PROJECT SUMMARY/ABSTRACT

This project is a proof of principle of *systems toxicology*, a new approach to chemical safety evaluation that integrates molecular, cellular, and physiological data in the context of a genetically diverse animal model to develop testable hypotheses about the key molecular events leading to adverse outcomes following chemical exposure. The project aims to capitalize on the potential of two powerful population-based model organism resources, the Collaborative Cross (CC) and Diversity Outbred (DO) mice, to study the role of genetics in conferring susceptibility to chemical exposures. Through an integrated set of experiments using arsenic exposure in mice and cell lines, the molecular genetic basis of toxicological responses will be evaluated. This project will test the hypothesis that genetic analysis in the context of a quantitative environmental perturbation will reveal multiple, novel, and diverse biochemical networks that respond to chemical exposure. The proposed integrated set of experiments will enable the discovery and validation of *adverse outcome pathways* through three specific aims. **Aim 1** will evaluate study designs for animal testing with genetically diverse DO mice including sample size requirements for toxicity evaluation. G x E genetic loci will be mapped and incorporated into predictive computational models, and testable hypotheses will be proposed for validation. **Aim 2** will conduct a parallel, population-level arsenic exposure study of *in vitro* primary cell cultures to identify genetic factors underlying susceptibility and resistance using physiologically informative cellular phenotypes. The data generated in the *in vitro* arsenic exposure study will allow determination of the extent to which cytotoxicity, genotoxicity, and oxidative stress in cellular assays are physiologically informative for the discovery of molecular pathways that drive susceptibility and/or response in the whole organism. **Aim 3** will identify key mechanisms in renal arsenic toxicity. This study will generate a model for the effect of arsenic exposure on the kidney to predict outcomes that are contingent on genetic background. Collectively, this new approach to toxicology using DO mice will address fundamental biological questions by combining chemical interventions with genetic variation. It will establish causal pathways across multiple levels of molecular and physiological outcomes to yield results with relevance to clinical translation.

PROJECT NARRATIVE

Regulatory toxicology seeks to create safety thresholds for chemical exposure in humans based on experimental studies in animals, but results of such studies may not accurately predict human sensitivity because they fail to accommodate the genetic diversity that exists across human populations. We will use population-based, genetically diverse mice to study the complex interplay between genetic variation and environmental factors that determine cellular and organismal responses to arsenic exposure. Through a novel statistical analysis of our data, we will account for individual genetic variation and provide a data-driven model that can be translated to risk assessment for human chemical exposure.

SPECIFIC AIMS

In response to RFA-ES-17-009, our overall goal is to demonstrate proof of principle for a *systems toxicology* approach, using a novel gene-by-environment (G x E) genetic mapping strategy, to identify genes responsible for individual variation in response to arsenic (As) exposure. Arsenic is a widespread environmental pollutant that is present in soil and groundwater. Exposure to As in drinking water imposes a significant burden on human health affecting hundreds of millions of people worldwide. Health outcomes related to As exposure in humans are complex and variable. Genetic variation is known to affect As metabolism and can be causally linked to adverse outcomes including cancer. While As toxicity has been extensively studied, many questions remain unanswered. This makes As an ideal model compound for developing new strategies for evaluation that can be applied more broadly. We propose to carry out an integrated set of experiments using As exposure in a genetically diverse mouse population (Diversity Outbred, DO) to evaluate study designs and develop new methods for safety assessment that can be broadly applied. We will directly compare findings obtained through *in vivo* and *in vitro* screens on the same animal genotypes. We will focus in-depth analysis on the kidney – a major target of As-induced damage – using a combination of clinical and molecular assays to identify biomarkers of exposure and tissue damage and to uncover mechanism that lead to adverse health outcomes. We will employ a novel validation strategy to verify the genetic- and biomarker-based predictions.

AIM 1. Develop new statistical methods for toxicity evaluation in a genetically diverse mouse model. Current experimental approaches to toxicity evaluation use genetically uniform animal models that do not reflect the range of variation in individual susceptibility seen in human populations. With the goal of establishing new models for testing, we will evaluate experimental design and sample size requirements for genetically diverse DO mice. We will develop new methods for dose-response and benchmark dose estimation that incorporate individual variation. This will lead to improved estimation of exposure risk in experimental animals and reliable extrapolation to risk in humans. We will map and identify genetic loci that are responsible for individual variation in susceptibility in DO mice using a novel G x E mapping strategy. We will incorporate omics data to develop computational models of the pathways that lead to adverse health outcomes. This will lead to the identification of predictive biomarkers and key molecular events that are potential targets for intervention and treatment in humans. We will develop methods to select animals and cell lines with specific genotypes in order to validate computational predictions.

AIM 2. Conduct a parallel, population-level arsenic exposure study *in vitro* to identify genetic pathways underlying susceptibility and resistance using physiologically informative cellular phenotypes. Genetically diverse panels of mouse cells offer scalability and economy for high throughput screening, capture the intrinsic factors that influence susceptibility, and are easily controlled in a mouse study (genetic variability, sex, age/life stage). We will create a large panel of primary fibroblast cultures from the same cohort of animals used for *in vivo* exposure. We will expose these cells to As using the same dose series as the *in vivo* experiment and we will generate quantitative measures of sensitivity and resistance using assays of cellular endpoints that are known to correlate with established modes of action for *in vivo* As toxicity.

AIM 3. Identify key mechanisms in renal arsenic toxicity using systems toxicology. Arsenic exposure leads to damage and dysfunction in multiple organs. One of the cell types with the highest accumulation and subsequent negative impact is the kidney proximal tubule cell (PTC). Therefore, the kidney is an excellent organ to study in order to identify key mechanisms of *in vivo* As toxicity. We will evaluate PTC function and kidney damage in the full cohort of 640 DO mice. We will include measurements for genotoxicity to enable direct comparison of our *in vivo* study with the *in vitro* study of Aim 2. We will select a subset of 240 DO mice from the most informative exposure groups for metabolomics, RNAseq, and DNA methylation assays. Using a systems-genetics approach we will develop models of *in vivo* As toxicity. Our findings will be validated in an independent cohort of outbred mice with reproducible genotypes.

IMPACT: Our study will provide new insights into the genetically determined range of As toxicity. It will serve as a paradigm for *systems toxicology* using compounds as intervention in combination with genetic variation to develop mechanistic hypotheses linking key molecular events to clinical outcomes.

RESEARCH STRATEGY

A. SIGNIFICANCE

New approaches to risk assessment will focus on mechanistic models of adverse outcomes. Regulatory science is shifting from an emphasis on endpoint testing to a mechanistically based approach of constructing adverse outcome pathways (AOPs) for toxicity[2]. An AOP is a model that describes a sequence of molecular and cellular changes that follow exposure to a toxic substance and ultimately result in illness or injury. Our efforts will leverage quantitative genetics to discover and build predictive computational models of AOPs. This will lead to future chemical safety reporting that is based on a mechanistic understanding of the biological processes that are perturbed by chemical exposure. The anticipated impact is a future regulatory paradigm that is consistent with the accepted framework of integrated approaches for testing and assessment, yet factors quantitative models of individual variation into the prediction of harmful effects of chemicals[3].

Arsenic is an ideal model compound for developing new approaches to toxicology. Inorganic arsenic (As) is an environmental contaminant that affects over 100 million people world-wide[4]. Most people are exposed to As via ingestion from drinking water and chronic exposure can lead to diseases of the skin, kidney, heart, and nerves, as well as cancer[5, 6]. The current As exposure limit for drinking water in the US is 10 μg/L[7]. While this level was chosen in part because of limitations in the ability of water treatments to achieve lower As levels, it is also a product of endpoint-based regulatory practices used to derive exposure thresholds for many environmental contaminants[8]. Many questions relating to As toxicity can best be addressed by *in vivo* exposure studies in model organisms. *In vivo* whole animal models of As exposure have shown evidence of carcinogenicity and damage to multiple organ systems including kidney. Adverse outcomes following As exposure in the mouse strongly parallel those that occur in humans, indicating that whole animal studies using mice can provide critical insight into the mechanisms of As-induced disease in humans[9].

The Diversity Outbred mice are an ideal reference population for systems toxicology. The majority of toxicity studies are conducted in a single inbred or isogenic strain. New study designs that introduce genetic variability into animal models are needed to estimate the extent of variation that can be expected in human populations. By design, isogenic strains lack genetic variation and thus will not reflect the range of genetically determined susceptibility present in a human population[10]. We propose to evaluate the utility of a unique, genetically diverse population of laboratory mice known as the Diversity Outbred (DO) population. DO mice carry over 52 million known genetic variants that intersect with 92% of exons and influence transcription for the majority of expressed genes[11]. Broad genetic diversity makes the DO a powerful genetic resource for systems biology and toxicology because we can combine genetic variants with 'omics' data to establish causal hypotheses linking genetic variation to the transcripts, proteins and metabolites that mediate clinical outcomes. We have successfully used DO mice in toxicity evaluations to identify genes that drive chemotherapy-induced myelosuppression and benzene-induced genotoxicity[12, 13]. In the proposed project, we will use DO mice to identify genes responsible for individual variation in response to As exposure using a novel gene-by-environment (G x E) genetic mapping strategy.

Another powerful advantage of conducting studies with DO mice is their relationship to the Collaborative Cross (CC) inbred strain panel. DO mice were derived from the CC progenitors and thus they share the same genetic polymorphisms. While there are too few CC strains to carry out G x E mapping, they provide a resource of reproducible genotypes for validation and mechanistic studies. Rather than use the CC strains directly – many breed poorly and are not robust – we propose to use F1 hybrid mice from pairs of CC strains (recombinant inbred intercross, RIX mice). RIX mice have similar levels of outbreeding to the DO and yet the same outbred RIX genotype can be replicated across many animals. There are thousands of possible RIX combinations, and they can be selected to have specific genotypes at multiple loci of interest. We propose to use RIX mice to validate computational predictions from AOPs that we discover in DO mice.

New analytical methods are needed to account for genetic diversity in setting safe exposure limits. Current methods for establishing safe exposure levels employ benchmark dose (BMD) estimation, which fits a dose-response (D-R) curve to data from isogenic animal experiments to obtain statistical confidence bounds for the BMD[14, 15]. Thresholds for human exposures are established by applying an ad hoc *adjustment factor* that lowers the BMD by 100-fold[16, 17]. The adjustment factor creates a buffer to account for inherent differences in susceptibility between the animal test population and a human population, as well as individual variation in sensitivity that exists within the genetically diverse human population[16, 17]. The arbitrary nature of the adjustment factor is problematic because an underestimated factor could put exposed populations at undue risk, while an overestimated factor could incur unnecessary regulatory costs to mitigate exposure that is

not dangerous, as well as create undue anxiety in a population that perceives a risk that does not actually exist. This warrants development of improved methods for determining exposure thresholds that can rationally account for genetic variation in susceptibility. We will develop and evaluate new statistical methods for BMD estimation using an empirical-Bayes framework that accounts for individual genetic variation and provides a data-driven approach to estimating adjustment factors[18, 19].

