Frontiers in Toxicogenomics in the Twenty-First Century—the Grand Challenge: To Understand How the Genome and Epigenome Interact with the Toxic Environment at the Single-Cell, Whole-Organism, and Multi-Generational Level

Douglas M. Ruden 1,2,3*, Katherine Gurdziel 1,2,3 and Michael Aschner 4

1 Department of Obstetrics and Gynecology, Wayne State University, Detroit, MI, United States, 2 C. S. Mott Center for Human Health and Development, Wayne State University, Detroit, MI, United States, 3 Institute of Environmental Health Sciences, Wayne State University, Detroit, MI, United States, 4 Department of Molecular Pharmacology, Albert Einstein College of Medicine, Bronx, NY, United States

Keywords: toxicogenomics, RNA-Seq, ScRNA-seq, organoids, multigenerational epigenetics, transgenerational epigenetics, ATAC-seq

INTRODUCTION

In 2011, we wrote the inaugural Grand Challenge for Frontiers in Toxicogenomics, a specialty section of Frontiers in Genetics. In the original Grand Challenge, we argued that the fields of “Genomics” and “Toxicology” needed to be merged into the new field of “Toxicogenomics.” Toxicology in the twentieth century involved animal testing and LD50 (Lethal Dose 50) levels where half the population of animals die from the toxin after a certain amount of time, such as in one week or one month. Toxicogenomics in the twenty-first century, in contrast, utilizes modern genetics, epigenetics, and molecular biology technologies.

Toxicogenomics in the twenty-first century should involve more research that utilizes whole-genome sequencing of model-organism or human-stem-cell-derived “organoids,” single-cell analyses, proteomics, complex genetics with conditional-lethal, knock-in or knock-out transgenes, and bioinformatics technologies.

Ideas for research topics in Toxicogenomics include:

- Organoid Toxicology Using Primary Cells and hESC-Derived Tissues.
- Single-Cell Transcriptomic Analyses of Toxicant-Exposed Tissues.
- Single-Cell Epigenomic Analyses of Toxicant-Exposed Tissues.
- Proteomics Technologies of Toxicant-Exposed Tissues.
- Multigenerational and Transgenerational Inheritance of Adaptive Epigenetic Changes of Toxicant-Exposed Model Animals and Humans.
- The Role of Extracellular Vesicles in Transmitting Signals Throughout the Organism After Toxicant Exposure.
- Imaging of Intracellular Alternations Caused By Toxicant Exposure.
- Toxicant-Specific eQTLs (Expression Quantitative Trait Loci).
- Exposures Measurements from Biobanked Tissues, Such as, Neonatal Dried Blood Spots and Shed Teeth.
- Toxicant Effects on the Microbiome.
- Sex-Specific Effects of Toxicants.
- Novel Methods—Second- and third-generation DNA sequencing technologies in Toxicogenomics, GWAS, etc.

In the following sections, we expand upon each of these topics. Hopefully, these topics will inspire future research and publications in these emerging fields in toxicogenomics.

ORGANOID TOXICOLOGY USING PRIMARY CELLS AND HUMAN EMBRYONIC STEM CELL (HESC)-DERIVED TISSUES

Organoid and organ-on-a-chip technologies are suitable replacement for animal models and will potentially impart multicellular in vitro organoid-related information on mechanisms of human toxicity and disease etiology. Organoids have been made from primary cell cultures, differentiated embryonic stem cells, or differentiated induced pluripotent stem cells (iPSC) for the liver (Cao et al., 2017), the heart (Beauchamp et al., 2015; Conant et al., 2017; Devarasetty et al., 2017), the brain (Hunsberger et al., 2015; Schwartz et al., 2015), the kidney (Astashkina et al., 2012; Chuah and Zink, 2017), the testes (Baert et al., 2017), the bladder (Janssen et al., 2013), and potentially multiple other organs and tissues (Xinari et al., 2015; Ingersoll et al., 2016; Skardal et al., 2016). Liver toxicity, indicated by liver cell death (Weber et al., 2016; Wong et al., 2016), and cardiac toxicity, indicated by changes in heart beat kinetics (Meattini et al., 2017; Yu et al., 2017), account for most drug candidate failures in human trials.

