INTEGRATIVE TOXICOGENOMICS: ANALYTICAL STRATEGIES TO AMALGAMATE EXPOSURE EFFECTS WITH GENOMIC SCIENCES

EDITED BY: Pierre R. Bushel and Weida Tong

frontiers Research Topics

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: researchtopics@frontiersin.org

INTEGRATIVE TOXICOGENOMICS: ANALYTICAL STRATEGIES TO AMALGAMATE EXPOSURE EFFECTS WITH GENOMIC SCIENCES

Topic Editors:
Pierre R. Bushel, National Institute of Environmental Health Sciences, United States
Weida Tong, National Center for Toxicological Research, United States
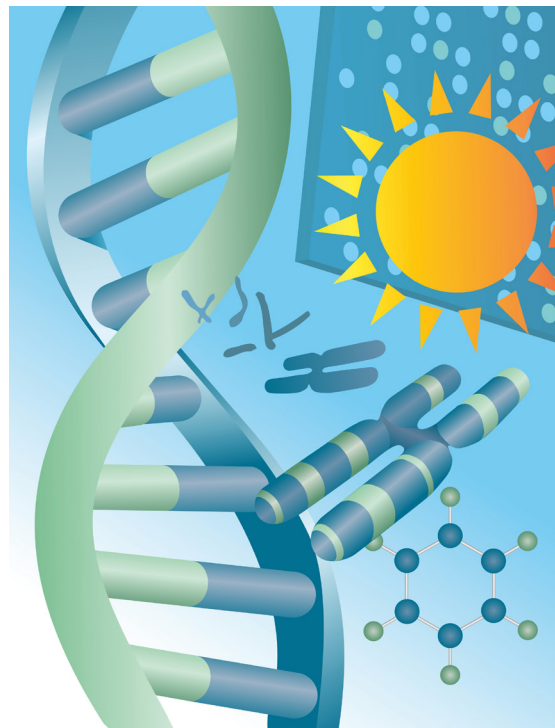
Cover image: Image Associates, Inc.

Toxicogenomics combines the use of toxicology and genomic sciences to elucidate chemical, toxic and environmental stressor effects on biological systems. Integrative toxicogenomics requires innovation in bioinformatics, statistics and systems toxicology and typically a combination of the utility of two of more of these disciplines to better understand molecular mechanisms involved in toxic responses. This Frontiers in Toxicogenomics Research Topic eBook focuses on integrative toxicogenomics more so at the late stage (analyzing each data set separately and then merging the results) and brings together analyses that combine gene expression (microarray, TempO-Seq or RNA-Seq) with other data (biological assays, clinical chemistry, therapeutic categories or molecular pathways) or highlights data analytics that leverage bioinformatics and statistics. The eight articles illustrate the state-of-art

in the field and the analysis of toxicogenomics data for a more comprehensive deduction of biological mechanisms and cellular functions associated with adverse outcomes from environmental exposures, chemicals and toxicants. However, it is clear that the field of integrative toxicogenomics needs considerably more attention paid to it in order to develop other clever ways of integrating the data for analysis.

Table of Contents

frontiers
in Genetics

Editorial: Integrative Toxicogenomics: Analytical Strategies to Amalgamate Exposure Effects With Genomic Sciences

Pierre R. Bushel[1]* and Weida Tong[2]

[1] *Microarray and Genome Informatics Group, Biostatistics and Computational Biology Branch, National Institute of Environmental Health Sciences, Durham, NC, United States,* [2] *Division of Bioinformatics and Biostatistics, National Center for Toxicological Research, U.S. Food and Drug Administration, Jefferson, AR, United States*

Editorial on the Research Topic

Integrative Toxicogenomics: Analytical Strategies to Amalgamate Exposure Effects With Genomic Sciences

Since the advent of toxicogenomics (TGx; Afshari et al., 1999; Nuwaysir et al., 1999), the expectation of its foreseeable impact to risk/safety assessment has given way to the reality that there is much that still needs to be done to realize its full potential. One of the advancements is in the area of integrative TGx which is focused on amalgamating (i.e., merging) diverse data resources with TGx data to gain a comprehensive understanding of toxicology. This requires innovation in bioinformatics, statistics and systems toxicology and typically a combination of the utility of two of more of these disciplines. In addition, integration of data takes place at one of three stages (Bushel, 2016): (1) early—combining separate data into a single matrix and analyzing it, (2) intermediate—keeping the data separate and applying a separate analytic function to each followed by combining of the functions or (3) late—analyzing each data set separately and then merging the results. This Frontiers in Toxicogenomics Research Topic focuses on integrative TGx at the late stage and brings together analyses that combine gene expression (microarray, TempO-Seq or RNA-Seq) with other data (biological assays, clinical chemistry, therapeutic categories or molecular pathways) or highlights data analytics that leverage bioinformatics and statistics. The eight articles illustrate the state-of-art in the field and the amalgamation of TGx data or incorporation of analytical methodologies for a more comprehensive deduction of biological mechanisms and cellular functions associated with adverse outcomes from environmental exposures and toxicants. However, it is clear that the field of integrative TGx needs considerably more attention paid to the merging and analysis of the data at the early and intermediate stages.