A parallel in vitro screen for integrated systems toxicology. *In vitro* toxicity testing can be used to assess sensitivity and susceptibility for compounds that act through cell-autonomous mechanisms thereby offering *in vitro* endpoints that correlate well with *in vivo*, tissue-level and organismal effects (disease). Therefore, it is theoretically possible that *in vitro–in vivo* correlates could be used for the discovery of G x E interactions. In a whole animal, a compound may be metabolized in the liver, transported to other organs, and undergo further metabolism prior to tissue injury. Thus, *in vitro* correlates may not exist or may be underdeveloped for many compounds. For inorganic As, its five metabolites have toxic effects *in vivo* and *in vitro*, and the correlation between *in vitro* and *in vivo* biomarkers of As toxicity has been demonstrated[20-22]. Arsenic is a known group I carcinogen associated with tumors of the lung, skin, bladder, liver and kidney. The genotoxic effects of As are the indirect result of DNA repair inhibition and increased oxidative stress, and include increased, unrepaired DNA breaks, micronuclei, aberrant sister chromatid exchange, and chromosome breakage / missegregation (aneuploidy). These genotoxic lesions lead to cytotoxicity and tumorigenesis in target tissues[5, 23]. Given these prior correlations, the As model offers a unique opportunity to use established *in vitro–in vivo* exposure correlates for parallel G x E discovery. We have selected three well-documented and interrelated endpoints that have considerable data to support a strong *in vitro–in vivo* correlation between cellular phenotypes and whole-animal physiological endpoints: genotoxicity, cytotoxicity, and oxidative stress. We will create a panel of primary cells from the same population of mice exposed to As in the *in vivo* arm of our study and we will use these correlates to determine effects of genetics on inter-individual phenotypic variation *in vitro*. By integrating these data with our *in vivo* data, we will identify shared pathways that determine inter-individual sensitivity.

The kidney is a model organ system for defining adverse outcome pathways. We will focus on the kidney for an in-depth study to identify genetic factors that determine metabolic changes due to As toxicity and derive AOPs**.** Once such factors are identified, we can validate whether they are organ-specific or drive As sensitivity in other organs. Because the kidney receives a high blood flow, it experiences a high level of exposure, making it particularly susceptible to chronic low levels of As in the blood. The cells of the proximal tubule (PTCs) are sensitive due to their high reabsorptive activity and anatomical position as the first renal epithelial cell to be exposed to filtered toxicants[24, 25]. The presence of transport systems that readily facilitate uptake and accumulation of chemicals within these cells as well as the large number of bioactivation enzymes that can generate reactive chemical species also make PTCs particularly susceptible[26]. Currently, it is not clear how changes in metabolism due to As lead to renal injury or how As metabolism is altered as renal injury progresses. There is a great need to understand the relationships between biomarkers and specific types of tissue injury[27]. Metabolic biomarkers in urine and blood can be evaluated as screening tools for exposure levels and As-induced organ damage that can be extrapolated to humans. We will use molecular omics data to identify AOPs that reveal mechanisms of *both* As transport *into PTCs and As metabolism within these cells*. Our findings will be applicable to As exposure in humans through the identification of candidate biomarkers and prioritization of the most relevant AOPs for risk assessment.

Systems toxicology – a powerful new discovery tool for fundamental biology. Genetics, as a discipline, achieves scientific insight by perturbing a system, *i.e.,* a developing organism, via the deletion or alteration of the function of a gene. With the addition and integration of high-throughput molecular biology technologies and next-generation statistical modeling, it is now within our reach to draw biological insights from the perturbation of entire genetic and biochemical networks. Hence, toxicology, which perturbs a system via one or more quantitative chemical interventions, is poised to become a fundamental new approach to biological sciences in the twenty-first century. The role of the CC and DO mice in this transformation will be as “model populations” across which primary data and concepts can be integrated and tested. By building a shared knowledgebase spanning multiple chemical interventions and species we can address both human and broader ecological issues (see **letter from Dr. Joseph Shaw**). Our project will serve as a proof of principle study for *systems toxicology* - an approach to biological inquiry that addresses fundamental mechanistic questions by using chemical interventions together with genetic variation to establish causal pathways across multiple levels of molecular, cellular, organ-specific, and whole-animal physiological outcomes.

B. INNOVATION

This project offers both methodological and conceptual innovations. We highlight four innovative aspects:

i) New approaches to risk assessment will incorporate genetic diversity. Introducing structured genetic diversity and developing new analysis methods will pave the way for regulatory changes, move away from arbitrary adjustment factors and toward data-driven and AOP-based evaluation of uncertainty. Our Bayesian inference methods will frame conclusions in terms of intuitive probability statements – avoiding the awkward and frequently misinterpreted logic of classical confidence intervals and p-values – to address multiple sources of uncertainty due to individual variation, statistical estimation error, and model selection.

ii) Parallel in vitro and in vivo screening will assess translatability. We will evaluate the power and validity of *in vitro* screening with genetically diverse cell-based systems. There is no direct evidence for connecting cell lines to whole animal results for systems toxicology. However, we will test the validity at a large scale using genetic diversity. As we cannot directly validate human cell lines using such an approach, the outcome of this study will be important for justifying use of genetically diverse human cell lines in toxicity evaluations.

iii) Employ novel validation strategies. The DO mouse and its inbred complement, the CC mouse, present a unique opportunity to validate findings through the creation of customizable and reproducible outbred genomes by crossing CC recombinant inbred strains (RIX). This RIX validation strategy has been discussed, but not yet fully realized. Through our study, we will set new standards and approaches for validation of causal hypotheses that emerge from mining large-scale data. In addition, we propose a cell-line validation strategy that is unique in that genetic modifications will be applied to multiple genetically diverse backgrounds, in contrast to modifying a single genetic background. These validation strategies will improve generalizability of findings and increase the likelihood of successful translation to humans.

iv) Introduce systems toxicology as an approach to study fundamental biology. Using the kidney as a model organ system with functional readouts over time through blood, urine, and ultimately tissue, we can identify chemically-induced molecular-level interactions, known as the Molecular Initiating Events (MIEs), and a series of higher order (molecular, cellular, tissue) Key Events that through a sequence of perturbations ultimately lead to an adverse outcome at the level of an individual. Omics technologies enable the responses of tens of thousands of genes and many of their eventual products (*i.e.,* metabolites) to be measured, thereby facilitating the discovery of molecular mechanisms within AOPs[28, 29]. Indeed omics-derived knowledge has been at the foundation of the derivation of some novel AOPs[30].

C. APPROACH

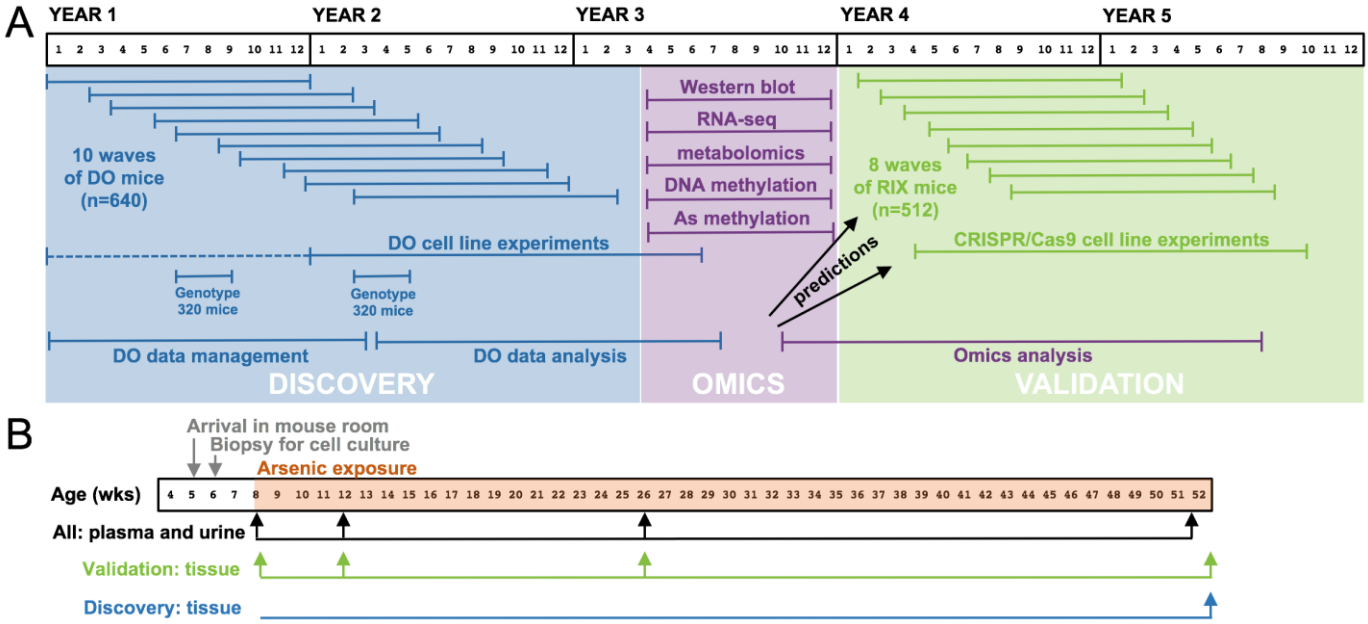

Figure 1. (**A**) Overview of the three phases of the proposed project. (**B**) Schedule of procedures and sample collection for each DO and RIX mouse.

Overview of Experimental Approach. The proposed work will proceed through discovery and validation phases, each with *in vivo* and *in vitro* components, and an omics data collection phase (**Figure 1**). Together, these experiments will address the three Specific Aims and will provide resources for future investigations.

The *in vivo* component of the *discovery phase.* 640 DO mice (320 males and 320 females) will be exposed to eight concentrations of sodium arsenite in drinking water. Mice will be randomized to exposure groups across a series of 10 birth-cohorts (waves) of 64 mice each. In total, across all waves, there will be 80 mice in each exposure group. Exposure will begin at eight weeks of age and mice will be evaluated for health status and weighed weekly during the course of treatment. Though we anticipate that no mice will experience such negative health effects given our careful selection of low exposure levels, animals that show >15% decrease in body weight or signs of severe suffering will be euthanized. At 8 weeks (pre-exposure) and again at 12, 26 and 52 weeks of age, plasma, urine, and fecal pellets will be collected. At 52 weeks of age, animals will be euthanized, the left kidney will be collected for histology and the right kidney will be flash frozen and homogenized. Liver, heart, brain, and eyes will be collected and stored for future studies.

Gene expression and metabolomics analysis. We will select a subset of 240 mice representing the range of exposures that are most informative for genetic mapping based on kidney function measurements. Specifically, we will select a high exposure group, in which the majority of animals are affected, and a low exposure group, in which few animals are severely affected, and one intermediate exposure group. We will obtain genome-wide transcriptome data by RNA-seq and metabolomics data by mass-spectrometry while we retain sufficient kidney tissue in aliquots suitable for validation experiments and additional analyses, such as genome-wide methylation analysis, should the opportunity arise.

The *in vitro* component of the *discovery phase.* We will establish primary fibroblast cell lines from each of the 640 mice that will be exposed to As in the discovery phase. We will expose 240 of these lines, corresponding to the selection of mice for omics analysis (above). These primary cell lines be evaluated for established modes of action for As toxicity *in vivo*, and are known to correlate well *in vivo* – cytotoxicity, genotoxicity, and oxidative stress. We will use these quantitative measurements of *in vitro* toxicity to identify genetic factors associated with As sensitivity and resistance, and we will identify candidate loci for validation.

The *in vivo* component of the *validation phase*: Analysis of discovery phase data will identify loci affecting sensitivity to adverse renal outcomes following As exposure. In the *in vivo validation phase*, we will prioritize these findings to select pairs of CC inbred strains to produce RIX cohorts. The RIX mice will carry specific combinations of alleles, perhaps at multiple loci, predicted to produce either sensitive or resistant genotypes by perturbing specific AOPs. Each cohort of 64 RIX mice (32 male and 32 female) will be exposed to As in drinking water at concentrations that elicited a robust response in genetically similar DO mice. We will conduct the same phenotyping as in the discovery phase. In addition, we will collect cross-sectional tissue samples from a subset of mice at 4 time points for multi-omics analysis. Crucially, this will enable us to examine the changes in gene expression and metabolism over time with the aim of distinguishing early from later-acting key events within the adverse outcome pathways and to identify the molecular initiating events.

