Genetic and epigenetic determinants of inter-individual variability in responses to toxicants
Lauren Lewis, Gregory E. Crawford, Terrence S. Furey and Ivan Rusyn

Abstract
It is well established that genetic variability has a major impact on susceptibility to common diseases, responses to drugs and toxicants, and influences disease-related outcomes. The appreciation that epigenetic marks also vary across the population is growing with more data becoming available from studies in humans and model organisms. In addition, the links between genetic variability, toxicity outcomes and epigenetics are being actively explored. Recent studies demonstrate that gene-by-environment interactions involve both chromatin states and transcriptional regulation, and that epigenetics provides important mechanistic clues to connect expression-related quantitative trait loci (QTL) and disease outcomes. However, studies of Gene × Environment × Epigenetics further extend the complexity of the experimental designs and create a challenge for selecting the most informative epigenetic readouts that can be feasibly performed to interrogate multiple individuals, exposures, tissue types and toxicity phenotypes. We propose that among the many possible epigenetic experimental methodologies, assessment of chromatin accessibility coupled with total RNA levels provides a cost-effective and comprehensive option to sufficiently characterize the complexity of epigenetic and regulatory activity in the context of understanding the inter-individual variability in responses to toxicants.

Addresses
1 Department of Veterinary Integrative Biosciences, Texas A&M University, College Station, TX, USA
2 Center for Genomic and Computational Biology and Department of Pediatrics, Division of Medical Genetics, Duke University, Durham, NC, USA
3 Department of Genetics, Department of Biology, Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

Corresponding author: Rusyn, Ivan (irusyn@cvm.tamu.edu)

1. Genetic variability
Estimation of the degree of inter-individual variability in the population is a required step in assessing the human health hazard posed by environmental chemicals. Indeed, the National Academies report Science and Decisions [1] called for the need to better “account for differences among humans in cancer susceptibility other than from possible early-life susceptibility.” Recent advances in the ability to conduct genome-wide association studies (GWAS) that identify quantitative trait loci (QTL) have enabled identification of genetic variants associated with important diseases [2]. It is clear that genetic variation influences the response of an individual to drugs and chemicals [3]. The blossoming field of personalized medicine now brings GWAS-enabled understanding of basic biology into clinical practice to determine how the knowledge of genetic variation can make therapies safer and more effective by tailoring selection and dosing of drugs for an individual patient [4].

GWAS that characterize effects of environmental toxicants on humans are usually based on epidemiological data, not controlled exposures [5]. This makes it a challenge to interpret findings from human cohorts exposed in the occupational or environmental settings. In addition, collection of tissues (with the exception of blood) from a wide variety of anatomical sites or developmental stages is not possible in humans that have been exposed to environmental toxicants. These limitations can be alleviated, at least partially, by the use of appropriate genetically-diverse laboratory animal-based model systems [6].

The mouse is a popular in vivo model for which genetic resources with publicly available genetic maps across dozens of strains are now available [7]. Mouse populations, such as the Collaborative Cross [8], provide an excellent testing system for evaluation of complexities in toxicokinetics and toxicodynamics [6,9–11]. In the past decade, it has been demonstrated convincingly that genetic diversity in the mouse can be used to identify sensitive sub-populations using a mouse model of the human population approach [12–25]. Most of the genetic variability among mouse strains has been focused on SNPs; however, variation in structure of DNA regions affecting DNA sequence length and/or orientation that includes deletions, insertions, copy-number gains,
inversions, and transposable elements, may also underpin susceptibility traits [26]. In addition, while inbred mouse strains are considered isogenic, intra-strain differences and their influence on experimental outcomes have been identified [27,28].

While advances in sequencing technologies, statistical genetics analysis methods and clinical trial designs have shown promise for the discovery of variants associated with drug response, interpretation of both human and mouse GWAS through identifying causal variants is a challenge, and the translation of the findings to the clinic and/or regulatory actions is slow. On the one hand, it remains difficult to interpret the outcomes of GWAS and validate genes underlying QTLs with certainty, due in part to not knowing which organs, tissues, and/or cell types any particular QTL is having a significant functional effect. On the other hand, the GWAS-driven attempts to disentangle treatment responders from non-responders via genetic predictors in pharmacogenetics studies have not been uniformly successful [29].