Dose-response and time-dependency are key features in toxicology. Nowadays, TGx study designs can easily incorporate at least one of these two features to assess gene expression-based point-of-departure and/or early biomarkers. Zhang et al. assess the toxic effect of doses of Zearalenone on cultured donkey granulosa cells (dGCs) by interrogating gene expression data from RNA sequence (RNA-Seq) analysis and visualization of apoptosis through a tunnel assay. The integrative analysis of the gene expression data, RT-qPCR and immunofluorescence staining of dGCs supports the dysregulation of apoptosis-related genes and induction of ovarian cancer-related genes via the PTEN/PI3K/AKT signaling pathway. Liu et al. leverage a previously developed methodology called pair ranking (PRank) to compare three preclinical rat TGx microarray data

sets for assessment of *in vitro* to *in vivo* extrapolation. They show that there was a high degree of agreement between the *in vivo* assay systems (24 h and 28 days) and similarity between the *in vitro* and the 28 days *in vivo* systems suggesting that a short-term *in vivo* assay system might be practical for some endpoints in order to save time and resources for drug safety evaluation and risk assessment. In addition, Souza et al. reverse engineer gene regulatory networks using *in vitro* human microarray gene expression data to reveal dose-dependent, chemical-specific mechanisms of action in stress-related biological networks. Although gene expression microarrays and RNA-Seq have a solid presence in certain areas of application, TGx studies have begun to explore targeted sequencing using the templated oligo sequencing detection assay (TempO-Seq™). House et al. report a bioinformatics pipeline to analyze dose-response gene expression data profiled with TempO-Seq in induced pluripotent stem cell-based cardiomyocytes. TempO-Seq has drawn significant attention in the TGx community due to its low cost, high throughput and easy to operate.

One of the most extensive applications of TGx is to gain an improved understanding of underlying mechanisms of treatment. While molecular initiative events for many chemicals are well-studied, their modes-of-action (MOAs) remain to be determined. Using RNA-Seq data for the aforementioned purpose can be challenging given the inherent noise. Lozoya et al. presents a leveraged signal-to-noise ratio (LSTNR) thresholding method to identify differentially expressed genes and reveal gene expression patterns especially from samples with very few replicates. Using their method to analyze RNA-Seq rat liver data generated through the MicroArray Quality Control phase III (MAQC3) SEquence Quality Control (SEQC) TGx study, they show that many of the chemicals cluster by MOA and that there are several genes that appear to function as biomarkers specific for chemicals with similar MOA. In addition, Hawliczek-Ignarski et al. investigate whether TGx profiles are able to group chemicals by MOA. As a proof-of-concept study, they tested the hypothesis with data generated after *in vitro* exposure of an established cell line to group chemicals with an uncoupling MOA. Furthermore, Funderburk et al. describe a weighted network analysis of rat liver gene expression from *in vivo* studies involving chemicals from several known MOAs (SEQC TGx study). They demonstrate that overlaps in toxicologic pathways by chemicals with different MOAs (receptor-mediated vs. non-receptor-mediated) reveal points of potential crosstalk between regulatory pathways.

Like mRNAs, microRNAs (short, non-coding RNA molecules roughly 22 nucleotides) are important in toxicology research as well because they are known to regulated gene expression (Bartel, 2004). Bisgin et al. evaluate bioinformatic tools and parameters for RNA-Seq analysis of microRNAs from the livers of rats exposed to thioacetamide at multiple doses and time points. They conclude that variation in the hairpin loop of microRNAs is small relative to the treatment effect and that normalization of the data introduced a large variation in differentially expressed microRNAs. In addition, they indicate that the miRDeep2 analysis tool was preferable over the other choices.