The *in vitro* component of the *validation phase:* Genetic analysis of discovery phase fibroblast data will identify loci that affect sensitivity to genotoxicity and oxidative stress. To determine the impact of gene variants on sensitivity or resistance to As exposure, we will employ CRISPR/Cas9 technology to genetically modify individual genotypes within our collection of DO fibroblasts. This approach will allow us to experimentally measure the effects of candidate susceptibility or resistance variants across any subset of DO genotypes with varying exposure responses. CRISPR/Cas9-edited and unedited, genotype matched control DO fibroblasts will be challenged with As and screened for cytotoxic, genotoxic, or oxidative stress responses. Because our collection of DO fibroblasts replicates the set of individual genotypes in our screen, our validation approach allows for direct manipulation and testing of predicted epistatic interactions in the context of G x E responses.

Rationale for exposure levels, period of exposure, and sample size in the Discovery cohort.

Dosage: The OECD Guidelines for the Testing of Chemicals recommends the use of at least three exposure levels and a concurrent control[31]. The highest exposure level should induce toxicity but not death or severe suffering and the lowest dose should aim to achieve no observed adverse effects. The EPA standard for As in drinking water is set at 10 ppb and was revised downward from 50 ppb in 2001[32]. We wanted to include both the new and old EPA standards and an exposure lower than these in the event that the 10 ppb standard might prove toxic to exceptionally sensitive individuals over the extended 10 months exposure period. Published short-term studies of mice exposed to 100 pm and higher showed body weight loss, anemia, and severe suffering[33]. Thus, our maximal exposure is 10 ppm. We wanted to include exposure levels in the range of typical human exposures in regions with high As levels in drinking water [0.01 to 0.1 ppm;[34]]. To include as many exposures as are practical to manage we chose the 8-exposure series (0, 0.005, 0.01, 0.05, 0.1, 0.5, 1 and 10 ppm).

While we can control the amount of As in the drinking water, we recognize that rodent diets often contain low levels of As. This occurs in the human population as well[35, 36]. The NTP uses the NTP2000 diet (Zeigler Brothers, Inc., Gardners, PA) in their two-year toxicity and carcinogenicity studies[37]. Each batch of this diet is assayed for a range of contaminants, including As, which must be below 0.5 ppm. We requested the certificate of analysis for the Dec. 2017 batch of NTP2000 and the As concentration was 0.236 ppm. We are continuing to investigate other diets with lower arsenic concentrations to minimize arsenic sources beyond what we supply in the drinking water. We will assay each batch for As and other environmental contaminants.

Exposure period: OECD guidelines recommend starting exposure prior to 9 weeks of age (in mice) and an exposure period of 90 days[31]. We wish to evaluate whether the 90-day exposure recommendation is adequate to predict the effects of chronic low-level exposure. Long-term chronic exposure is characteristic of human exposures and the emergence of kidney disease and genotoxicity *in vivo* (*e.g.,* cancer) are long-term endpoints. We selected a 10 months exposure because beyond the age of 1 year, we can expect to see increased rate of death by natural causes in DO mice (Churchill, unpublished data).

Sample size: Multiple considerations factored into the sample sizes proposed for different phases of the study (*e.g.,* **see Vertebrate Animals**). One of the goals of Aim 1 is to determine the required sample sizes for toxicity evaluation with DO mice – therefore we designed a study that exceeds the typical standards with the goal of obtaining data that can be down-sampled to empirically evaluate statistical power at smaller sample sizes. OECD guidelines recommend at least 10 male and 10 female animals per group. An additional consideration, relevant to genetically diverse mice, is to ensure that we observe at least a few animals from exceptional subpopulations. We evaluated multiple scenarios for genetic risk. For example, a monogenic recessive locus present at allele frequency of ¼ (two founders) is expected to arise in four or more individuals in a group of 80 with probability 0.74. While one can construct scenarios that we are unlikely to detect, using 80 animals per group will enable us to observe rare subgroups of highly susceptible animals. For omics analysis, 240 animals will be selected from the most informative exposure groups (discussed above). Based on our experience with omics traits in DO mice, this will provide sufficient power for mapping[38].

Handling and Disposal of Arsenic: JAX is committed to the health and safety of its employees through the policies described in **Facilities and Resources** ("Disposal and Handling of Arsenic and Arsenic-containing Materials") and the **letter from Michael Watkins**, manager of JAX EHSS.

AIM 1. Develop new statistical methods for toxicity evaluation in a genetically diverse mouse model.

Rationale: The use of animal testing data to establish human safety standards entails a high degree of uncertainty, which is currently addressed by the application of adjustment factors[16]. This uncertainty derives from three sources. **Variability** between individual humans arises from genetic makeup, sex, age, diet, and other environmental exposures. This variation, which is intrinsic to humans, is largely controlled in isogenic animal experiments and thus necessitates the intra-species component of the adjustment factor. **Statistical uncertainty** can be reduced by collecting more data but we have to be cognizant of costs and the ethical use of animals[39]. **Extrapolation** from an animal model to humans, which necessitates the inter-species component of the adjustment factor, is best addressed through better understanding of the molecular basis of inter-species differences in the genes and enzymes involved in AOPs[28]. Aim 1 will address these challenges by evaluating study designs that incorporate genetic diversity and by developing new methods to discover and validate AOPs using the data generated in Aims 2 and 3.

Establish standards for power and sample size in toxicity evaluations using DO mice. Our *in vivo* study of 640 DO mice will include clinical evaluations of kidney function and measurements of As-species in plasma and urine. The latter will be evaluated as measures of internal exposure. The *in vitro* study of cell lines from the same animals provides an opportunity to observe many individual genotypes across multiple exposure levels. Outcomes will include measures of cytotoxicity and genotoxicity. We will test all measured outcomes for D-R trends and to identify sex differences. We will obtain genotypes to estimate kinship and heritability of all outcomes.

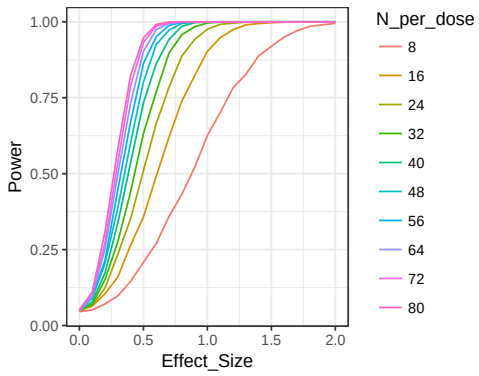

Figure 2: Power to detect a linear trend with dose. We simulated dose-response data across eight exposure levels and estimated power using a linear trend test. Effect size is the difference in response between the lowest and highest exposure levels expressed in standard deviation (SD) units. The SD reflects all sources of variance including genetic variance. Sample sizes range from 8 to 80 animals per group.

We will estimate variance components including the contributions of co-housing, wave, assay batch, sex, body weight, and residual error. Evaluation of these sources of variance across a wide range of outcomes will provide a resource of information to inform future study design and sample size determinations for DO experiments (**Figure 2**). Our study data will be down-sampled to empirically evaluate sample size recommendations.

Develop new method for benchmark dose estimation. Dose-response curves relate exposure (dose) to quantitative outcomes. A variety of shapes can be obtained by fitting the 4-parameter family of D-R curves $y = a[c-(c-1)\exp(x/b)^d$ [(40)]. The model parameters have natural interpretations: *a* is baseline response, *b* is sensitivity to dose, *c* is the maximum effect over baseline, and *d* is the curvature. The models are nested such that setting *a*, *c*, or *d* to zero or *b* to infinity produces different families of curves. Classical approaches to D-R modeling apply nested hypothesis testing to select a best model; newer methods use model averaging to account for uncertainty about the shape of the D-R curve[(41, 42)]. We will develop empirical-Bayes analysis methods that share advantages of model averaging while allowing us to evaluate the importance of each model parameter. Empirical-Bayes analysis of complex, non-linear regression models has only recently become practical due to development of software and modeling languages such as STAN[(43)]. Using STAN, we will develop hierarchical models that define the D-R parameters as a function of sex, genotype, or other covariates to determine if an outcome has a sex-specific baseline or a genotype-dependent sensitivity. We will advocate for the adoption of Bayesian predictive intervals as reporting standards for quantifying uncertainty in risk assessment (**Figure 3**). Models implemented in STAN can address challenging features of data such as variance heterogeneity (heteroscedasticity) with relative ease compared to classical statistical modeling. We will compare Bayes estimates to current standards for BMD estimation using BMDL 2.7 software[(14)].

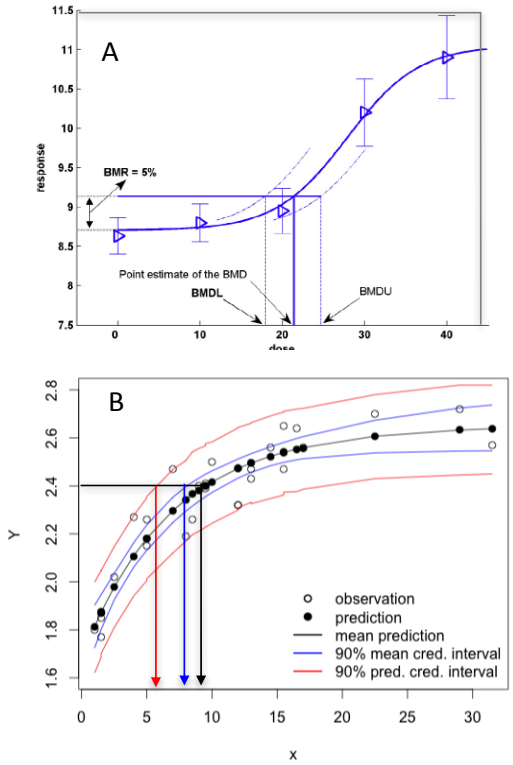

Figure 3: Classical and Bayesian approaches to BMD estimation. (**A**) Classical benchmark dose estimation is illustrated schematically (1). A benchmark response (BMR) level is defined as a percentage increase over baseline. Then a point estimate of the benchmark dose (BMD) is obtained by finding the point on the x-axis where the height of the dose-response curve meets the BMR. Statistical variation around the fitted dose-response curve is taken into account to obtain a lower confidence bounds (BMDL) for the BMD. The BMDL is an estimate of the dose at which we are confident that the ***average response*** of the population will fall below the BMR. Classical BMD estimation does not account for individual genetic variability. (**B**) To illustrate the idea behind Bayesian benchmark dose estimation, we consider a growth curve fitted to data using STAN software (https://www.r-bloggers.com/non-linear-growth-curves-with-stan/). Assuming the BMR is 2.4 (in y-axis units), we obtain a point estimate of BMD (black) as before; a lower credible bound (blue), which is analogous to the BMDL, accounts for statistical estimation error; and a predictive credible bound (red) accounts for individual variability, including genetic variation in the study population. The predictive bound is an estimate of the dose at which we are confident that 90% of ***individual responses*** will fall below the BMR.

Map G x E loci to discover genes involved in AOPs. Genetic mapping of G x E interaction is analogous to the problem of quantitative trait locus (QTL) mapping with interacting covariates[(44)]. Our goal is to identify genetic loci with polymorphic alleles that change the shape of the dose-response relationship. To establish that a locus is a G x E QTL, we fit a nested series of models relating the outcome to exposure and to genotype. We apply statistical tests to establish that 1) there is a trend with increasing exposure, 2) there is a genotype effect (*i.e.,* an additive QTL), and 3) there is a genotype by exposure interaction. The defining character of a G x E QTL is that the effects of exposure will differ depending on the genotype. We will map G x E loci by testing at each locus in a genome-wide scan and applying a multiple-testing adjustment using permutation analysis[(45)]. We will initially use a linear trend test using R/qtl2 software[(46)]. Next, we will introduce genetic effects into D-R modeling (above) to determine which aspects of the D-R are under genetic control, *e.g.,* sensitivity QTL vs. maximal responses QTL. Genes at these loci will be evaluated for inclusion in AOPs.