2. Linkages between genetic, transcriptional, and epigenetic variability

Comprehensive maps of human and mouse regulatory DNA were recently published by the ENCODE (Encyclopedia of DNA Elements) Consortium [30], mouse ENCODE [31], and the Roadmap Epigenomics Project [32]. These studies comprehensively characterized the location and relationships between chromatin accessibility, histone modifications, chromatin looping, transcription, DNA methylation and the occupancy of sequence-specific factors. The wide spectrum of different cultured cell lines and tissues that were assayed have identified over a million common and cell-type specific gene regulatory elements. Genome-wide chromatin accessibility analyses, originally performed by DNase-seq [33] and more recently by ATAC-seq [34], have become invaluable approaches for mapping the genomic location of transcriptionally-active chromatin. While consortia such as ENCODE and Roadmap have identified large numbers of putative regulatory elements, little is known about how these elements are affected by variation in genetics, sex, or exposure to individual or complex combinations of environmental stimuli.

Studies in a large and genetically heterogeneous collection of human lymphoblast cell lines (LCLs) [32] and tissues [35] have identified heritable variation in gene expression across humans. These expression quantitative trait loci (eQTL) studies have been complemented by a more limited number of chromatin studies that have identified QTLs that impact DNaseI sensitivity (dsQTL; [36]), histone modifications in chromatin (cQTLs; [37]), DNA methylation [38], and transcription factors binding sites [37,39]. These studies demonstrate the versatility and complexity of gene regulation, whereby modulation of gene expression is executed by different elements forming intricate networks that include changes in chromatin activity. In addition, these studies show how genetic variants identified in GWAS can be linked through a regulatory network to the associated gene. For example, it was shown that both locally and distally acting genetic variants exhibit strong influence on expression and chromatin [37,40]. It was also found that two-thirds of local eQTLs were also local dsQTLs or cQTLs [36], which means that the variation in chromatin is associated with variation in the expression levels of nearby genes. At the same time, a total of 15% of proximal histone QTLs were associated with changes in chromatin states at distal genomic regions with which they interact physically [41]. These data show that specific genetic variants modulating regulatory element activity may concurrently affect local and distal chromatin modifications and gene expression.

While population variability in DNA- and chromatinerelated epigenetic marks is well recognized, it has been shown that variability in miRNA expression in the population may be negligible as compared to the genetically-determined variability in mRNA expression [20,42]. Specifically, few eQTLs were observed for miRNAs in various tissues in population studies in mice [13,43,44]. The stability of miRNA expression in a genetically diverse population suggests that miRNAs may be a much more reliable population-wide biomarker of the effects of chemicals on epigenetic mechanisms of toxicity, as compared to changes in DNA methylation, chromatin and/or histone modifications. Indeed, chemical-induced disruptions in miRNA expression, a phenomenon established for a large number of toxicants, is recognized as an important toxicity mechanism [45]. Post-transcriptional regulation of mRNA levels by miRNAs is not a true epigenetic process. For the remainder of this review, though, we include miRNAs when discussing the epigenome for the sake of simplicity as their primary function is to regulate gene expression.

3. Environmental agents cause toxicity through epigenetic mechanisms

Epigenetic reprogramming has been proposed as an integral part of the “genome instability” enabling characteristic of cancer cells [46] and it is well established that chemical carcinogens may affect the cellular epigenetic state [47]. Changes in DNA methylation, histone/chromatin remodeling, and altered expression of miRNAs represent the most frequently reported toxicant-induced alterations of the epigenome [48]. Because of the potential impact of these epigenotoxic
effects on gene expression patterns and, consequently, on the toxicity phenotypes, epigenetic changes have been proposed as biomarkers of carcinogen exposure and effect [49,50].

