix, D. J. (2009). Profiling Chemicals Based on Chronic Toxicity Results from the US EPA ToxRef Database. *Environ Health Persp* **117**, 392-399, Doi 10.1289/Ehp.0800074.

Martin, M. T., Knudsen, T. B., Reif, D. M., Houck, K. A., Judson, R. S., Kavlock, R. J. and Dix, D. J. (2011). Predictive Model of Rat Reproductive Toxicity from ToxCast High Throughput Screening. *Biol Reprod*, biolreprod.111.090977 [pii] 10.1095/biolreprod.111.090977.

Moody, G. C., Griffin, S. J., Mather, A. N., McGinnity, D. F. and Riley, R. J. (1999). Fully automated analysis of activities catalysed by the major human liver cytochrome P450 (CYP) enzymes: assessment of human CYP inhibition potential. *Xenobiotica* **29**, 53-75, 10.1080/004982599238812.

Novic, M. and Vracko, M. (2010). QSAR models for reproductive toxicity and endocrine disruption activity. *Molecules* **15**, 1987-99, 15031987 [pii] 10.3390/molecules15031987.

Peraza, M. A., Cromey, D. W., Carolus, B., Carter, D. E. and Gandolfi, A. J. (2006). Morphological and functional alterations in human proximal tubular cell line induced by low level inorganic arsenic: evidence for targeting of mitochondria and initiated apoptosis. *J Appl Toxicol* **26**, 356-367, Doi 10.1002/Jat.1149.

Perreault, S. D., Klinefelter G.R., Clegg E. (2008). Assessment of Male Reproductive Toxicity. Informa Healthcare, USA.

Sakamoto, J. and Hashimoto, K. (1986). Reproductive toxicity of acrylamide and related compounds in mice--effects on fertility and sperm morphology. *Arch Toxicol* **59**, 201-5.

Sakamoto, J. and Hashimoto, K. (1988). Effects of N,N'-methylene-bis-acrylamide (MBA) on mouse germ cells--sperm count and morphology, and testicular pathology. *Arch Toxicol* **62**, 54-9.

Scialli, A. R. (2008). The challenge of reproductive and developmental toxicology under REACH. *Regul Toxicol Pharmacol* **51**, 244-50, S0273-2300(08)00073-1 [pii] 10.1016/j.yrtph.2008.04.008.

Shen, M., Beguin, C., Golbraikh, A., Stables, J. P., Kohn, H. and Tropsha, A. (2004). Application of predictive QSAR models to database mining: identification and experimental

validation of novel anticonvulsant compounds. *J Med Chem* **47**, 2356-64, 10.1021/jm030584q.

Toda, K., Okada, T., Miyaura, C. and Saibara, T. (2003). Fenofibrate, a ligand for PPARalpha, inhibits aromatase cytochrome P450 expression in the ovary of mouse. *J Lipid Res* **44**, 265-70, 10.1194/jlr.M200327-JLR200
M200327-JLR200 [pii].

Tropsha, A. and Golbraikh, A. (2007). Predictive QSAR modeling workflow, model applicability domains, and virtual screening. *Curr Pharm Des* **13**, 3494-504.

Witorsch, R. J. and Thomas, J. A. (2010). Personal care products and endocrine disruption: A critical review of the literature. *Crit Rev Toxicol* **40**, 1-30, Doi 10.3109/10408444.2010.515563.

CHAPTER 6

GENERAL DISCUSSION

The combined results of this research demonstrate that bioactivity profiling from high throughput screening (HTS) data can be used to accurately predict reproductive toxicity potential and be used to increase chemical testing efficiency. In arriving at a forward predictive model of reproductive toxicity with the capacity to change chemical testing decision making, foundational work had to be performed to ensure data quality, modeling transparency, and repeatability. The foundational methodologies employed included standardization of toxicological vocabularies, database development, HTS data analysis and informatics, classification modeling, and validation procedures.