Apply mediation analysis to evaluate causal AOPs. Bayesian Networks (BNs) are graphical representations of causal hypotheses involving both measured and latent (unobserved) variables[(47)]. BNs have been applied in expression QTL (eQTL) analysis where the local (or cis) eQTL acts as an instrumental variable to anchor and direct causal relationships[(48)]. An AOP can be expressed as a set of causal relationships among variables.

Therefore we can evaluate AOPs using the formal inference structure of BNs. To illustrate, we consider a hypothetical example based on the following prior observations. Methylation of As occurs in liver and kidney, which are the most relevant tissues for metabolism and clearance of As[49, 50]. Differences between sexes and across human populations indicate the presence of functional polymorphisms in As methylation pathways[17]. Variants identified in human As methyltransferase (AS3MT) have been associated with As sensitivity[51, 52]. Arsenic exposure may also influence AS3MT expression. Returning to DO mice, *As3mt* gene expression has a strong local-eQTL (**Figure 4A,B**) in both kidney and liver– interestingly, the liver eQTL has different allelic effects (data not shown). We can construct a BN model to represent hypotheses about the role of *As3mt* in mediating kidney damage. We hypothesize that expression of *As3mt* drives ASMT activity, which in turn affects levels of kidney damage. We use methylated As in urine and KIM1 protein as biomarkers for the unobserved processes. We note that our model predicts that the biomarkers should have G x E QTL at the *As3mt* locus, because the effect of genotype on, e.g., KIM1 is dependent on As exposure. The specific causal hypothesis indicated by the first arrow in (**Figure 4C**) can be evaluated by fitting a regression model to any of the three outcomes and including *As3mt* expression as a covariate. If the hypothesis is true, the residual correlation of *As3mt genotype* with each outcome should fall to near zero after accounting for *As3mt* gene expression. This inference is possible despite the potential confounding effects of Arsenic exposure because of the properties of *As3mt* genotype as an instrumental variable[48]. Complete evaluation of a BN/AOP would involve additional steps and careful checks for data quality. We may reject, refine, or tentatively accept the model as part of an evolving AOP leading up to the final step – experimental validation.

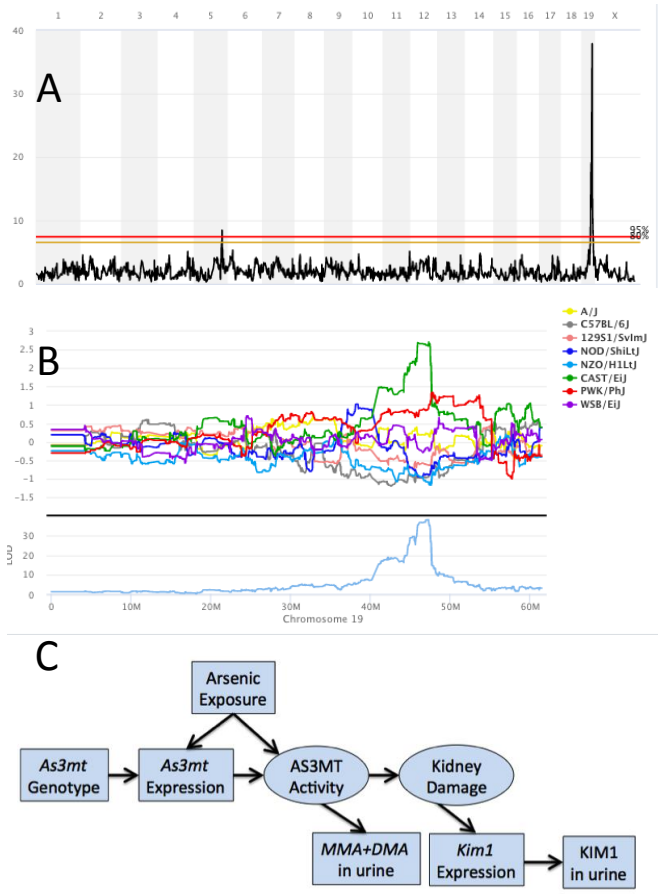

Figure 4: Using genetic mediation analysis to construct an AOP. Genetic mediation analysis starts with the identification of local gene expression QTL (eQTL) that co-locates with QTL for one or more outcomes of interest. In this hypothetical example, we consider *As3mt expression*, which has a significant eQTL in kidney (A) (Korstanje and Churchill, unpublished data). The additive effects of founder alleles (B) indicate that the CAST haplotype carries a high expressing allele. This pattern should also be observed in outcome variables *MMA+DMA in urine*, *Kim1 gene expression,* and *KIM1 in urine*. These outcomes are used as biomarkers for unobserved processes *As3mt activity* and *Kidney Damage*.

Predictive validation of genetic loci in a reproducible outbred mouse model. We will use G x E mapping and BNs to develop AOPs and derive testable hypotheses, *i.e.,* models that make specific predictions about the response to As exposure in animals or cells with known genotypes. To validate an AOP hypothesis *in vivo*, we propose to create RIX mice with specific genotypes at one or more loci of interest. These *in vivo* validation experiments are resource intensive and we will carefully consider the relevance to human health as a criterion for choosing which AOPs to validate. In contrast, validation using primary cell lines, for appropriate outcomes, can be carried out on a larger scale. We will use our DO fibroblasts, and as needed introduce genetic modifications using the CRISPR/Cas9 system. The latter will enable direct comparison of cells with and without specific modifications in one or more genetic backgrounds (Aim 2).

Expected results, potential pitfalls, and possible solutions. These experiments will deliver a substantial body of high quality data to inform future study designs. The development of computational AOPs requires that we identify G x E QTL for physiological or cell-based outcomes with relevance to human health. While we can be confident that the omics data will yield such QTL, physiological data can require large sample size for QTL mapping – while 640 mice should be sufficient, we can develop BNs with suggestive QTL effects. It is important that the As exposures should hit the “sweet spot” where at least some, but not all, mice are showing adverse outcomes. To address this concern we will monitor the *in vivo* experiment throughout but specifically at the first 90-day evaluation. If the dose range appears to be too low or too high, we have an opportunity to adjust it at this stage with minimal impact on the overall experiment.

AIM 2. Conduct a parallel, population-level arsenic exposure study *in vitro* to identify genetic pathways underlying susceptibility and resistance using physiologically informative cellular phenotypes.

Rationale: It is now possible to generate panels of cell lines representing individual mouse genotypes from genetic reference populations. Such 'parallel' panels capture the intrinsic factors of genetic variability, sex, age/life stage, etc., that determine organismal susceptibility to exposure and that are easily controlled in a mouse study, while offering the scalability and economy of an *in vitro* system. Even for *in vitro* outcomes that mirror known physiological endpoints at the organismal level (*e.g.,* cell proliferation → hyperplasia, or genotoxicity → genomic instability → cancer), it is not known to what extent *in vitro* G x E discovery will recapitulate *in vivo* G x E discovery. The goal of Aim 2 is to create a panel of 640 genetically unique primary fibroblast cultures from the full Discovery cohort of DO mice, and to perform a parallel As dose response study in a subset of this panel. In addition to providing a panel of cells that are genetically identical to the mice our Discovery cohort, primary fibroblasts offer distinct advantages over transformed cell lines, including a closer resemblance to the source tissues, and low mutation burden. We will expose this panel of cells to the same range of As doses that will be provided to the animals and we will employ a series of well-established assays to assess cytotoxicity, genotoxicity, and oxidative stress in exposed cells. These three endpoints represent established modes of action for As toxicity *in vivo*, are known to correlate well *in vivo* – and under comparable acute As exposures – and they are particularly relevant to kidney PTC responses to As exposure *in vitro* and *in vivo*[53].

Preliminary data: To inform relevant As dosing ranges for our *in vitro* exposure, we looked to previous studies that explored the cytotoxic and genotoxic effects of inorganic As and its five metabolites on transformed cell lines (*i.e.,* HEK293, CHO cells) at dose ranges similar to those proposed in our mouse study[21, 23, 54]. Dopp et al. showed that inorganic trivalent As and dimethylated trivalent As are the As forms are most cytotoxic to cultured fibroblasts, with 100% cytotoxicity at 100 μM and 10 μM, respectively, after 24 hours of exposure in culture media. Moreover, inorganic trivalent As genotoxicity was detectable at doses as low as 0.5 μM, demonstrating that genotoxic effects of As can be detected at sub-cytotoxic doses. However, to our knowledge, there are no published dose response data for arsenite in *primary* fibroblast cultures from laboratory mice. Since primary fibroblasts have significantly different growth and metabolic properties compared to transformed cell lines, we tested the viability of primary fibroblasts in the context of our proposed arsenite dose series. We found that unlike transformed human and rodent cell lines, significant loss of viability occurred at doses as low as 0.5 ppm (~4 μM arsenite), and there were significant differences in response between the outbred and inbred fibroblasts (**Figure 5**). Our data suggest that at least a subset of the DO fibroblasts will be more resistant to arsenite. Additionally, our data show that the proposed range of doses provides sufficient granularity at sub-cytotoxic doses for our proposed *in vitro* study.

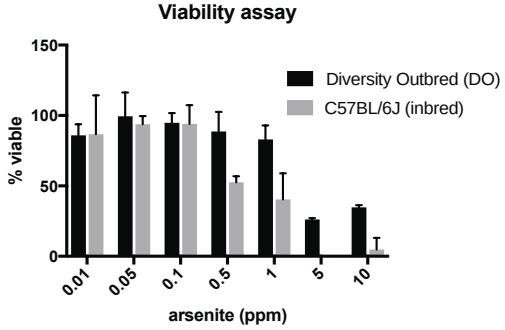

Figure 5. Primary fibroblast cultures derived from a single DO and a single C57BL/6J male were thawed, grown for 16 hours, and then exposed to a series of arsenite concentrations in culture medium for 24 hours. Post-exposure, viability was assessed using the LIVE/DEAD assay (ThermoFisher, L3224) and data were normalized to the 0 ppm (0 μM) group.

Experimental Methods:

<u>*Establishment of fibroblast cultures*</u>. We will establish primary fibroblast cultures from sterile tail biopsies obtained from the 640 DO mice in the Discovery cohort *prior* to As exposure. Using protocols that are routinely employed in the Reinholdt laboratory, tail biopsies will be cultured and fibroblasts expanded such that a minimum of $1x10^7$ cells is cryopreserved from each animal[55, 56]. In accordance with standard protocols in the Reinholdt laboratory for newly created primary cell lines, we will use low-density SNP genotyping to confirm that genotype of each line matches the genotype of the individual DO from which it was derived (thereby minimizing sample mix-ups) and we will confirm that each cell line is mycoplasma free (taking advantage of PCR-based testing services offered by the Diagnostic Services group within JAX Comparative Medicine and Quality – **see Facilities and Resources**).