SINGLE-CELL TRANSCRIPTOMIC ANALYSES OF TOXICANT-EXPOSED TISSUES

Transcriptional profiling of various cell types in tissues and their responses to toxicants will impart cell population-specific molecular characterization, specifically identifying target cells and their sensitivity. Typical RNA-seq experiments from tissues are done from populations of hundreds of thousands or millions of cells, but the heterogeneity of the cells are missed. Single-cell RNA sequencing (scRNA-seq) was developed to study tumor heterogeneity (Müller and Diaz, 2017), but it has been used recently to study heterogeneity in other areas, such as, in the immune system (Yang et al., 2017), the brain (Müller and Diaz, 2017; Ofengeim et al., 2017), the retina (Macosko et al., 2015; Quadrato et al., 2017), in embryonic development (Mohammed et al., 2017), and in many other tissues (reviewed in Picelli, 2017). Originally, scRNA-seq involved fluorescence activated cell sorting (FACS) or manually dissecting single cells and performing the sequencing reactions in microtiter dishes. However, microfluidics systems have been developed by Fluidigm, Inc., that can sequence either 96 single cells or 800 single cells at a time (www.fluidigm.com). So-called drop-seq technologies were developed that can sequence tens-of-thousands of single cells at a time using single cells captured in femtoliter oil droplets (Macosko et al., 2015). Here is a link for a video on the origins of the drop-seq technique (https://www.youtube.com/watch?v=vL7ptq2Dcf0). The company 10X Genomics have developed a Chromium™ system to perform drop-seq on as many as 10,000 to 100,000 cells at a time, to a depth of up to 50,000 reads per cell (www.10xgenomics.com). Combining scRNA-seq and toxicology can potentially determine how the toxicants affect the distribution of cell types in a tissue.

SINGLE-CELL EPIGENOMIC ANALYSES OF TOXICANT-EXPOSED TISSUES

Drop-seq and other microfluidics technologies described in the previous section have revolutionized the field of single-cell transcriptomics, but these technologies can also be used to study epigenomics at the single-cell level. In 2013, William Greenleaf’s laboratory developed assay for transposase-accessible chromatin using sequencing (ATAC-seq) (Buenrostro et al., 2013). This technique can map accessible chromatin by utilizing bacterial Tn5 transposons pre-loaded with primers that can be used for next-generation DNA sequencing. The transposons insert the primers into regions of open chromatin, and bioinformatics approaches, such as, CENTIPEDE, can be used to identify transcription factor motifs in the open chromatin regions (Pique-Regi et al., 2011). Recently, ATAC-seq has been used for mapping the accessible genome of individual cells (scATAC-seq) using the Fluidigm C1™ programmable microfluidics platform (Buenrostro et al., 2015). Combining scATAC-seq with toxicology can be used to identify transcription factors that change their binding characteristics in the presence of toxicants.

PROTEOMICS TECHNOLOGIES OF TOXICANT-EXPOSED TISSUES

Large scale analysis of gene expression at the protein level will allow unprecedented opportunity to understand toxic mechanisms and/or modes of action and signaling pathways, and identify novel biomarkers of exposure. Post-translational changes, such as phosphorylation, are often induced by toxicants (Caruso et al., 2014). As proteomics technology improves dramatically, smaller and smaller samples can be analyzed for larger and larger numbers of proteins. A top-down approach in proteomics characterizes the entire proteome for changes induced by a toxicant. For example, the changes in the phosphoproteome induced by mercury exposure has been studied in immune cells (Caruso et al., 2014). A PubMed search using the key terms “proteomics” and “toxicology” identifies over 300 published papers going back to 1998. A recent review of the field discusses the role of proteomics to identify adverse outcome pathways for chemical risk assessment (Brockmeier et al., 2017).
MULTIGENERATIONAL AND TRANSGENERATIONAL INHERITANCE OF ADAPTIVE EPIGENETIC CHANGES OF TOXICANT-EXPOSED MODEL ANIMALS AND HUMANS