In Vivo – Capturing Traditional Reproductive Toxicity Information

Reproductive toxicity, as defined in this research, is chemical effects on reproductive performance (i.e., fertility and fecundity), the reproductive tract, and/or sexual development. Reproductive toxicity has been routinely assessed, using laboratory animal studies, in the chemical risk assessment process for over forty years. For environmental chemicals, the multi-generation reproduction and fertility study in rats has been the primary tool for assessing reproductive toxicity potential in humans. Unfortunately, these expensive and animal intensive studies have only been conducted on a fraction of environmentally relevant chemicals and the studies have been dispersed across many filing cabinets, computers, and file management systems over the years. Limited efforts have been made to make these data

useful beyond individual chemical risk assessments and to normalize study design and effect nomenclature. Before advancements in reproductive toxicity testing decision-making could be made using computational modeling and alternative methods, the scientific community needed to fully understand the body of information that existed. The Toxicity Reference Database (ToxRefDB) was the first large-scale effort to capture and house the dispersed library of reproductive toxicity information in a consistent and transparent manner (as described in Chapter 2).

The ToxRefDB reproductive toxicity study dataset has opened the doors to the broad evaluation of the multi-generation reproductive toxicity test with goals of refining the study design, reducing animal use with alternative methods, or replacing the study altogether. The emphasis on using an accepted controlled vocabulary for capturing the study and effect information permits the integration of analyses and the common interpretation of results across many laboratories, study authors, study reviewers, study designs, and over forty years of data. A retrospective analysis, primarily using ToxRefDB, comparing first and second generation relative sensitivity has helped push forward the adoption of an extended one-generation protocol (Piersma et al., 2010), while a similar description of the data was used in defining reproductive toxicants for developing a predictive model (as described in Chapter 4). In the past, these analyses would not have been connected or comparable in any way, thus hindering the ability to evaluate the impact of either analysis on refining toxicity testing strategies. As alternative test methods for reproductive toxicity continue to be developed (Schenk et al., 2010), transparent and consistent definition and use of chemical training sets need to be used in order to adequately and uniformly assess new methods and approaches.

In Vitro – Analyzing and Interpreting High Throughput Screening Data

High throughput screening (HTS) has the ability to tests thousands of chemicals in concentration response across hundreds of *in vitro* molecular and cellular assays for tens of thousands of dollars as compared to the hundreds of thousands of dollars to run a single two-generation reproductive study. Much like structural features, HTS assays provide uniform data for which to characterize a large chemical library, but are anchored to molecular and cellular events often tied to known biological pathways. HTS, as often performed by the pharmaceutical industry, has its own unique challenges including the screening of millions of chemicals as opposed to thousands as is being done in HTS application toward toxicity testing. However, pharmaceutical screening is not typically performed in concentration response and aims to identify a select set of very potent and efficacious chemicals against a single molecular target as lead candidates for further evaluation. HTS application toward toxicity testing has generated and reinvigorated areas of research including performing broad spectrum biological activity profiling, routinely running HTS in concentration response mode, and discovering the appropriate informatics and analysis techniques to handle and interpret the large quantity of data (as described in Chapter 3). The U.S. EPA ToxCast research program has produced the largest publically available broad spectrum HTS dataset and a few unique themes and lessons have emerged from this diverse dataset, including the difference between statistical and biological relevance and the need for common but flexible data analysis workflows.

Generally, a four-parameter hill model can be fit to the concentration response data producing a model that estimates the baseline or bottom response, maximal or top response, the slope of the response, and the 50% activity concentration or AC50. Data pre-processing,

parameter constraining, confounding, statistical and biological considerations all play a role into the final set of conditions used to define active or inactive concentrations of a particular chemical-assay combination. Specific considerations include, but are not limited to; data normalization methods, positive and negative control performance and availability, baseline response adjustments, plate or well variations, response saturation (e.g., full versus partial agonist), active concentration extrapolation, outlier detection and removal, response directionality, non-monotonic response detection, cytotoxicity filtering and interpretation, fluorescence and other artifact detection, statistical versus biological significance, and target specificity. Taking into account many of these considerations enables summary statistical outputs to be compared appropriately across assays and assay technologies and to be used in downstream modeling applications. These complex and often dependent confounders support the idea that no single assay is "truth" and that incorporating orthogonal or parallel assays provides much needed testing redundancy helping ensure an accurate characterization of a chemical's potential to interact with a particular biological target or pathway. Obtaining accurate activity calls and accurately annotating the assays prepares the large data-set for integration. Pathway-based analysis has been the focus of many researchers in the genomic-era. However, mapping roughly 500 assays to genes and then to pathways results in hundreds to thousands of potential pathways with high levels of redundancy. To lower the complexity of data integration, assays were mapped to gene or aggregated to related gene-sets (e.g., all cytochrome P450 inhibition assays) reducing the dataset down from over 500 unique assays to less than 100 gene or gene-sets. The result of this data reduction process creates an increasingly independent dataset, i.e., decreased correlation between input features, and assists in the practical long-term implementation of downstream models and tools. The

model's features are genes or gene-sets and do not necessarily rely on a single assay to be available and allows for seamless upgrades to the system as new and improved assay technology becomes available.