One of the first examples of the linkages between environment, epigenetics and phenotypes were studies of in utero exposure to environmental agents that can also disrupt the epigenome. The agouti mouse model was used to demonstrate that environmental factors may affect the fetal epigenome [51]. Using this mouse model, maternal exposure to the endocrine disruptor BPA causes loss of DNA methylation at key loci, resulting in a shift in coat color of offspring [52]. Normal methylation patterns can then be restored with maternal dietary supplementation using methyl donors like folic acid. Evidence shows along with DNA methylation, variable histone modifications affect the inter-individual epigenetic variation of this metastable epiallele [53].

Another prominent example of how environmental toxicants may have epigenetic effects are studies on the mechanism of carcinogenesis for metals [54]. Arsenic, a ubiquitous environmental contaminant, disrupts the normal epigenome transforming the epigenetic landscape to reflect that of a cancer cell [55]. Exposure to metals like arsenic causes significant epigenetic modifications such as changes in global histone methylation levels [56]. After exposure to arsenic compounds, human lung carcinoma A549 cells showed an increase in global levels of H3K4me3 and H3K9me2 [57]. Similarly, human peripheral blood nuclear cells extracted from subjects exposed to high levels of arsenic in water had an increase in H3K9me2 levels [58].

Environmental contaminants can also alter gene expression by epigenetically reprogramming tissues. Neonatal BPA exposure increases H3K4me3 levels in promoters of genes associated with prostate cancer through activation of histone methyltransferase MLL1 [59]. Although there was no difference in basal expression of levels of BPA reprogrammed genes, once challenged with hormone treatment, there is enhanced gene-specific transcription. It is thought the change in levels of H3K4me3 primes these genes for an enhanced response. Additionally, evidence suggests that BPA exposure during prostate development could epigenetically reprogram the expression of Scgb2a1 in the adult prostate [60].

Epigenetic changes may be a consequence of DNA damage [61], or may be part of the non-genotoxic mechanisms of carcinogenesis [62]. The interplay between chemical-induced DNA damage response and transcription, DNA replication, and repair has only recently been linked to chromatin dynamics, especially to histone modifications and post-repair chromatin restoration at the sites of DNA damage [63]. For example, a local response to DNA double-stranded breaks gives rise to chromatin condensation which spreads at least over several Kb from the damage sites and can induce epigenetic silencing of the nearby genes [64]. In addition, it was shown that levels of the heterochromatin-associated histone modification H3K9me3 accounted for more than 40% of mutation rate variation, providing striking evidence that mutation rates in cancer genomes are related closely to chromatin organization [65]. Besides that, DNA repair can cause local chromatin state transitions eventually resulting in prolonged inactivation of transcription via not yet fully established gene silencing mechanisms. Modulation of the epigenetic status of damaged genes potentially expands the field of DNA damage into the sphere of regulation of gene expression [61]. While the interest in the role of epigenome in toxicity mechanisms is growing, the genotoxicity of chemicals has been more thoroughly studied and characterized, as evidenced by a systematic review of published studies of genotoxic carcinogens that investigated epigenetic endpoints [66].

DNA methylation is another key epigenetic mechanism, regulating both gene expression and chromatin stability. DNA methylation studies have been recently combined with RNA-seq and ChIP-seq to identify the role of the changes in the methylome in disease pathogenesis [67]. DNA methylation and genetic polymorphisms have important concomitant regulatory effects on transcription factor-driven gene expression [68]. Aberrant DNA methylation patterns due to exposure to environmental chemicals are also well-characterized. Exposure to benzene, metals, and traffic pollution are all examples of toxicants that can have an effect on DNA methylation [69–72].