Studies on epigenetic transgenerational inheritance will impart new information on how, and the mechanisms by which toxicants may affect multi- and trans-generational inheritance of abnormal developmental phenotypes include epigenetic mis-regulation in germ cells. In 2005, Michael Skinner's laboratory published a seminal paper in this field by showing that rats exposed to the pesticide vinclozolin have behavioral alterations for multiple generations (Anway et al., 2005; Guerrero-Bosagna et al., 2010; Crews et al., 2014). In 2013, Bruce Blumberg's laboratory showed that the anti-fouling agent tributyltin (TBT) is a multigeneration obesogen (Chamorro-García et al., 2013; Janesick et al., 2014). Recently, the Blumberg lab extended these studies showing a male obesity phenotype in the F4 generation, making the results truly transgenerational since the germ cells did not have direct exposure to TBT (Chamorro-García et al., 2017). In 2015, we (Ruden) showed evidence for multigenerational epigenetic inheritance in humans by showing that DNA methylation changes associated with maternal exposure to lead can be transmitted to the grandchildren (Sen et al., 2015). Other laboratories are investigating the role of small RNAs in the multigeneration inheritance of the obesity phenotype (Cropley et al., 2016). Carvan et al. have demonstrated that mercury-induced epigenetic transgenerational inheritance of abnormal neurobehavior is correlated with sperm epimutations in zebrafish (Carvan et al., 2017). These are all important studies because they can determine the impact of toxicant exposure on future generations in a non-mutagenic manner.

THE ROLE OF EXTRACELLULAR VESICLES IN TRANSMITTING SIGNALS THROUGHOUT THE ORGANISM AFTER TOXICANT EXPOSURE

Extracellular vesicles (EVs) carry small RNA and protein cargos and have recently been shown to signal to different parts of the body after injury. For example, astrocyte-shed extracellular vesicles have been shown to regulate the peripheral leukocyte response to inflammatory brain lesions in an endocrine-like manner (Dickens et al., 2017). Several studies have characterized exosome cargoes that correlate with different brain diseases (Dorsett et al., 2017; Levy, 2017; Selmaj et al., 2017). There have been very few studies linking toxicant exposures to changes in exosomes, but a recent study did a comparative analysis of microRNA and mRNA expression profiles in cells and exosomes under toluene exposure (Lim et al., 2017). We believe that the role that toxicants have on exosome cargoes for endocrine signaling throughout the body is an important emergent field in toxicology research.

IMAGING OF INTRACELLULAR ALTERNATIONS CAUSED BY TOXICANT EXPOSURE

Toxicants can cause changes intracellularly, such as, by altering the cytoskeleton or damage the mitochondria, lysosomes, or other organelles. Confocal microscopy, and other high-resolution 3D-imaging techniques, can be used to study intracellular alterations caused by toxicant exposure (Fretaud et al., 2017). Intracellular changes can be studied by fluorescent fusion proteins, such as, green fluorescent protein (GFP) fusions (Walmsley, 2008), or fluorescent molecules, such as, quantum dots (Guo and Liu, 2017; Zhu et al., 2017), or fluorescent DNA dyes (Edward, 2012). Advanced microscopic approaches will always be important in developing toxicogenomic research technologies.

TOXICANT-SPECIFIC EQTLS (EXPRESSION QUANTITATIVE TRAIT LOCIS)

Quantitative traits are phenotypes, such as, height or weight, that vary in a population, usually in a normal distribution. Studies on expression quantitative trait loci (eQTLs) will contribute to variation in gene expression levels and changes in SNPs in response to specific toxicants. In 2009, we (Ruden) published the first paper identifying eQTLs that are induced by developmental lead exposure in Drosophila (Ruden et al., 2009). More recent studies by the Mackay laboratory has identified QTLs involved in lead and cadmium toxicity in the Drosophila model (Zhou et al., 2017). The advantages of using Drosophila in toxicogenomics studies are that it has a short generation time (about 10 days) and that over 205 strains have been sequenced in the Drosophila Genetic Reference Panel (DGRP) (Mackay et al., 2012). There are over five million single-nucleotide polymorphisms (SNPs) in the DGRP collection, and the genome is about 10-fold smaller than the human genome, thereby making the SNP-density in Drosophila, about 1 SNP per ∼50 bp, higher than in the entire human sequenced population (5 to 10 SNPs per kb). The DGRP and the power of Drosophila genetics has been underutilized in the toxicogenomics field, but we believe that this model will increasingly show it importance in identifying conserved pathways for toxicant exposure.