Model Development – Combing HTS and Animal Toxicity Data

Computational modeling is an interface between reference *in vivo* reproductive toxicity data and *in vitro* assay (i.e. HTS) data. Reproductive toxicity is an aggregated multi-modal and multi-effect outcome. No single assay has the ability to broadly identify reproductive toxicants. Computational modeling allows one to explore the complex relationships between *in vivo* observations and networks of *in vitro* activity. One of the more simplistic computational modeling approaches is the development of a classification model, which aims to accurately classify or predict an outcome based a training set with known outcomes. The training set for modeling reproductive toxicity was the set of chemicals in the ToxCast library with high quality reproductive toxicity data (as described in Chapter 4). The initial inputs into the model were the hundreds of ToxCast assays that were collectively mapped to genes and the aggregate activity across the assays per gene provided the quantitative inputs into the model. The assay-gene combinations were further filtered based on a feature selection process that evaluated the statistical association to the training set data. The filtered gene set was then weighted in a multivariate model using linear discriminate analysis (LDA) and five-fold cross-validation. Many other approaches and methods could have been deployed, but our observation has been that using complex machine learning algorithms have a tendency to over-fit the data lowering the outputted model's ability to be externally predictive. The resulting internal model performance statistics were greater than or

equal to 75% balanced accuracy and there was no significant difference between the training and test set accuracies.

For the purposes of clarity and in the context of the predictive reproductive toxicity model, the terms ‘external validation’ and ‘forward validation’ are distinguished from each other and defined. External validation is the testing of the predictive model’s performance using data generated alongside data used in the training of the model, but not used in the developing the initial model. Forward validation is the testing of the predictive model’s performance using data generated independent of the data used in the training of the model and for which the results were not previously known. Among the chemicals selected for external validation, the model provided accurate predictions for 16 of the 21 chemicals. The five chemicals with inaccurate predictions provide valuable insight into potential limitations or gaps of the model. Interestingly, the 5 chemicals had a common phenotypic profile with respect to reproductive toxicity causing reduced early offspring survival, particularly litter size decrease with little to no accompanying effects on reproductive performance or reproductive tract pathology. The rLOAEL for all 5 chemicals was set at the high dose tested based on the early offspring survival effects and the parental and offspring LOAEL were set at the lower dose levels. Based on the inclusive definition used for defining a positive for reproductive toxicity for model development all 5 were considered positive, but lack evidence of specific fertility-related or developmentally sensitive reproductive outcomes. Nonetheless, a gap in model predictivity has been identified and could potentially be filled using additional assay technologies, physical chemical properties and structural descriptors, or acute or short-term *in vivo* studies.

Biological Plausibility – Key Targets and Pathways Identified

The model development process identified biologically plausible features and pathways from over 500 assays mapped to less than 100 genes or gene-sets and spanning many reproductive relevant modes-of-action. PPARα activity was clearly associated with reproductive toxicity, with all 10 PPARα agonists in the training set causing reproductive toxicity. Although a mechanistic link between PPAR activity and fertility or other reproductive impairments remains unclear (Peraza *et al.*, 2006), the role of PPAR in steroid metabolism and its activity in reproductive tissues infers that it is a plausible target for disruption of endocrine signaling and altered gametogenesis. AR and ERα activity was also associated with reproductive toxicity. The ToxCast receptor profiling identified most if not all the known anti-androgenic and estrogenic chemicals in the current dataset, but the causal relationship between reproductive toxicity and steroid receptor activity, absolute and relative potency and efficacy, needs to be explored further. CYP enzyme inhibition, as compared to gene induction, was significantly more associated with reproductive toxicity. Alterations in steroid metabolism through CYP induction have been previously associated with reproductive impairment (Goetz *et al.*, 2007), however the non-specific inhibition of CYPs may be a surrogate for a chemical's capacity to disturb steroid metabolism including inhibition of key CYPs involved in steroidogenesis (e.g., Cyp19 and Cyp17). Related to CYP activity, PXR interestingly displayed a negative correlation/association with reproductive toxicity. In general, PXR lowered the false positive rate of the model by lowering the model score of chemicals with non-specific and low potency nuclear receptor activity. Robust PXR activity is an indication of potent xeno-sensing and potentially rapid metabolism. A major component of the model not directly related to nuclear receptor biology and