<u>*In vitro exposure and screening*</u>. To promote randomization and to manage costs, *in vitro* As exposure and cellular assays will be conducted after all fibroblast cultures are archived. We will expose 240 of these cultures to As doses representing the range of exposures that are most informative for PTC response and injury in Aim 3. Cells will be thawed, plated in triplicate onto three 96-well plates, and exposed for 24 hours to our proposed

series of eight exposure levels of sodium arsenite in culture medium as shown in **Figure 6** (0, 0.01, 0.05, 0.1, 0.5, 1, 5 and 10 ppm). After exposure, cells will be harvested and assayed for cytotoxicity, genotoxicity, and oxidative stress. Cytotoxicity will be assessed via the percentage of viable cells using MTT viability assay (ThermoFisher, V13154), which is a colorimetric assay in which a tetrazolium dye is reduced to formazan via metabolic NAD(P)H-dependent oxidoreductase in viable cells. In this assay, the amount of dye conversion is directly proportional to proliferative capacity. An additional measure of cytotoxicity will be through TUNEL (terminal deoxynucleotidyl transferase dUTP nick end labeling) assays (ThermoFisher, C10245) which detect cells with apoptosis-driven DNA fragmentation. Genotoxicity will be assessed by assaying genomic DNA integrity via the single-cell DNA electrophoresis 'comet' assay (Trevigen, 4253-096). Unlike other assays for genotoxicity, the comet assay is inexpensive and detects a wide variety of DNA lesions (double-strand breaks, single-strand breaks, excision repair sites, and crosslinking sites). Oxidative stress will be assessed via a superoxide dismutase assay (ThermoFisher, EIASODC), which is an enzymatic colorimetric assay that detects the activity of the free radical scavenger superoxide dismutase (SOD). We have selected these assays on the basis of the following criteria: ability to multiplex in a 96-well plate (or array format), price per sample, published *in vitro–in vivo* correlation in toxicology studies, and an *in vivo* correlate in peripheral blood or in kidney PTCs (per Aim 3). The formats of all but the comet assay will allow for high-throughput data collection with a SpectraMax i3, multi-label fluorescent plate reader. Commercial comet assays are available in 96-well formats, and we will use a Nikon Ti2-E with automated image capture and high-content image analysis capability (available through the JAX Light Microscopy core). For quality control, each assay will include negative (untreated) and positive controls (*i.e.,* cisplatin treatment for genotoxicity assays, mitomycin C / DNAse I for cytotoxicity assays, and hydrogen peroxide for the oxidative stress assay). Additionally, fibroblasts will be tested within the first year to ensure robustness across a subset of variable genotypes and growth characteristics (*i.e.*, doubling time, see pitfalls below). For all assays, statistical analyses will be in accordance with established methods for each assay, and manufacturer's recommendations.

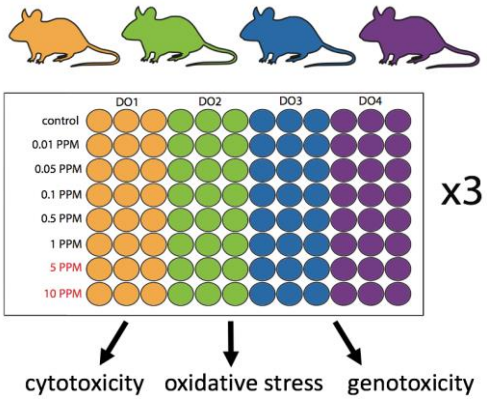

Figure 6. Experimental design for *in vitro* arsenic exposure. Adult fibroblast cultures plated in triplicate will be exposed to an eight-dose series of sodium arsenite for 24 hours and then assessed for cytotoxicity, genotoxicity, and oxidative stress.

<u>*CRISPR/Cas9 validation*</u>. Our genetic analyses in Aims 1 and 2 will reveal candidate genes and variants that are associated with sensitivity or resistance to As exposure. Candidate genes will be tested by simple CRISPR/Cas9 deletion of critical exons, while individual candidate variants will be tested by reciprocal CRISPR/Cas9 KI of SNPs that confer susceptibility into cell lines with otherwise resistant genotypes / phenotypes and vice versa. Since we will also have *in vivo* correlates for some of the *in vitro* endpoints (*e.g.,* TUNEL labeling of PTCs, and metabolomic data that could reveal oxidative stress protein signatures), we will prioritize candidates that are common across our *in vivo / in vitro* analyses. The Reinholdt laboratory has experience with the design and implementation of CRISPR/Cas9 engineering technology and uses CRISPR/Cas9-based validation routinely in their role as the validation core for the NIDA-funded JAX Center for Systems Neurogenetics of Addiction (Core C, **see Reinholdt biosketch**)[57]. Engineered cell lines will be tested for relevant endpoints as described above. CRISPR/Cas9 knock out (KO) of candidate genes or knock in (KI) of candidate variants will be confirmed and all potential off target sites (regions with three or fewer polymorphisms in the guide sequence) will be screened by PCR. The impact of KO of candidate genes or KI of candidate regulatory sequences will be confirmed by RT-PCR of the deleted gene or of nearby genes.

<u>*Expected results, potential pitfalls, and possible solutions*</u>. Our *in vitro* As exposure study will allow us to determine how measures of cytotoxicity, genotoxicity, and oxidative stress are associated with particular genotypes for the discovery of molecular pathways that drive susceptibility or resistance. Common pathways that emerge in our *in vitro* and *in vivo* studies will provide key benchmark data for application of cellular assays for toxicogenomic discovery, and will inform the extent to which these assays can supplant animal-intensive, *in vivo* genetic studies. Conversely, pathways that are uniquely discovered in our *in vitro* assays will support the value of parallelized *in vitro–in vivo* approaches as there may be exposure thresholds that are negative *in vivo* but positive *in vitro* in the context of chronic exposure (or vice versa).

Based on our previous experience culturing primary fibroblasts from DO mice, we expect there will be variability in growth characteristics (*e.g.,* doubling time) that may complicate screening. Therefore, as we

establish each of the cultures in Year 1, we will score them on the basis of population doubling time (# cells harvested / # cells plated / days in culture). Cohorts of cells will be plated on the basis of this score – grouping on the basis of doubling time – and screening will be staggered to avoid batch effects. If differences in growth rate remain a concern despite this approach, we will use growth rate scores as a co-variate in our genetic analyses.

For each of our endpoints (cytotoxicity, genotoxicity, and oxidative stress) there are several robust, widely used commercially available assay kits that can be employed, and we do not anticipate any difficulty in their application to our samples. We selected our specific assays on the basis of throughput, cost, and prior art, but we could switch commercial assay platforms if any issues arise. In some cases, we have introduced redundancy. For example, both MTT and TUNEL assays will be used to assess viability, however one assay provides a measure of proliferative capacity using viable cell counts (MTT) and the other, % cell death via apoptosis. Variability in growth rate across genotypes could impact our broad application of these assays. Thus, in Year 1 we will test each assay across a subset (~20) of newly established DO fibroblast cultures as described above. If an assay proves difficult to implement across our diverse fibroblast panel, we will employ substitute assays according to the criteria listed above as the budget allows.

AIM 3. Identify key mechanisms in renal arsenic toxicity using systems toxicology.

Rationale: Arsenic exposure leads to problems in multiple organs, but one of the cell types with the highest accumulation and subsequent negative impact is the kidney proximal tubule cell (PTC)[25, 26]. Therefore, this cell type is an excellent model to identify and study the AOPs of As toxicity. Practical constraints do not allow us to directly assay isolated PTCs; thus, we will perform whole-kidney assays that address PTC-specific functions, a similar approach we have used to study the impact of lithium on the collecting duct principal cells of the kidney[58, 59] (de Groot and Korstanje, under review). Based on the literature, we have generated a model of how As exposure impacts PTCs (**Figure 7**)[60-62]. After both genetic and environmental perturbations (as described in Aim 1), we will carry out omics analysis on kidney samples to ensure that we capture all anticipated as well as any unexpected molecular changes (gene expression, protein expression, metabolites) in the system and adjust our model accordingly. Using this approach, we can identify AOPs that can be validated by measuring the molecular response of the kidney to As exposure in a given genetic background and comparing this to the response as predicted by our model.

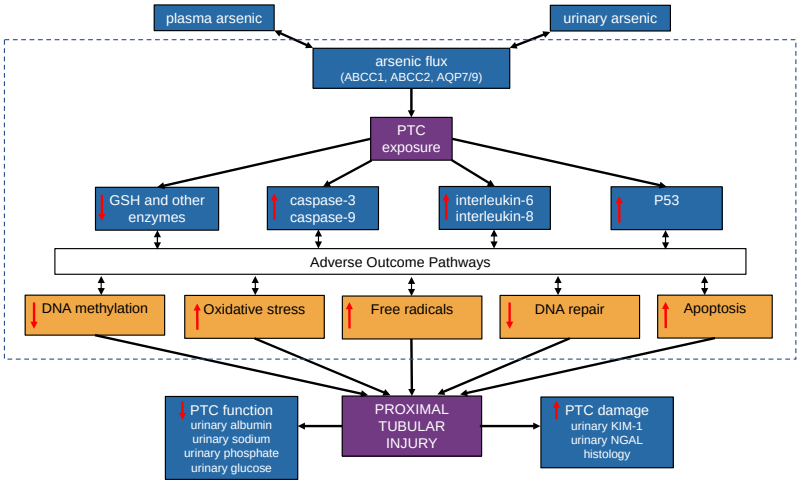

Figure 7: Model of arsenic impact on PTCs. The amount of arsenic that the PTC is exposed to is determined by the amount of *plasma arsenic* that is filtered by the kidney and leaves the body as *urinary arsenic*. Several transporters are responsible for the actual flux in the cell. This impact of **PTC exposure to arsenic** can be evidenced by changes in DNA methylation, oxidative stress, and apoptosis. These changes can lead to proximal tubular injury, which manifests as a reduced **PTC function** and **PTC damage.**

Experimental Methods:

For the full Discovery cohort of 640 DO mice, we will assess three outcomes in the context of the kidney. We will estimate **PTC exposure to arsenic** by measuring As^{III}, As^{V}, and total As in plasma and urine at 2, 3, and 12 months of age by Hydride Generation Atomic Absorption spectrometry (**see letter from Dr. Amirbahman**). At 2, 3, and 12 months, **PTC function** will be assessed through measurement of key proteins and metabolites in urine that are either exclusively or predominantly regulated by the PTC, including albumin, sodium, phosphate, and glucose. We will follow our established methods using a Beckman Coulter AU680 chemistry analyzer[63, 64]. To assess **PTC damage**, we will use several approaches**:** 1) measure urine levels of KIM1 and NAG, two proteins that are excreted after PTC damage[65], by ELISA at 2, 3, and 12 months; 2) perform image analysis on Periodic acid-Schiff (PAS)-stained sections of the kidneys that will be collected at 12 months and quantify damage using our machine learning algorithms (Sheehan et al, AJP Renal, in revision); and 3) TUNEL staining to detect apoptotic cells, which will allow us to compare our *in vivo* data with the TUNEL assay data from the cultured cells that have the same genetic background (see Aim 2).

We will select 240 mice from three exposures for in-depth characterization. Our aim is to select the exposure groups that are maximally informative for G x E QTL mapping based on measures of PTC function and PTC injury. For each of the 560 mice in the non-control exposure groups, we will determine its PTC status as impaired or not impaired based on comparison to the control group. If any of the functional measurements falls beyond the 90th percentile of the control value, the mouse will be labeled as *impaired*. We will select a *low exposure group* as the lowest exposure for which at least 10% of mice are impaired. We will select a *high exposure* group as the highest exposure for which not more than 90% of mice are impaired. An *intermediate exposure group* will be selected as the exposure level that is nearest to the middle of the low exposure and high exposure groups. These parameters may require adjustment based on actual outcome distributions but we will keep clinical norms in mind when determining whether or not an animal is functionally impaired.