EXPOSURE MEASUREMENTS FROM BIOBANKED TISSUES, SUCH AS NEONATAL DRIED BLOOD SPOTS AND SHED TEETH

Every state in the USA collects neonatal dried bloodspots (NDBS) from every child born in that state for screening for genetic diseases. Michigan and California have tens of millions of NDBS stored in biobanks and are available to biomedical research. We (Ruden) have used this resource to correlate...
grandmaternal exposure to lead with the grandchildren’s DNA methylation pattern (Sen et al., 2015). Teeth are an important emerging resource in toxicology research. Manish Arora, who is both a dentist and a toxicologist, have shown that teeth can be used to study the effects of heavy metal exposure during childhood and the later development of schizophrenia (Modabbernia et al., 2016), and autism spectrum disorder (ASD) (Arora et al., 2017). In the ASD study, monozygotic and dizygotic twins discordant for ASD were used to determine whether fetal and postnatal heavy metal exposure increases ASD risk. They found that ASD cases have reduced uptake of essential elements manganese and zinc, and higher uptake of the neurotoxant lead (Arora et al., 2017). It is important to note that teeth can be used to determine gestational exposures because teeth develop during the second trimester, and there is a distinct physical landmark that develops in the teeth at birth. Arora et al. compare teeth to trees with rings, and laser-scanning mass-spectrometry approaches can be used to determine when and how much lead exposure occurs at the resolution of one or a few weeks. More and more biobanks are starting to collect teeth, such as, in the Environmental Influences on Child Health Outcomes (ECHO) biobank (www.nih.gov/echo), to utilize teeth as an important resource in toxicology studies.

**NOVEL METHODS—SECOND- AND THIRD-GENERATION DNA SEQUENCING TECHNOLOGIES IN TOXICOGENOMICS, GWAS, ETC**

New methods are being developed every year that affect the field of toxicogenomics. DNA methylation can be studied by using a chemical trick—bisulfite converts unmethylated cytosines to uracil, but does not convert methylated cytosine as efficiently. However, other DNA modifications have been identified, such as, N6-mA, which can be studied best by single-molecule DNA-sequencing technologies, such as, the Pacific Biosciences SMART-seq™ technology (Flusberg et al., 2010). How toxicants might affect N6-mA or other base modifications is an unexplored area of toxicogenomics research. Other technologies are just beginning to be used in toxicology research, such as, genome-wide association (GWAS) analyses. The Mackay laboratory have used the Drosophila Genome Reference Panel (DGRP) to conduct a GWAS analysis on the toxicology of lead and cadmium (Zhou et al., 2016). GWAS studies in humans have also been used to characterize human genetic susceptibility for environmental chemical risk assessment (Yang et al., 2012; Mortensen and Euling, 2013). More studies such as, these are needed to expand toxicogenomics research in the twenty-first century.

**CONCLUSIONS AND FUTURE STUDIES IN TOXICOGENOMICS**

Perhaps the greatest challenge facing the fields of toxicology and toxicogenomics in the future is to produce, process, curate, archive, and analyze immense genomic, proteomics, and image datasets at the multi-generational, lifespan, and single-cell levels (Figure 1). To take just one example, a single scRNA-seq experiment that analyzes the transcriptome of 100,000 cells ± toxicant exposure can generate in 1 week over 10 terabytes (TB) of raw sequencing data on the NovaSeq™ next-generation DNA sequencing platform from Illumina and can take over a week to align the data to a reference genome on a typical 1,000-node supercomputer cluster (www.illumina.com). In one year, this would generate ~0.5 petabytes (PB) of raw sequencing data. Clearly, this exceeds the resources and computer expertise at most universities, for just this one type of experiment.