xenobiotic/steroid metabolism was GPCR binding. Numerous GPCR binding assays were significantly associated with reproductive toxicity. Those chosen to represent the GPCR family were selected for statistical, and not biological, reasons as there is limited literature information on their role in reproduction in contrast to their well characterized role in nervous system function. Platforms measuring EGFR, TGF-β1 and NF-κB activity were also associated with reproductive toxicity. All three gene products have been shown to modulate the relative sensitivity of developmental toxicants, especially aryl-hydrocarbon receptor signaling (Abbott *et al.*, 2003; Tian *et al.*, 1999) and may be indicative of altered xenobiotic metabolism, cellular proliferation, cell-cell signaling or potential epigenetic effects (Tian, 2009; Tian *et al.*, 2002). Overall the key targets in the model identify plausible modes of action leading to reproductive toxicity covering anti-androgenic, estrogenic, cholesterol/steroid metabolism, limited coverage of disruption of steroidogenesis, and altered xenobiotic metabolism modes of action.

Validation – Evaluating the Developed Model

Forward validation of the predictive reproductive toxicity model ultimately measures the stability of the entire system. The model was previously shown to be stable using cross-validation (75% accuracy across the test sets) and robust using external validation (75% accuracy). Forward validation not only further tests the external predictivity of the model across an increasingly diverse chemical set, but also tests the reproducibility and overall quality of the input data (i.e., HTS data). The additional 297 chemicals run across the entire assay set used in developing the model produced 77 chemicals with high quality reproductive toxicity study data. Key differences between the external validation dataset and the forward validation dataset included the fact that the cell-free assays were all run in single-point versus

full concentration response and the rodent cytochrome P450 inhibition (CYP) assays were replaced with human assays due to cost reduction and testing efficiency. These real-world factors truly test the applicability of the predictive model and interesting appeared to have little impact on model performance. Using ToxCast Phase I data, the single-point activity calls were roughly 90% concordant with concentration response activity calls and the human CYP assays were roughly 70% correlated to the rodent. Additionally, applying the *in vitro* activity filter of 2% (i.e., 2 out of every 100 assays were active for a particular chemical) removed 15 of the 77 chemicals from the forward validation chemical set. In applying the model to the forward validation chemical set with and without the activity filter the model remained highly accurate with 77% accuracy (72% balanced accuracy) and 70% accuracy (69% balanced accuracy), respectively. The maintained high level of accuracy progressing from training, to testing, to external validation, and finally to forward validation demonstrates the applicability of the model to classify the reproductive toxicity potential of chemicals.

Application – Developing a Tool for Chemical Testing

There currently is great need for a tool that can inform chemical testing decisions especially with regards to reproductive toxicity testing due to heavy animal use, high cost, and increasing concern for chemicals effects on reproductive development (Scialli, 2008). The forward validated predictive model of reproductive toxicity would undoubtedly be improved by incorporation of other available models, alternative test methods and institutional knowledge, especially in regards to supplementing recognized weaknesses in the model including steroidogenesis and germ-cell mutagenicity (Cordelli et al., 2007; Hareng et al., 2005). The validation status and transparency of the various alternative methods, tools or

models varies, but currently would be difficult to integrate due to discordant chemical coverage and a lack of data availability at this time. Nonetheless, evaluating the current model in isolation gives great insight into its potential impact on chemical testing decision-making. Three case studies were derived varying the statutory context varying from requiring a MGR study, to requesting a MGR study, to no authority to request a study. Using the full chemical dataset with reproductive toxicity information (N=381chemicals), the first case study tested the ability of the model to accurately identify the most likely negative chemicals in order to apply a waiver to those chemicals. Based on optimized model sensitivity and negative predictivity, waivers could be applied to 67 chemicals out of the 381 total, a roughly 20% reduction in reproductive testing and total cost. In contrast, case study 2 has the statutory authority to request studies when evidence suggests the need for a study. Therefore, the model was optimized to identify most likely positive chemicals resulting in 52 of the 381 chemicals resulting in MGR study requests. Each study costs upwards of $750,000, but the increased efficiency in running studies that result in important positive findings potentially being used in chemical risk assessment would save roughly $90,000 per chemical across all chemicals that would fall under case study 2. Case study 3 provides the biggest challenge because of the limited authority to request studies and the current number of chemicals that fall into this category. In starting to tackle the need for assessing these chemicals the model was applied by attempting to optimize the confidence in both negative and positive predictions. Therefore, the model was applied to the lowest and highest scoring chemicals with a high *in vitro* activity filter resulting in 25% of the chemicals being classified as highly likely to be a reproductive toxicant (27 of 28 chemicals positive) or not (59 of 67chemicals negative). The remaining 75% of chemicals could be prioritized for further testing or