For these selected 240 DO mice, we will define the response to As exposure through five specific analyses. ***First,* to quantify levels of key proteins** in the model that are relevant to As exposure (**Figure 7**), we will perform western blots of kidney lysates to detect: the transporters ABCC1, ABCC2, AQP7, AQP9 that mediate As uptake and excretion; apoptosis regulators caspase-3 and caspase-9; inflammatory cytokines interleukin-6 and interleukin-8; and p53, a cell cycle regulator that triggers apoptosis in response to DNA damage and cellular stress. Multiplex western blots will be performed using validated antibodies that are detected by DyLight fluorescent secondary antibodies. ***Second*, to quantify gene expression**, we will subject isolated whole-kidney RNA to genome-wide 2x150 bp sequencing on an Illumina NextSeq, obtaining a minimum of 50 million reads per sample. Although 2x150 bp sequencing costs more than 2x75 bp or 2x100 bp sequencing, it has the advantage of allowing more reliable identification of RNA isoforms from alternative splicing and RNA editing. ***Third*, to quantify metabolites**, we will conduct an unbiased analysis of metabolites, including those associated with oxidative stress, methylation pathways, and previously reported As-induced metabolic perturbations. Both hydrophilic and hydrophobic metabolites will be extracted from kidney using methanol/ water/MTBE[66, 67], and a single, pooled quality-control sample will be prepared and periodically analyzed to quantify analytical reproducibility for each detected metabolite[68]. The resulting hydrophilic and hydrophobic phases from each sample will be analyzed by ultra-high-performance liquid chromatography tandem mass spectrometry (UHPLC-MS) applying four complementary assays: two hydrophilic interaction chromatography (HILIC) assays for hydrophilic metabolites (positive and negative ion modes) and two reversed phase C_{30} assays for lipids (positive and negative ion modes). Polar metabolites will be extracted from urine by dilution in acetonitrile, and analyzed using the same two HILIC assays as above. All analyses will be performed applying a Thermo Scientific Ultimate3000 UHPLC system coupled to a Thermo Scientific electrospray Q Exactive mass spectrometer with full scan data to be collected by applying a high mass resolution of 70,000 (FWHM at m/z 200). Online MS/MS data will also be acquired to aid in metabolite identification[69]. All raw data will be processed using XCMS[70]. Metabolite annotation will be performed applying PUTMEDID_LCMS for metabolite feature integration and derivation of molecular formula, and MS/MS data will be searched against our in-house library of >400 metabolites and external mass spectral libraries (*e.g.,* mzCloud)[71]. ***Fourth*, we will quantify arsenic methylation**, which occurs in several tissues including the kidney[49, 50]. The species found in urine are inorganic As^{III} + As^{V} (10–30%), monomethylated (MMA as MMA^{III} + MMA^{V}) (10–20%), and dimethylated (DMA as DMA^{III} + DMA^{V}) (60–80%). Differences in the levels of these species between males and females, and between different populations, suggests the presence of functional genetic polymorphisms in As methylation[17]. In the early 2000s, an As methyltransferase (AS3MT) was identified. In humans, variants in this gene have been associated with variation in As sensitivity[72]. We will measure the different As species in mouse urine using inductively coupled plasma mass spectrometry. ***Fifth*, we will quantify DNA methylation** for several genes that have been reported to become hypermethylated as a result of As exposure[73]. Specifically, we will measure the methylation status of *Gprin1*, *Ptprj*, and *Osbp2* using bisulfite sequencing.

Genetic analysis and identification of key factors involved in arsenic exposure: We will identify genetic determinants of As-related injury by performing genetic mapping on each trait. We will fit a model in which the trait is regressed on sex, dose and experimental cohort with a mixed-effects term to adjust for the error-correlation between related individuals[74]. We will also search for G x E interactions by performing genetic mapping with dose as a covariate that interacts with genotype. Significance thresholds are determined via resampling of the trait data[45]. We will identify candidate genes within loci that are above the α=0.05 significance level by performing two analyses. First, we impute the DO founder SNPs onto each DO genome and perform association mapping to identify genetic variants that alter the protein structure[74]. Second, we will perform mediation analysis by regressing the trait on the expression of each kidney gene in our dosed cohorts of DO mice[38]. We will also search for genes with constitutive expression in untreated mice that may influence

susceptibility to As-related toxicity[13]. We have extensive liver and kidney expression and proteomics data[38] (Churchill and Korstanje, unpublished) from untreated DO mice and we will search for genes that have eQTL in the same location and with a similar pattern of allele effects as the trait QTL.

Systems biology of arsenic exposure: In our computational and mathematical modeling of the complex biological system of As exposure, we will start with a model derived from the literature on what is known about the effect of As exposure on the PTC (**Figure 7**). By measuring phenotypes that give us information on the different parts of the model in a large number of genetically diverse DO mice (which provide systematic perturbations in the model), the model can be further refined. Also, candidate genes, responsible for individual variation in response to As exposure, identified by the genetic analysis described above can be integrated into the model and their candidacy can be tested (see below).

Our study will generate a model for the impact of arsenic exposure on the kidney in a mammalian system that can predict PTC function and damage given a known genetic background. We will assess the robustness of our model by testing its ability to predict the readouts of As exposure in genetically diverse RIX mice (see Aim 1). As the genotype of each RIX combination is known at high resolution and their genomes encompass the diversity of the DO mouse population, we can determine the presence of the identified candidate genes in each mouse and use our model to predict for each genotype the impact of As exposure on PTC function and PTC damage. These mice will then be subjected to chronic As exposure over a range of doses as described for the Discovery cohort, allowing us to compare the model predictions to real measures of these parameters following actual exposure as we will do in the Validation cohort.

Expected results, potential pitfalls, and possible solutions. The data generated in our systems toxicology studies will establish a model of the mechanisms through which As enters the cell, changes metabolism, and leads to damage. They will allow us to identify genes, AOPs, and biomarkers that are involved in individual variation in response to As exposure of the kidney. We have been studying the molecular changes with age and disease in the mouse kidney, in particular the PTC, for many years and all techniques are well established in our lab. We also have a lot of experience in managing large cohort studies with complex sample collection schedules. We therefore do not anticipate any difficulties in sample collection, processing, and measurements.

Although collecting and homogenizing the whole kidney instead of isolating PTCs for our studies will introduce some additional variation, this choice is driven by practicality as the number of animals in the study would make it hard to isolate the PTCs from all animals in a consistent way. The variation among different isolates would be much greater than the variation caused by measurements in total kidney instead of isolated PTCs. However, PTCs are the most abundant cell type in the kidney and the most affected by As exposure[25, 26]. Also, many of the measurements we perform are PTC-specific. For example, KIM1 is a protein that is only expressed by the PTC and is only detected in urine upon damage of the PTC. We therefore expect our whole-kidney results to correlate well with PTC biology.

SEX AND BIOLOGICAL VARIABLES: Animals of both sexes will be used in all aspects of the proposed experiments.

RIGOR AND REPRODUCIBILITY: Measures undertaken by the team to ensure scientific rigor have been described in the experimental approach section as well as in the "**Resource Sharing Plan"** and in "**Authentication of Key Biological and/or Chemical Resources."** Main factors considered by our team were sex as a biological variable, positive and negative controls, and statistical analysis of data.

Deliverables of the proposed project. Our study to evaluate arsenic (As) toxicity using genetically diverse DO mice will deliver:
1) Guidelines for sample size and other parameters that can inform the design of new toxicity evaluations using DO mice.
2) New statistical methods for benchmark dose estimation using an empirical-Bayes framework that accounts for individual genetic variation and provides a data-driven model for determining adjustment factors to improve risk assessment accuracy.
3) Evaluation of *in vitro* screening methods to predict whole-animal exposure outcomes and inform screens using human cell lines.
4) Molecular biomarkers of arsenic exposure and arsenic-induced kidney damage that can be translated to application in humans.
5) Validated AOPs that describe mechanisms of exposure, metabolism, and transport of arsenic in the kidney. These predictive computational models of response to arsenic exposure in DO mice will help to prioritize treatment and risk assessment in humans.
6) Community resources of data, banked tissues, and primary cells with extensive study data to provide a common platform for the evaluation of new analysis methods. There is already expressed interest in our data from the toxicology community (**see letter from Dr. Joseph Shaw**). Our tissues will further understanding of whole-organism effects of arsenic exposure while our iPSC-programmable primary fibroblast panel captures genetic and cellular diversity for future systems toxicogenomic studies.
7) A proof of principle study for *systems toxicology* – using chemical exposures as interventions together with genetic variation to establish causal pathways across multiple levels of molecular, cellular, organ-specific, and whole-animal clinical outcomes.

BIBLIOGRAPHY & REFERENCES CITED

1. Brandon EF, Bulder AS, van Engelen JG, Mahieu CM, Mennes WC, Pronk ME, Rietveld AG, van de Ven BM, Ten Voorde SE, Wolterink G, Slob W, Zeilmaker MJ, Bessems JG. Does EU legislation allow the use of the Benchmark Dose (BMD) approach for risk assessment? Regulatory toxicology and pharmacology : RTP. 2013;67(2):182-8. Epub 2013/07/23. doi: 10.1016/j.yrtph.2013.07.005. PubMed PMID: 23871753.
2. Vinken M. The adverse outcome pathway concept: a pragmatic tool in toxicology. Toxicology. 2013;312:158-65. Epub 2013/08/28. doi: 10.1016/j.tox.2013.08.011. PubMed PMID: 23978457.
3. NRC. Toxicity Testing in the 21st Century: A Vision and a Strategy. Washington, DC: The National Academies Press; 2007. 216 p.
4. Pierce BL, Tong L, Argos M, Gao J, Farzana J, Roy S, Paul-Brutus R, Rahaman R, Rakibuz-Zaman M, Parvez F, Ahmed A, Quasem I, Hore SK, Alam S, Islam T, Harjes J, Sarwar G, Slavkovich V, Gamble MV, Chen Y, Yunus M, Rahman M, Baron JA, Graziano JH, Ahsan H. Arsenic metabolism efficiency has a causal role in arsenic toxicity: Mendelian randomization and gene-environment interaction. Int J Epidemiol. 2013;42(6):1862-71. Epub 2014/02/19. doi: 10.1093/ije/dyt182. PubMed PMID: 24536095; PMCID: PMC3887566.
5. Bustaffa E, Stoccoro A, Bianchi F, Migliore L. Genotoxic and epigenetic mechanisms in arsenic carcinogenicity. Arch Toxicol. 2014;88(5):1043-67. Epub 2014/04/03. doi: 10.1007/s00204-014-1233-7. PubMed PMID: 24691704.
6. NRC. Critical Aspects of EPA's IRIS Assessment of Inorganic Arsenic: Interim Report. Washington, DC: The National Academies Press; 2013. 127 p.
7. National Primary Drinking Water Regulations; Arsenic and Clarifications to Compliance and New Source Contaminants Monitoring, Stat. Parts 9, 141 and 142 (1/22/2001, 2001).
8. WHO. Guidelines for Drinking Water Quality, 4th Edition,. Geneva, Switzerland: World Health Organization, 2011.
9. Tokar EJ, Benbrahim-Tallaa L, Ward JM, Lunn R, Sams RL, 2nd, Waalkes MP. Cancer in experimental animals exposed to arsenic and arsenic compounds. Crit Rev Toxicol. 2010;40(10):912-27. Epub 2010/09/04. doi: 10.3109/10408444.2010.506641. PubMed PMID: 20812815; PMCID: PMC3076186.
10. Harrill AH, McAllister KA. New Rodent Population Models May Inform Human Health Risk Assessment and Identification of Genetic Susceptibility to Environmental Exposures. Environmental health perspectives. 2017;125(8):086002. Epub 2017/09/09. doi: 10.1289/EHP1274. PubMed PMID: 28886592.
11. Srivastava A, Morgan AP, Najarian ML, Sarsani VK, Sigmon JS, Shorter JR, Kashfeen A, McMullan RC, Williams LH, Giusti-Rodriguez P, Ferris MT, Sullivan P, Hock P, Miller DR, Bell TA, McMillan L, Churchill GA, de Villena FP. Genomes of the Mouse Collaborative Cross. Genetics. 2017;206(2):537-56. Epub 2017/06/09. doi: 10.1534/genetics.116.198838. PubMed PMID: 28592495; PMCID: PMC5499171.
12. Gatti DM, Weber SN, Goodwin NC, Lammert F, Churchill GA. Genetic background influences susceptibility to chemotherapy-induced hematotoxicity. The pharmacogenomics journal. 2017. Epub 2017/06/14. doi: 10.1038/tpj.2017.23. PubMed PMID: 28607509; PMCID: PMC5729066.
13. French JE, Gatti DM, Morgan DL, Kissling GE, Shockley KR, Knudsen GA, Shepard KG, Price HC, King D, Witt KL, Pedersen LC, Munger SC, Svenson KL, Churchill GA. Diversity Outbred Mice Identify Population-Based Exposure Thresholds and Genetic Factors that Influence Benzene-Induced Genotoxicity. Environmental health perspectives. 2015;123(3):237-45. doi: 10.1289/ehp.1408202. PubMed PMID: 25376053; PMCID: 4348743.
14. E.P.A. US. Benchmark Dose Software. 2.7 ed: U.S.E.P.A.; 2018.
15. EFSA Scientific Committee;Hardy AB, D.; Halldorsson, T.; Jeger, M.J.;Knutsen, K.H.; More, S.; Mortensen, A.; Naegeli, H.; Noteborn, H.;Ockleford, C.; Ricci, A.; Rychen, G.; SilanoV.; Solecki, R.; Turck, D.;Aerts, M.; Bodin, L.; Davis, A.; Edler LG-R, U.; Sand, S.;Slob, W.; Bottex, B.; Abrahantes, J.C.; Marques, D.C.;Kass G.; Schlatter, J.R. Update: use of the benchmark dose approach in risk assessment. European Food Safety Authoriy Journal. 2017;15(1). Epub 11/17/2016. doi: doi: 10.2903/j.efsa.2017.4658.