In the new and emerging field of toxicogenomics, no single person has the expertise or the time to effectively bring these components together. As technologies continue to improve, success in advancing toxicogenomics will depend on collaborations across large trans-disciplinary and multi-disciplinary groups. Experts who generate genomic and epigenomic data are needed to produce the raw genomic data for subsequent analysis. As the typical equipment for next-generation sequencing is in the $1 million range, it is beyond what an individual investigator can afford. Therefore, core facilities, such as, genomics and proteomics cores, are needed to be expanded and modernized at universities to meet the needs of the entire research community. In this model, research done
FIGURE 1 | Toxicogenomics Experiments in the Twenty-First Century. (A) Multigenerational exposure toxicogenomics involves studying the effects of toxicant exposures in pregnant women (F0) on her babies (F1) and on her grandchildren (F2) who were exposed as germline stem cells in the pregnant mothers. (B) Lifespan exposure toxicogenomics involves studying the effects of toxicant exposures anytime during the life of an organism, from babies to adults to the elderly. (C) Tissue exposure and single-cell RNA-seq (scRNA-seq) involve exposing organisms or 3D organoids to toxicants and then analyzing the gene expression pattern in all of the cells individually. The figure shows a t-Distributed Stochastic Neighbor Embedding (t-SNE) analysis of hypothetical scRNA-seq data. The t-SNE analysis is a technique for dimensionality reduction that is well suited for the visualization of high-dimensional datasets (Taskesen and Reinders, 2016). Each cluster corresponds to one cell type in the tissue being analyzed.

by the individual investigators in their own laboratories will be reduced, as research in core facilities expands.

Bioinformaticians and computer scientists are needed to provide programming and computer science expertise to efficiently process, curate, archive, and analyze vast genomic datasets, and to effectively utilize existing high performance computing resources. Newer and better high performance computing resources need to be improved on an annual basis to meet the expanding demand.

Modelers, mathematicians, and computer scientists will be needed to work with the bioinformaticians to develop new ways to not only process data to a usable format, but also to image and visualize data. For example, we (Ruden) have worked with computer scientists to develop the SnpEff (Cingolani et al., 2012b) software, which is used to characterize the effects of SNPs in whole-genome sequencing data, and has been cited over 1,700 times since it came out in 2012. As of May/June 2017, this highly cited paper received enough citations to place it in the top 1% of the academic field of Molecular Biology & Genetics based on a highly cited threshold for the field and publication year. Also in our laboratory, the SnpSift (Cingolani et al., 2012a) software was developed about the same as SnpEff to identify individual causative SNPs in Drosophila genetic screens. As a third example, CloudAligner (Nguyen et al., 2011), a MapReduce based tool to align read onto reference genome using cloud computing resources, such as the Amazon Cloud™, has been developed in a collaboration between Weisong Shi in the Computer Science Department and our laboratory (Ruden).

It might be more cost effective for most universities to use computing resources on the cloud rather than purchasing their own computer systems. Similarly, it might be more cost effective for investigators to explore companies outside of the university to
FUNDING

This work was supported by National Institutes of Health (NIH) [grant number 5R01ES012933 (DR), 5P30ES020957 (DR), 1U1G3OD023285 (DR), 5R01ES10563 (MA), and 5R01ES07331 (MA)]. NIH Center grant P30 CA022453 to the Karmanos Cancer Institute at Wayne State University, and the Perinatology Research Branch of the National Institutes of Child Health and Development at Wayne State University.

REFERENCES

Anway, M. D., Cupp, A. S., Uzunmcu, M., and Skinner, M. K. (2005). Epigenetic transgenerational actions of endocrine disruptors and male fertility. Science 308, 1466–1469. doi: 10.1126/science.1108190

Arora, M., Reichenberg, A., Willfors, C., Austin, C., Gennings, C., Berggren, S., et al. (2017). Fetal and postnatal metal dysregulation in autism. Nat. Commun. 8, 15493. doi: 10.1038/ncomms15493

Ashokkumar, A. I., Mann, B. K., Prestwich, G. D., and Grainger, D. W. (2012). A 3-D organoid kidney culture model engineered for high-throughput nephropathy assays. Biomaterials 33, 4700–4711. doi: 10.1016/biomaterials.2012.02.063

Baert, Y., De Kock, J., Alves-Lopes, J. P., Söder, O., Stuckenborg, J. B., and Goossens, E. (2017). Primary human testicular cells self-organize into organoids with testicular properties. Stem Cell Rep. 8, 30–38. doi: 10.1016/j.stemcr.2016.11.012

Beauchamp, P., Moritz, W., Kelm, J. M., Ullrich, N. D., Agarvorka, I., Anson, B. D., et al. (2015). Development and characterization of a scaffold-free 3D spheroid model of induced pluripotent stem cell-derived human cardiomyocytes. Tissue Eng. C Methods 21, 852–861. doi: 10.1089/ten.tec.2014.0376