evaluation based on their model scores. The three cases studies cover the statutory context of the vast majority of chemicals, in particular environmental chemicals. Application of the model to chemicals within each case study showed clear increases in testing efficiency, animal use efficiency, cost reduction, and the characterization of previously untested chemicals.

With the availability of a forward validated classification model predicting reproductive toxicity, the bottleneck of uncharacterized chemicals can be evaluated either through improvements in the overall statutory authority to request MGR studies or in the ability to quantitatively identify reproductive toxicants. If the statutory authority to request MGR studies were improved, then the current model in concert with other models, alternative methods and institutional knowledge could identify with fairly good accuracy and efficiency all chemicals that a MGR should be requested. If the latter were improved to the point of accurate adverse dose predictions, then the model could drastically decrease the need for MGR studies and be used in the assessment of the majority of environmental chemicals. To do this, improvements in HTS assay reproducibility, metabolic capacity, mode-of-action coverage, reverse toxicokinetics, and overall model accuracy would need to be made. Placing the classification model into a systems modeling context will begin to address these next generation of research questions. For now, the forward validated predictive model of reproductive toxicity can go a long way in improving reproductive chemical testing efficiency and decision-making.

References

Abbott, B. D., Buckalew, A. R., DeVito, M. J., Ross, D., Bryant, P. L. and Schmid, J. E. (2003). EGF and TGF-alpha expression influence the developmental toxicity of TCDD: Dose response and AhR phenotype in EGF, TGF-alpha, and EGF plus TGF-alpha knockout mice. *Toxicol Sci* **71**, 84-95.

Cordelli, E., Fresegna, A. M., D'Alessio, A., Eleuteri, P., Spano, M., Pacchierotti, F. and Villani, P. (2007). ReProComet: a new in vitro method to assess DNA damage in mammalian sperm. *Toxicol Sci* **99**, 545-52, kfm191 [pii]
10.1093/toxsci/kfm191.

Goetz, A. K., Ren, H. Z., Schmid, J. E., Blystone, C. R., Thillainadarajah, I., Best, D. S., Nichols, H. P., Strader, L. F., Wolf, D. C., Narotsky, M. G., Rockett, J. C. and Dix, D. J. (2007). Disruption of testosterone homeostasis as a mode of action for the reproductive toxicity of triazole fungicides in the male rat. *Toxicol Sci* **95**, 227-239, DOI 10.1093/toxsci/kfl124.

Hareng, L., Pellizzer, C., Bremer, S., Schwarz, M. and Hartung, T. (2005). The Integrated Project ReProTect: A novel approach in reproductive toxicity hazard assessment. *Reprod Toxicol* **20**, 441-452, DOI 10.1016/j.reprotox.2005.04.003.

Peraza, M. A., Cromey, D. W., Carolus, B., Carter, D. E. and Gandolfi, A. J. (2006). Morphological and functional alterations in human proximal tubular cell line induced by low level inorganic arsenic: evidence for targeting of mitochondria and initiated apoptosis. *J Appl Toxicol* **26**, 356-367, Doi 10.1002/Jat.1149.

Piersma, A. H., Rorije, E., Beekhuijzen, M. E., Cooper, R., Dix, D. J., Heinrich-Hirsch, B., Martin, M. T., Mendez, E., Muller, A., Paparella, M., Ramsingh, D., Reaves, E., Ridgway, P., Schenk, E., Stachiw, L., Ulbrich, B. and Hakkert, B. C. (2010). Combined retrospective analysis of 498 rat multi-generation reproductive toxicity studies: on the impact of parameters related to F1 mating and F2 offspring. *Reprod Toxicol*, S0890-6238(10)00351-5 [pii]
10.1016/j.reprotox.2010.11.013.