16. Dorne JL, Renwick AG. The refinement of uncertainty/safety factors in risk assessment by the incorporation of data on toxicokinetic variability in humans. Toxicol Sci. 2005;86(1):20-6. Epub 2005/04/01. doi: 10.1093/toxsci/kfi160. PubMed PMID: 15800035.

17. Loffredo CA, Aposhian HV, Cebrian ME, Yamauchi H, Silbergeld EK. Variability in human metabolism of arsenic. Environ Res. 2003;92(2):85-91. Epub 2003/07/12. PubMed PMID: 12854687.

18. Abdo N, Xia M, Brown CC, Kosyk O, Huang R, Sakamuru S, Zhou YH, Jack JR, Gallins P, Xia K, Li Y, Chiu WA, Motsinger-Reif AA, Austin CP, Tice RR, Rusyn I, Wright FA. Population-based in vitro hazard and concentration-response assessment of chemicals: the 1000 genomes high-throughput screening study. Environmental health perspectives. 2015;123(5):458-66. Epub 2015/01/27. doi: 10.1289/ehp.1408775. PubMed PMID: 25622337; PMCID: PMC4421772.

19. Bhat VS, Meek MEB, Valcke M, English C, Boobis A, Brown R. Evolution of chemical-specific adjustment factors (CSAF) based on recent international experience; increasing utility and facilitating regulatory acceptance. Crit Rev Toxicol. 2017;47(9):729-49. Epub 2017/07/07. doi: 10.1080/10408444.2017.1303818. PubMed PMID: 28681680.

20. Marchiset-Ferlay N, Savanovitch C, Sauvant-Rochat MP. What is the best biomarker to assess arsenic exposure via drinking water? Environment international. 2012;39(1):150-71. doi: 10.1016/j.envint.2011.07.015. PubMed PMID: 22208756.

21. Dopp E, Hartmann LM, von Recklinghausen U, Florea AM, Rabieh S, Zimmermann U, Shokouhi B, Yadav S, Hirner AV, Rettenmeier AW. Forced uptake of trivalent and pentavalent methylated and inorganic arsenic and its cyto-/genotoxicity in fibroblasts and hepatoma cells. Toxicol Sci. 2005;87(1):46-56. doi: 10.1093/toxsci/kfi218. PubMed PMID: 15947026.

22. De Vizcaya-Ruiz A, Barbier O, Ruiz-Ramos R, Cebrian ME. Biomarkers of oxidative stress and damage in human populations exposed to arsenic. Mutat Res. 2009;674(1-2):85-92. doi: 10.1016/j.mrgentox.2008.09.020. PubMed PMID: 18984063.

23. Jha AN, Noditi M, Nilsson R, Natarajan AT. Genotoxic effects of sodium arsenite on human cells. Mutat Res. 1992;284(2):215-21. PubMed PMID: 1281272.

24. Madden EF, Fowler BA. Mechanisms of nephrotoxicity from metal combinations: a review. Drug Chem Toxicol. 2000;23(1):1-12. Epub 2000/03/11. doi: 10.1081/DCT-100100098. PubMed PMID: 10711385.

25. Brown MM, Rhyne BC, Goyer RA. Intracellular effects of chronic arsenic administration on renal proximal tubule cells. J Toxicol Environ Health. 1976;1(3):505-14. Epub 1976/01/01. doi: 10.1080/15287397609529349. PubMed PMID: 1246090.

26. Peraza MA, Cromey DW, Carolus B, Carter DE, Gandolfi AJ. 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. 2006;26(4):356-67. Epub 2006/06/15. doi: 10.1002/jat.1149. PubMed PMID: 16773642.

27. Carlin DJ, Naujokas MF, Bradham KD, Cowden J, Heacock M, Henry HF, Lee JS, Thomas DJ, Thompson C, Tokar EJ, Waalkes MP, Birnbaum LS, Suk WA. Arsenic and Environmental Health: State of the Science and Future Research Opportunities. Environmental health perspectives. 2016;124(7):890-9. Epub 2015/11/21. doi: 10.1289/ehp.1510209. PubMed PMID: 26587579; PMCID: PMC4937867.

28. Brockmeier EK, Hodges G, Hutchinson TH, Butler E, Hecker M, Tollefsen KE, Garcia-Reyero N, Kille P, Becker D, Chipman K, Colbourne J, Collette TW, Cossins A, Cronin M, Graystock P, Gutsell S, Knapen D, Katsiadaki I, Lange A, Marshall S, Owen SF, Perkins EJ, Plaistow S, Schroeder A, Taylor D, Viant M, Ankley G, Falciani F. The Role of Omics in the Application of Adverse Outcome Pathways for Chemical Risk Assessment. Toxicol Sci. 2017;158(2):252-62. Epub 2017/05/20. doi: 10.1093/toxsci/kfx097. PubMed PMID: 28525648.

29. Campos B, Biales AD, Brown JB, Gallagher K, Henry TR, Marshall S, Sappington KG, Viant MR, Whale G, Colbourne JK. The Challenge: How omics technologies can enhance chemical safety regulation: Perspectives from academia, government, and industry. Environmental Toxicology and Chemistry. 2018;In Press.

30. Antczak P, White TA, Giri A, Michelangeli F, Viant MR, Cronin MT, Vulpe C, Falciani F. Systems Biology Approach Reveals a Calcium-Dependent Mechanism for Basal Toxicity in Daphnia magna. Environ Sci Technol. 2015;49(18):11132-40. Epub 2015/08/06. doi: 10.1021/acs.est.5b02707. PubMed PMID: 26244374.

31. Development OfEC-oa. OECD Guidline for the Testing of Chemicals. Repeated Dose 90-day Oral Toxicity Study in Rodents: Organisation for Economic Co-operation and Development; 1998. p. 1-10.

32. Drinking Water Standard for Arsenic, (2001).

33. Perry MR, Wyllie S, Raab A, Feldmann J, Fairlamb AH. Chronic exposure to arsenic in drinking water can lead to resistance to antimonial drugs in a mouse model of visceral leishmaniasis. Proc Natl Acad Sci U S A. 2013;110(49):19932-7. Epub 2013/10/30. doi: 10.1073/pnas.1311535110. PubMed PMID: 24167266; PMCID: PMC3856816.

34. Nigra AE, Sanchez TR, Nachman KE, Harvey D, Chillrud SN, Graziano JH, Navas-Acien A. The effect of the Environmental Protection Agency maximum contaminant level on arsenic exposure in the USA from 2003 to 2014: an analysis of the National Health and Nutrition Examination Survey (NHANES). Lancet Public Health. 2017;2(11):e513-e21. Epub 2017/12/19. doi: 10.1016/S2468-2667(17)30195-0. PubMed PMID: 29250608; PMCID: PMC5729579.

35. Kozul CD, Nomikos AP, Hampton TH, Warnke LA, Gosse JA, Davey JC, Thorpe JE, Jackson BP, Ihnat MA, Hamilton JW. Laboratory diet profoundly alters gene expression and confounds genomic analysis in mouse liver and lung. Chem Biol Interact. 2008;173(2):129-40. Epub 2008/04/09. doi: 10.1016/j.cbi.2008.02.008. PubMed PMID: 18396267.

36. Kozul-Horvath CD, Zandbergen F, Jackson BP, Enelow RI, Hamilton JW. Effects of low-dose drinking water arsenic on mouse fetal and postnatal growth and development. PLoS One. 2012;7(5):e38249. Epub 2012/06/14. doi: 10.1371/journal.pone.0038249. PubMed PMID: 22693606; PMCID: PMC3365045.

37. Rao GN. New diet (NTP-2000) for rats in the National Toxicology Program toxicity and carcinogenicity studies. Fundam Appl Toxicol. 1996;32(1):102-8. Epub 1996/07/01. PubMed PMID: 8812243.

38. Chick JM, Munger SC, Simecek P, Huttlin EL, Choi K, Gatti DM, Raghupathy N, Svenson KL, Churchill GA, Gygi SP. Defining the consequences of genetic variation on a proteome-wide scale. Nature. 2016;534(7608):500-5. doi: 10.1038/nature18270. PubMed PMID: 27309819; PMCID: PMC5292866.

39. Burden N, Mahony C, Muller BP, Terry C, Westmoreland C, Kimber I. Aligning the 3Rs with new paradigms in the safety assessment of chemicals. Toxicology. 2015;330:62-6. Epub 2015/05/02. PubMed PMID: 25932488.

40. Slob W. Dose-response modeling of continuous endpoints. Toxicol Sci. 2002;66(2):298-312. Epub 2002/03/16. PubMed PMID: 11896297.

41. Chiu WA, Slob W. A Unified Probabilistic Framework for Dose-Response Assessment of Human Health Effects. Environmental health perspectives. 2015;123(12):1241-54. Epub 2015/05/27. doi: 10.1289/ehp.1409385. PubMed PMID: 26006063; PMCID: PMC4671238.

42. Slob W, Bakker MI, Biesebeek JD, Bokkers BG. Exploring the uncertainties in cancer risk assessment using the integrated probabilistic risk assessment (IPRA) approach. Risk Anal. 2014;34(8):1401-22. Epub 2014/04/29. doi: 10.1111/risa.12194. PubMed PMID: 24766324.

43. Carpenter B, Gelman A, Hoffman MD, Lee D, Goodrich B, Betancourt M, Brubaker M, Guo J, Li P, Riddell A. Stan: A Probabilistic Programming Language. 2017. 2017;76(1):32. Epub 2017-01-11. doi: 10.18637/jss.v076.i01.