4. Environmental effects on the epigenome in the context of genetic variability

There is now overwhelming evidence that connects genetic variability and epigenetic marks and that chemical exposures can exert toxicity through epigenetic mechanisms; yet less is known about how the effects on the epigenome may vary in the population. Intriguing novel insights into linkages between genotoxic and epigenetic mechanisms of carcinogenesis, and the role of genetic variability among individuals have been provided by studies of a classical genotoxic carcinogen 1,3-butadiene. It is a genotoxic chemical with DNA damaging effects that vary among genetically distinct individuals. Butadiene is a major industrial chemical used in the production of synthetic rubbers and polymers. It is also a ubiquitous environmental contaminant that is found in cigarette smoke and automobile exhaust. IARC has classified butadiene as a known human carcinogen [73]. It is well established that the mechanism of carcinogenicity is due to
butadiene’s reactive metabolites. These epoxides interact directly with DNA and form mutagenic DNA adducts. Butadiene also elicits an epigenetic response, causing significant loss of global DNA methylation as well as a decrease in H3K9, H3K27, and H3K20 trimethylation in C57BL/6J mouse liver [74].

Using a mouse population-based model it was shown that inter-individual (e.g., inter-strain) differences exist in both genotoxic and epigenotoxic effects of 1,3-butadiene exposure and that the chromatin remodeling response is at least one mechanism for the inter-strain differences in 1,3-butadiene-induced DNA damage [14]. Specifically, it was shown that 1,3-butadiene alters bulk chromatin histone mark levels resulting in strain-specific abundances of these marks. In particular, CAST/EiJ and C57BL/6J mice, two genetically distinct strains, exhibited basal and treatment-induced differences in overall levels of these histone marks.

5. Opinion: Studies of Gene × Environment × Epigenome in genetically diverse populations can provide a mechanistic explanation for expression and QTL effects

Studies of Gene × Environment × Epigenome seek to uncover not only relationships between environment exposures and gene expression levels, but also to determine the epigenetic gene regulatory mechanisms altered by these exposures that contribute to the observed expression changes (Fig. 1). These studies further extend the complexity of the experimental designs and require choosing the most informative epigenetic readouts to interrogate multiple individuals, exposures, tissue types and toxicity phenotypes (Fig. 2).

One possibility would be to perform ChiP-seq for many different histone marks or transcription factors, but this would be costly since multiple experiments would be required for each sample. It may also be impractical as these assays typically require non-trivial amounts of sample material limiting the number of experiments that could be performed, thus requiring an uninformed selection of assays to be made. RNA-seq protocols are able to comprehensively quantify multiple types of RNA transcripts, including protein-coding mRNAs, microRNAs (miRNAs), long non-coding RNAs (lncRNAs), enhancer RNAs (eRNAs), and circular RNAs, (circRNAs), and chemical modifications of these molecules [75,76]. These data are able to characterize variation in transcription generation and RNA processing, including alternative transcript initiation and splicing, providing detailed information about transcriptional outputs that affect cellular function. Therefore, a more cost-effective and unbiased approach would be to assess chromatin accessibility coupled with various RNA levels to comprehensively characterize epigenetic and transcriptional activity in the context of inter-individual variability in responses to toxicants.

The ideal way to identify gene regulatory elements would be a single empirical method that exploits a feature held in common between all functional classes of regulatory elements [77]. Identification of “accessible chromatin” through techniques like DNase-seq or ATAC-seq are currently the closest to this, because nucleosome depletion at active regulatory regions is a conserved feature of eukaryotic chromatin [78,79]. The sequence-specific binding of transcription factors and the recruitment of active histone marks at promoters, enhancers, and insulators is typically associated with nucleosome depletion in eukaryotic cells [80,81]. DNase-seq and ATAC-seq reveal these active regulatory elements by preferential digestion of or transposon insertion into nucleosome-depleted regions [34,80,82]. Until recently, DNase-seq was one of the few methods for assessing chromatin accessibility [83–86], and was used in the analysis of >100 human cell types and tissues [85,87].