Schenk, B., Weimer, M., Bremer, S., van der Burg, B., Cortvrindt, R., Freyberger, A., Lazzari, G., Pellizzer, C., Piersma, A., Schafer, W. R., Seiler, A., Witters, H. and Schwarz, M. (2010). The ReProTect Feasibility Study, a novel comprehensive in vitro approach to detect reproductive toxicants. *Reprod Toxicol* **30**, 200-218, DOI 10.1016/j.reprotox.2010.05.012.

Scialli, A. R. (2008). The challenge of reproductive and developmental toxicology under REACH. *Regul Toxicol Pharmacol* **51**, 244-50, S0273-2300(08)00073-1 [pii]
10.1016/j.yrtph.2008.04.008.

Tian, Y., Ke, S., Denison, M. S., Rabson, A. B. and Gallo, M. A. (1999). Ah receptor and NF-kappa B interactions, a potential mechanism for dioxin toxicity. *J Biol Chem* **274**, 510-515.

Tian, Y. A. (2009). Ah receptor and NF-kappa B interplay on the stage of epigenome. *Biochem Pharmacol* **77**, 670-680, DOI 10.1016/j.bcp.2008.10.023.

Tian, Y. N., Rabson, A. B. and Gallo, M. A. (2002). Ah receptor and NF-kappa B interactions: mechanisms and physiological implications. *Chem-Biol Interact* **141**, 97-115, Pii S0009-2797(02)00068-6.

CHAPTER 7

FUTURE DIRECTIONS

The development of a classification model predictive of rodent reproductive toxicity has provided the chemical testing community with a tool for prioritizing and targeting testing decisions for chemicals with unknown toxicities. Practical implementation of this tool requires large scale testing of thousands of chemicals across, at minimum, the assays used in the model. As additional models are generated and integrated with each other, hundreds of high throughput screening (HTS) assays will be required and be used to test, refine and add to the high throughput predictive toolbox. The cost to run over 500 assays in roughly 10 distinct technologies is roughly $30,000 per chemical in contrast to the greater than twenty million dollars it costs to run the battery of animal toxicity studies used in pesticide registration. The nominal costs to run such a large library of assays makes expansion of the current assay set reasonable, especially if data generation and analysis costs are shared or distributed among government regulatory agencies and chemical companies. The current predictive model of reproductive toxicity is a good example of how we will continue to refine and expand the model.

Assay Development to Fill Biological Gaps

The predictive model has recognized data gaps in terms of known and unknown modes of toxic action, assay coverage, and pathway redundancy. The model contains non-specific cytochrome P450 inhibition as a feature predictive of reproductive toxicity potential

yet provides little specific indication of a chemical's potential to disrupt steroidogenesis. To cover the steroidogenic mode-of-action, multiple assays may be required but a good start would be the incorporation of the H295R steroidogenesis assay currently used in the Endocrine Disrupter Screening Program (EDSP) and that is currently being converted into a higher-throughput 96-well format (Hecker and Giesy, 2008; Hecker et al., 2011). Another identified mode-of-action gap in the current model is the detection of germ-cell mutagens and clastogens. Currently, a number of assays are available, but lack throughput for testing thousands of chemicals (Cordelli et al., 2007). Future research in advancing the throughput and mechanistic coverage of assays or models targeting the identification of germ-cell damaging chemicals is needed and is an area in which structure-based modeling may be a useful tool (Grindon et al., 2008). The identification and filling of biological gaps in the model will only increase the understanding and relationships between features in the model and the resulting predictions. An initial step in relating the biological responses, predicted outcome and the actual adverse outcome is the use of pharmacokinetic information. The assumption would be that the biological features in the model would have to be activated at or below concentrations that cause the adverse reproductive outcome. Currently, a concept of reverse pharmacokinetics, a combination of *in vitro* hepatic clearance, plasma protein binding and a simple pharmacokinetic model, is being used to estimate the administered dose required to achieve a steady-state concentration in the blood equivalent to the active concentrations observed in a particular HTS assay (Rotroff et al., 2010). In short, one can convert micromolar active concentrations to milligram per kilogram of bodyweight per day (mg/kg/day). Early work has shown that relationship exist between HTS targets linked to a particular mode-of-action and whole animal toxic outcomes for chemicals with a known

modes-of-action (Judson et al., 2011). However, additional research is needed to incorporate this reverse toxicokinetic information into the model development process and using the information to corroborate and link biological activity with toxicological activity.