44. Solberg LC, Baum AE, Ahmadiyeh N, Shimomura K, Li R, Turek FW, Churchill GA, Takahashi JS, Redei EE. Sex- and lineage-specific inheritance of depression-like behavior in the rat. Mamm Genome. 2004;15(8):648-62. Epub 2004/10/01. doi: 10.1007/s00335-004-2326-z. PubMed PMID: 15457344; PMCID: PMC3764448.

45. Churchill GA, Doerge RW. Empirical threshold values for quantitative trait mapping. Genetics. 1994;138(3):963-71. Epub 1994/11/01. PubMed PMID: 7851788; PMCID: 1206241.

46. Broman KW. R/qtl2. Madison, WI2018. p. R/qtl2 (aka qtl) is a reimplementation of the QTL analysis software R/qtl, to better handle high-dimensional data and complex cross designs.

47. Friedman N, Linial M, Nachman I, Pe'er D. Using Bayesian networks to analyze expression data. J Comput Biol. 2000;7(3-4):601-20. Epub 2000/12/07. doi: 10.1089/106652700750050961. PubMed PMID: 11108481.

48. Sheehan NA, Didelez V, Burton PR, Tobin MD. Mendelian randomisation and causal inference in observational epidemiology. PLoS Med. 2008;5(8):e177. Epub 2008/08/30. doi: 10.1371/journal.pmed.0050177. PubMed PMID: 18752343; PMCID: PMC2522255.

49. Drobna Z, Walton FS, Harmon AW, Thomas DJ, Styblo M. Interspecies differences in metabolism of arsenic by cultured primary hepatocytes. Toxicol Appl Pharmacol. 2010;245(1):47-56. Epub 2010/02/09. doi: 10.1016/j.taap.2010.01.015. PubMed PMID: 20138079; PMCID: PMC2862857.

50. Healy SM, Casarez EA, Ayala-Fierro F, Aposhian H. Enzymatic methylation of arsenic compounds. V. Arsenite methyltransferase activity in tissues of mice. Toxicol Appl Pharmacol. 1998;148(1):65-70. Epub 1998/02/18. PubMed PMID: 9465265.

51. Antonelli R, Shao K, Thomas DJ, Sams R, 2nd, Cowden J. AS3MT, GSTO, and PNP polymorphisms: impact on arsenic methylation and implications for disease susceptibility. Environ Res. 2014;132:156-67. Epub 2014/05/06. doi: 10.1016/j.envres.2014.03.012. PubMed PMID: 24792412.

52. Pierce BL, Kibriya MG, Tong L, Jasmine F, Argos M, Roy S, Paul-Brutus R, Rahaman R, Rakibuz-Zaman M, Parvez F, Ahmed A, Quasem I, Hore SK, Alam S, Islam T, Slavkovich V, Gamble MV, Yunus M, Rahman M, Baron JA, Graziano JH, Ahsan H. Genome-wide association study identifies chromosome 10q24.32 variants associated with arsenic metabolism and toxicity phenotypes in Bangladesh. PLoS Genet. 2012;8(2):e1002522. doi: 10.1371/journal.pgen.1002522. PubMed PMID: 22383894; PMCID: 3285587.

53. Gong X, Ivanov VN, Davidson MM, Hei TK. Tetramethylpyrazine (TMP) protects against sodium arsenite-induced nephrotoxicity by suppressing ROS production, mitochondrial dysfunction, pro-inflammatory signaling pathways and programed cell death. Arch Toxicol. 2015;89(7):1057-70. doi: 10.1007/s00204-014-1302-y. PubMed PMID: 24961358; PMCID: 4377316.

54. Dong JT, Luo XM. Arsenic-induced DNA-strand breaks associated with DNA-protein crosslinks in human fetal lung fibroblasts. Mutat Res. 1993;302(2):97-102. PubMed PMID: 7684511.

55. Czechanski A, Kim H, Byers C, Greenstein I, Stumpff J, Reinholdt LG. Kif18a is specifically required for mitotic progression during germ line development. Developmental biology. 2015;402(2):253-62. doi: 10.1016/j.ydbio.2015.03.011. PubMed PMID: 25824710; PMCID: 4450139.

56. Czechanski A, Byers C, Greenstein I, Schrode N, Donahue LR, Hadjantonakis AK, Reinholdt LG. Derivation and characterization of mouse embryonic stem cells from permissive and nonpermissive strains. Nature protocols. 2014;9(3):559-74. doi: 10.1038/nprot.2014.030. PubMed PMID: 24504480; PMCID: 4112089.

57. Peterson KA, Beane GL, Goodwin LO, Kutny PM, Reinholdt LG, Murray SA. CRISPRtools: a flexible computational platform for performing CRISPR/Cas9 experiments in the mouse. Mamm Genome. 2017;28(7-8):283-90. doi: 10.1007/s00335-017-9681-z. PubMed PMID: 28280930; PMCID: 5591755.

58. de Groot T, Damen L, Kosse L, Alsady M, Doty R, Baumgarten R, Sheehan S, van der Vlag J, Korstanje R, Deen PMT. Lithium reduces blood glucose levels, but aggravates albuminuria in BTBR-ob/ob mice. PLoS One. 2017;12(12):e0189485. Epub 2017/12/16. doi: 10.1371/journal.pone.0189485. PubMed PMID: 29244860; PMCID: PMC5731748.

59. Alsady M, de Groot T, Kortenoeven MLA, Carmone C, Neijman K, Bekkenkamp-Grovenstein M, Engelke U, Wevers R, Baumgarten R, Korstanje R, Deen PMT. Lithium induces aerobic glycolysis and glutaminolysis in collecting duct principal cells. Am J Physiol Renal Physiol. 2017:ajprenal002972017. Epub 2017/10/27. doi: 10.1152/ajprenal.00297.2017. PubMed PMID: 29070571.

60. Lu Y, Yuan H, Deng S, Wei Q, Guo C, Yi J, Wu J, Li R, Wen L, He Z, Yuan L. Arsanilic acid causes apoptosis and oxidative stress in rat kidney epithelial cells (NRK-52e cells) by the activation of the

caspase-9 and -3 signaling pathway. Drug Chem Toxicol. 2014;37(1):55-62. Epub 2013/07/16. doi: 10.3109/01480545.2013.806532. PubMed PMID: 23848144.

61. Tokumoto M, Lee JY, Fujiwara Y, Uchiyama M, Satoh M. Inorganic arsenic induces apoptosis through downregulation of Ube2d genes and p53 accumulation in rat proximal tubular cells. J Toxicol Sci. 2013;38(6):815-20. Epub 2013/11/12. PubMed PMID: 24212999.

62. Robles-Osorio ML, Sabath-Silva E, Sabath E. Arsenic-mediated nephrotoxicity. Ren Fail. 2015;37(4):542-7. Epub 2015/02/24. doi: 10.3109/0886022X.2015.1013419. PubMed PMID: 25703706.

63. Tian C, Gagnon LH, Longo-Guess C, Korstanje R, Sheehan SM, Ohlemiller KK, Schrader AD, Lett JM, Johnson KR. Hearing loss without overt metabolic acidosis in ATP6V1B1 deficient MRL mice, a new genetic model for non-syndromic deafness with enlarged vestibular aqueducts. Hum Mol Genet. 2017;26(19):3722-35. Epub 2017/09/22. doi: 10.1093/hmg/ddx257. PubMed PMID: 28934385.

64. Korstanje R, Deutsch K, Bolanos-Palmieri P, Hanke N, Schroder P, Staggs L, Brasen JH, Roberts IS, Sheehan S, Savage H, Haller H, Schiffer M. Loss of Kynurenine 3-Mono-oxygenase Causes Proteinuria. J Am Soc Nephrol. 2016;27(11):3271-7. Epub 2016/11/02. doi: 10.1681/ASN.2015070835. PubMed PMID: 27020856; PMCID: PMC5084883.

65. Han WK, Waikar SS, Johnson A, Betensky RA, Dent CL, Devarajan P, Bonventre JV. Urinary biomarkers in the early diagnosis of acute kidney injury. Kidney Int. 2008;73(7):863-9. Epub 2007/12/07. doi: 10.1038/sj.ki.5002715. PubMed PMID: 18059454; PMCID: PMC2586909.

66. Mirbahai L, Southam AD, Sommer U, Williams TD, Bignell JP, Lyons BP, Viant MR, Chipman JK. Disruption of DNA methylation via S-adenosylhomocysteine is a key process in high incidence liver carcinogenesis in fish. J Proteome Res. 2013;12(6):2895-904. Epub 2013/04/25. doi: 10.1021/pr400195u. PubMed PMID: 23611792.

67. Matyash V, Liebisch G, Kurzchalia TV, Shevchenko A, Schwudke D. Lipid extraction by methyl-tert-butyl ether for high-throughput lipidomics. Journal of lipid research. 2008;49(5):1137-46. Epub 2008/02/19. doi: 10.1194/jlr.D700041-JLR200. PubMed PMID: 18281723; PMCID: PMC2311442.

68. Dunn WB, Broadhurst D, Begley P, Zelena E, Francis-McIntyre S, Anderson N, Brown M, Knowles JD, Halsall A, Haselden JN, Nicholls AW, Wilson ID, Kell DB, Goodacre R, Human Serum Metabolome C. Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. Nature protocols. 2011;6(7):1060-83. Epub 2011/07/02. doi: 10.1038/nprot.2011.335. PubMed PMID: 21720319.

69. Mullard G, Allwood JW, Weber R, Brown M, Begley P, Hollywood KA, Jones M, Unwin RD, Bishop PN, Cooper G, Dunn WB. A new strategy for MS/MS data acquisition applying multiple data dependent experiments on Orbitrap mass spectrometers in non-targeted metabolomic applications. Metabolomics. 2015;11(5):1068–80. doi: https://doi.org/10.1007/s11306-014-0763-6.

70. Smith CA, Want EJ, O'Maille G, Abagyan R, Siuzdak G. XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. Anal Chem. 2006;78(3):779-87. Epub 2006/02/02. doi: 10.1021/ac051437y. PubMed PMID: 16448051.

71. Brown M, Wedge DC, Goodacre R, Kell DB, Baker PN, Kenny LC, Mamas MA, Neyses L, Dunn WB. Automated workflows for accurate mass-based putative metabolite identification in LC/MS-derived metabolomic datasets. Bioinformatics. 2011;27(8):1108-12. Epub 2011/02/18. doi: 10.1093/bioinformatics/btr079. PubMed PMID: 21325300; PMCID: PMC3709197.

72. Engstrom KS, Vahter M, Fletcher T, Leonardi G, Goessler W, Gurzau E, Koppova K, Rudnai P, Kumar R, Broberg K. Genetic variation in arsenic (+3 oxidation state) methyltransferase (AS3MT), arsenic metabolism and risk of basal cell carcinoma in a European population. Environ Mol Mutagen. 2015;56(1):60-9. Epub 2014/08/27. doi: 10.1002/em.21896. PubMed PMID: 25156000; PMCID: PMC4322484.

73. Rea M, Eckstein M, Eleazer R, Smith C, Fondufe-Mittendorf YN. Genome-wide DNA methylation reprogramming in response to inorganic arsenic links inhibition of CTCF binding, DNMT expression and cellular transformation. Sci Rep. 2017;7:41474. Epub 2017/02/06. doi: 10.1038/srep41474. PubMed PMID: 28150704; PMCID: PMC5288714.

74. Gatti DM, Svenson KL, Shabalin A, Wu LY, Valdar W, Simecek P, Goodwin N, Cheng R, Pomp D, Palmer A, Chesler EJ, Broman KW, Churchill GA. Quantitative trait locus mapping methods for diversity outbred mice. G3 (Bethesda). 2014;4(9):1623-33. doi: 10.1534/g3.114.013748. PubMed PMID: 25237114; PMCID: 4169154.