While DNase-seq and ATAC-seq are both designed to measure accessible chromatin, DNase-seq technique is a derivative of the DNaseI hypersensitivity assay first described nearly 40 years ago [88]. Therefore, there is a massive amount of literature identifying and characterizing DNaseI hypersensitive sites, which supports the value of this method [77]. ATAC-seq was devised more recently, and has been shown to identify largely similar, but not identical, regions to that of DNase-seq [89]. The ATAC-seq protocol is less technically challenging and faster which facilitated many more labs to adopt the ATAC-seq over DNase-seq. It is not surprising that there are some differences between ATAC-seq and DNase-seq, and while some of this may be due to
Hypothetical study designs to evaluate toxicant-induced effects on the epigenome in (A) single genetic background, or (B) population-wide experimental animal model.
different sequence biases of the Tn5 transposase or the DNase enzyme, it is not yet clear what other factors may contribute to the differences [90,91]. An additional advantage of ATAC-seq is the ability to characterize much smaller amounts of cells or tissue [89], which is essential for experiments where there is no renewable resource, such as biobanked tissues. While a modified version of DNase-seq exists for small samples [92], this protocol has not been as widely adopted as ATAC-seq or single cell ATAC-seq [93–95]. However, generation of ATAC-seq libraries is more expensive due to the cost of Tn5 transposase and the need for requiring deeper sequencing (up to 50% of ATAC-seq reads come from the mitochondrial genome). Analysis of DNase-seq and ATAC-seq data is fairly similar.

We posit that data from strand-specific RNA-seq and ATAC-seq represents a cost-effective approach to identify and resolve regions of active or repressed enhancers and promoters. A common challenge among all regulatory element assays is determining the target gene(s) being regulated. Chromatin conformation capture assays provide the best evidence for interactions between regulatory elements and target genes. While these assays are cost-prohibitive to run on all samples, data has been generated from an increasing number of diverse cell-types and tissues through individual labs and within the ENCODE consortium that are being compiled and visualized in browsers [96]. In addition, software such as TargetFinder [97] can predict gene targets in specific samples using accessible chromatin data. In conjunction with RNA-seq data, alterations in chromatin accessibility correlated with expression changes in target genes either due to differences in genetic background or in response to environmental stimuli provide evidence of regulatory element function. ATAC-seq also does not immediately reveal what factors are bound in accessible chromatin regions. The DNA sequence binding preferences for an increasing number of factors is being defined [98], which can be used to predict which factors are present. Transcription factor footprinting using accessible chromatin data [99] provide additional evidence for sites where factors are bound. Admittedly, ChIP-seq data for particular factors and histone modifications provide more direct evidence of regulatory element function. Analysis of chromatin accessibility data can reveal which additional assays may provide the most relevant additional information.

Concentrating on just these two assays makes feasible the generation of data from sufficient samples of genetically diverse backgrounds to further link gene expression and epigenetic changes to genetic variability. Along with environmental phenotype data of interest, the stage is set for a full Gene × Environment × Epigenome analysis. The optimal study design would include not only post-environmental exposure data, but also matched pre-exposure data from control subjects. Susceptibility to damaging effects of toxicants may be due not only to genetic variability driving differential responses, but also may depend on baseline transcriptional and epigenetic states, also influenced by genetic background, that may be better primed in certain individuals to defend against injury [100].

Author declaration
We wish to confirm that there are no known conflicts of interest associated with this publication and there has been no significant financial support for this work that could have influenced its outcome.

We confirm that the manuscript has been read and approved by all named authors and that there are no other persons who satisfied the criteria for authorship but are not listed. We further confirm that the order of authors listed in the manuscript has been approved by all of us.

We confirm that we have given due consideration to the protection of intellectual property associated with this work and that there are no impediments to publication, including the timing of publication, with respect to intellectual property. In so doing we confirm that we have followed the regulations of our institutions concerning intellectual property.

We understand that the Corresponding Author is the sole contact for the Editorial process (including Editorial Manager and direct communications with the office). He is responsible for communicating with the other authors about progress, submissions of revisions and final approval of proofs. We confirm that we have provided a current, correct email address which is accessible by the Corresponding Author and which has been configured to accept email from IRusyn@cvm.tamu.edu.

References


Genetic and epigenetic determinants of inter-individual variability Lewis et al. 57