Neuroendocrine Systems Modeling

To advance the reproductive modeling effort being classification modeling, the effort will need to be put into a system modeling framework. One such approach is the development of virtual tissues or systems as is currently being done with the liver and embryo. Directly related to the prediction of reproductive outcomes, would be the modeling of the hypothalamic-pituitary (HP) axis, the primary initiator of the onset of puberty and key in the maintenance of gonadal (G), thyroid (T), and adrenal (A) function. Many of the key normal physiological events involved in pubertal onset, sexual maturation, steroidogenesis, and oogenesis have been well described based on observed hormonal, protein and gene changes (Hoyer, 2010; Perreault, 2008). However, the exact sequence of these events and the ability of chemicals to perturb these well orchestrated feedback mechanisms have yet to be elucidated in full. Currently, a variety of in vivo and in vitro assays exist that explore chemical perturbations of the HP(GTA)-axis, including estrogenicity, androgenicity, aromatase, steroidogenesis, thyroid function, puberty, and sexual maturation and will continue to be refined, scaled-up, and validated. The majority of these mechanisms are covered in the current predictive model of reproductive toxicity, but dose- and time-relationships can be explored further within the virtual HP(GTA)-axis. The ultimate goal would be to have a working model of the HP(GTA)-axis that would incorporate high-throughput testing results and other data to make detailed predictions of an uncharacterized

chemical's effect on reproduction and sexual development with life-stage, dose response, and mechanistic specificity.

Extension to Predicting Human and Ecological Reproductive Toxicity

The systems modeling approaches also help progress from predicting animal toxicity to human toxicity and extrapolating to ecologically-relevant species. Very limited causal evidence exists pertaining to chemical effects on human reproduction, thus the inability to predict human reproductive toxicity directly, to date. However, through the EPA ToxCast research program marketed and failed pharmaceutical compounds have been included into the chemical library and the respective pharmaceutical companies have contributed pre-clinical and clinical data. Efforts are underway to make the clinical data useful and to begin to bridge between HTS, pre-clinical, and clinical outcomes. The number of chemicals with high quality clinical fertility or other reproductive parameters tested will most likely be limited, but nonetheless may provide hints as to the human relevancy of some of the targets in the current or future models. In addition to predicting human toxicity, research efforts are needed to evaluate the challenges in extending models predictive of mammalian toxicity to ecologically-relevant species, including reptiles, amphibians, fish, and birds. Work similar to ToxRefDB has been performed for pesticide registration studies for ecological risk assessments (Russom, 2002). The data from ECOTOX could provide valuable bridging data by comparing, for example, avian reproduction study data to rat multigeneration reproductive study data, but additional work is needed to make these data comparable. Much attention has been placed on testing for endocrine disruption for both human and ecological risk assessments (Ankley et al., 2009). However, few ecologically-relevant species have been the target of endocrine screening especially for assays outside of estrogen and androgen receptor

activity. Assay development needs to be focused on these other species as well as expanding beyond estrogen and androgen assays into screening for HP(GTA) effects (Watanabe et al., 2009). Lessons learned from developing predictive reproductive toxicity models of ecological species, especially if done so from a population perspective, could have great impact on the use of predictive models towards human health risk assessment. Likewise, performing systems modeling determining best practices for specific extrapolation from rodent to human in HTS predictive models will greatly impact approaches applied in environmental toxicology. Using computational toxicology as the interface between typically distinct risk assessment processes should help advance both regulatory sciences or even merge them into one with an eye toward full life-cycle assessments.

Developing an Integrated Testing Strategy

An early step toward assessing chemicals in an integrated fashion is the combining of predictive toxicity models and alternative test methods as part of an overarching integrated testing strategy. Similar efforts to develop predictive models of prenatal developmental toxicity using HTS data have shown promise and the compatibility with the predictive model of reproductive toxicity will make the integration of the two models straightforward. These HTS-derived models will enable the evaluation of the vast majority of environmental chemicals with the initial capacity to prioritize chemical testing or make other simple chemical testing decisions. In addition to the systems modeling work that needs to be performed to add dose, time, and mechanistic relevance to the classification models, equal efforts in developing and validating alternative test methods for evaluating developmental and reproductive toxicity in alternative test species or medium throughput complex assay systems are required. Ideally, the systems modeling and the test method development will

feed off of each other with the systems modeling identifying molecular and cellular mechanisms to focus assay development and for the complex assay systems to provide data feeding systems modeling efforts. As the mechanisms leading to toxicity continue to be elucidated and the ability to assay those mechanisms becomes available, the need and reliance on animal testing will decline. The focus and usefulness of chemical-by-chemical risk assessment will diminish and the integrated evaluation of chemicals in the context of pathways, co-exposures, lifestyle and communities will be the drivers of chemical safety assessments.