Brockmeier, E. K., Hodges, G., Hutchinson, T. H., Butler, E., Hecker, M., Tollefsen, K. E., et al. (2017). The role of omics in the application of adverse outcome pathways for chemical risk assessment. Toxicol Sci. 158, 252–262. doi: 10.1093/toxsci/kfx097

Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y., and Greenleaf, W. J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat. Methods 10, 1213–1218. doi: 10.1038/nmeth.2688

Buenrostro, J. D., Wu, B., Litzenburger, U. M., Willfors, C., Austin, C., Gennings, C., Berggren, S., et al. (2017). Fetal and postnatal metal dysregulation in autism. Nat. Commun. 8, 15493. doi: 10.1038/ncomms15493

Chamorro-Garcia, R., Sahu, M., Abbey, R. J., Jauede, J., Pham, N., and Blumberg, B. (2013). Transgenerational inheritance of increased fat depot size, stem cell reprogramming, and hepatic steatosis elicited by prenatal exposure to the obesogen tributyltin in mice. Environ. Health Perspect. 121, 359–366. doi: 10.1289/ehp.1205701

Conant, G. E., Eaton, S. A., Aiken, A., Young, P. E., Giannoulato u, E., Ho, J. W., et al. (2016). Male-lineage transmission of an acquired metabolic phenotype induced by grand-paternal obesity. Mol. Metab. 5, 699–708. doi: 10.1016/j.molmet.2016.06.008

Devarasetty, M., Forsythe, S., Shupe, T. D., Soker, S., Bishop, C. E., Atala, A., et al. (2017). Optical tracking and digital quantification of beating behavior in bioengineered human cardiac organoids. Biosensors 7:E24. doi: 10.3390/bios7030024

Dickens, A. M., Tovar-Y-Romo, L. B., Yoo, S. W., Trout, A. L., Bae, M., Kamnoge, M., et al. (2017). Astrocyte-shed extracellular vesicles regulate the peripheral leukocyte response to inflammatory brain lesions. Sci. Signal. 10, eaau7696. doi: 10.1126/scisignal.aau7696

Dorsett, C. R., McGuire, J. L., Niedzielko, T. L., DePasquale, E. A., Meller, J., Floyd, C. L., et al. (2017). Traumatic brain injury induces alterations in cortical glutamate uptake without a reduction in glutamate transporter-1 protein expression. J. Neurotrauma 34, 220–234. doi: 10.1089/neu.2015.4372

Edward, R. (2012). Red/far-red fluorescing DNA-specific anthraquiones for nucl:cyto segmentation and viability reporting in cell-based assays. Methods Enzymol. 505, 23–45. doi: 10.1016/B978-0-12-388448-0.00010-3

Flusberg, B. A., Webster, D. R., Lee, J. H., Travers, K. J., Olivares, E. C., Clark, T. A., et al. (2010). Direct detection of DNA methylation during single-molecule, real-time sequencing. Nat. Methods 7, 461–465. doi: 10.1038/nmeth.1459

Freauté, M., Rivière, L., Job, É., Gay, S., Lareyre, J. J., Joly, J. S., et al. (2017). High-resolution 3D imaging of whole organ after clearing: taking a new look at the zebrafish testis. Sci. Rep. 7:43012. doi: 10.1038/srep43012

García-González, C., Setti, M., Sottile, S., Mancini, M., and Skinner, M. K. (2010). Epigenetic transgenerational actions of vinclozolin on promoter regions of the sperm genome. PLoS ONE 5:e13100. doi: 10.1371/journal.pone.0013100

Gonzalez, M. L., Snyder, B. J., Blumberg, B. S., and Skinner, M. K. (2005). Conduct next-generation sequencing experiments, such as, with the company GeneWiz (www.genewiz.com), which can provide services from library preparation and sequencing to detailed bioinformatics analyses. Clearly, new ways of thinking about how to conduct toxicogenomics experiments is needed as the field develops and new technologies emerge.

AUTHOR CONTRIBUTIONS

DR: Wrote the first draft; KG: Edited and added a section; MA: Edited and commented on all drafts.