References

Ankley, G. T., Bencic, D. C., Breen, M. S., Collette, T. W., Conolly, R. B., Denslow, N. D., Edwards, S. W., Ekman, D. R., Garcia-Reyero, N., Jensen, K. M., Lazorchak, J. M., Martinovic, D., Miller, D. H., Perkins, E. J., Orlando, E. F., Villeneuve, D. L., Wang, R. L. and Watanabe, K. H. (2009). Endocrine disrupting chemicals in fish: developing exposure indicators and predictive models of effects based on mechanism of action. *Aquat Toxicol* **92**, 168-78, S0166-445X(09)00039-3 [pii]
10.1016/j.aquatox.2009.01.013.

Cordelli, E., Fresegna, A. M., D'Alessio, A., Eleuteri, P., Spano, M., Pacchierotti, F. and Villani, P. (2007). ReProComet: a new in vitro method to assess DNA damage in mammalian sperm. *Toxicol Sci* **99**, 545-52, kfm191 [pii]
10.1093/toxsci/kfm191.

Grindon, C., Combes, R., Cronin, M. T., Roberts, D. W. and Garrod, J. F. (2008). Integrated decision-tree testing strategies for developmental and reproductive toxicity with respect to the requirements of the EU REACH legislation. *Altern Lab Anim* **36**, 65-80.

Hecker, M. and Giesy, J. P. (2008). Novel trends in endocrine disruptor testing: the H295R Steroidogenesis Assay for identification of inducers and inhibitors of hormone production. *Anal Bioanal Chem* **390**, 287-91, 10.1007/s00216-007-1657-5.

Hecker, M., Hollert, H., Cooper, R., Vinggaard, A. M., Akahori, Y., Murphy, M., Nellemann, C., Higley, E., Newsted, J., Laskey, J., Buckalew, A., Grund, S., Maletz, S., Giesy, J. and Timm, G. (2011). The OECD validation program of the H295R steroidogenesis assay: Phase 3. Final inter-laboratory validation study. *Environ Sci Pollut Res Int* **18**, 503-15, 10.1007/s11356-010-0396-x.

Hoyer, P. B. (2010). Female Reproductive Toxicology. pp. 339-345. Elsevier, Oxford.

Judson, R. S., Kavlock, R. J., Setzer, R. W., 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**, 451-62, 10.1021/tx100428e.

Perreault, S. D., Klinefelter G.R., Clegg E. (2008). Assessment of Male Reproductive Toxicity. Informa Healthcare, USA.

Rotroff, D. M., Wetmore, B. A., Dix, D. J., Ferguson, S. S., Clewell, H. J., Houck, K. A., LeCluyse, E. L., Andersen, M. E., Judson, R. S., Smith, C. M., Sochaski, M. A., Kavlock, R. J., Boellmann, F., Martin, M. T., Reif, D. M., Wambaugh, J. F. and Thomas, R. S. (2010). Incorporating Human Dosimetry and Exposure into High-Throughput In Vitro Toxicity Screening. *Toxicol Sci* **117**, 348-358, DOI 10.1093/toxsci/kfq220.

Russom, C. L. (2002). Mining environmental toxicology information: web resources. *Toxicology* **173**, 75-88, S0300483X02000239 [pii].

Watanabe, K. H., Li, Z., Kroll, K. J., Villeneuve, D. L., Garcia-Reyero, N., Orlando, E. F., Sepulveda, M. S., Collette, T. W., Ekman, D. R., Ankley, G. T. and Denslow, N. D. (2009). A computational model of the hypothalamic-pituitary-gonadal axis in male fathead minnows exposed to 17alpha-ethinylestradiol and 17beta-estradiol. *Toxicol Sci* **109**, 180-92, kfp069 [pii]
10.1093/toxsci/kfp069.