From this small collection of articles, the power and promise of integrative toxicogenomics may seem reassuring. The desire is that as toxicogenomic studies broaden in scope and design, and the complexity of the data increases exponentially, new cutting-edge research and development of novel methodologies to manage, integrate, and analyze massive amounts of data will continue to evolve.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

ACKNOWLEDGMENTS

REFERENCES

Afshari, C. A., Nuwaysir, E. F., and Barrett, J. C. (1999). Application of complementary DNA microarray technology to carcinogen identification, toxicology, and drug safety evaluation. *Cancer Res.* 59, 4759–4760.

Bartel, D. P. (2004). MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 116, 281–297. doi: 10.1016/S0092-8674(04)00045-5

Bushel, P. R. (2016). "Bioinformatics of Genomics in Assessment of Cancer," in *Toxicogenomics in Predictive Carcinogenicity*, eds M. D. Waters and R. S. Thomas (Cambridge: Royal Society of Chemistry), 442–483.

Nuwaysir, E. F., Bittner, M., Trent, J., Barrett, J. C., and Afshari, C. A. (1999). Microarrays and toxicology: the advent of toxicogenomics. *Mol. Carcinog.* 24, 153–159.

frontiers
in Genetics

Dose and Time Dependencies in Stress Pathway Responses during Chemical Exposure: Novel Insights from Gene Regulatory Networks

Terezinha M. Souza*, Jos C. S. Kleinjans and Danyel G. J. Jennen

Department of Toxicogenomics, GROW School for Oncology and Developmental Biology, Maastricht University, Maastricht, Netherlands

Perturbation of biological networks is often observed during exposure to xenobiotics, and the identification of disturbed processes, their dynamic traits, and dose–response relationships are some of the current challenges for elucidating the mechanisms determining adverse outcomes. In this scenario, reverse engineering of gene regulatory networks (GRNs) from expression data may provide a system-level snapshot embedded within accurate molecular events. Here, we investigate the composition of GRNs inferred from groups of chemicals with two distinct outcomes, namely carcinogenicity [azathioprine (AZA) and cyclophosphamide (CYC)] and drug-induced liver injury (DILI; diclofenac, nitrofurantoin, and propylthiouracil), and a non-carcinogenic/non-DILI group (aspirin, diazepam, and omeprazole). For this, we analyzed publicly available exposed *in vitro* human data, taking into account dose and time dependencies. Dose–Time Network Identification (DTNI) was applied to gene sets from exposed primary human hepatocytes using four stress pathways, namely endoplasmic reticulum (ER), NF-κB, NRF2, and TP53. Inferred GRNs suggested case specificity, varying in interactions, starting nodes, and target genes across groups. DILI and carcinogenic compounds were shown to directly affect all pathway-based GRNs, while non-DILI/non-carcinogenic chemicals only affected NF-κB. NF-κB-based GRNs clearly illustrated group-specific disturbances, with the cancer-related casein kinase *CSNK2A1* being a target gene only in the carcinogenic group, and opposite regulation of NF-κB subunits being observed in DILI and non-DILI/non-carcinogenic groups. Target genes in NRF2-based GRNs shared by DILI and carcinogenic compounds suggested markers of hepatotoxicity. Finally, we indicate several of these group-specific interactions as potentially novel. In summary, our reversed-engineered GRNs are capable of revealing dose dependent, chemical-specific mechanisms of action in stress-related biological networks.

Keywords: gene regulatory networks, network inference, toxicity pathways, hepatotoxicity, transcription networks

INTRODUCTION

In the last few years, investigation of biological processes disturbed by chemical exposure and its potential adverse effects has become the main goal in toxicological assessments targeting hazard identification and/or drug development. This strategy encompasses two steps: the first being the identification of such processes/pathways and the second, the estimation of dose–response,

dynamic relationships that may define the boundaries between adaptive and adverse responses (Bhattacharya et al., 2011; Middleton et al., 2017).

Although a substantial amount of mechanistic information has been gained from applying high-throughput (HT) technologies (e.g., transcriptomics, proteomics, and metabolomics) to exposed *in vitro* models, the ability of current methods to address the aforementioned challenges has proven insufficient. Take, for instance, pathway analysis – omnipresent in HT studies as means to identify biological processes affected across different conditions. First, functional annotation does not reflect the diversity of the human genome, with the repertoire of pathways across multiple databases being either comprised of general processes (e.g., metabolism, signaling) or very specific responses (e.g., drug-related pathways) (Khatri et al., 2012). The choice of arbitrary thresholds for fold change and/or statistical significance, as well as stratification of the input gene list by direction of expression are an additional source of variation that influence the output qualitatively and quantitatively.

A more realistic portrayal potentially loaded with novel mechanistic insights can be achieved by reversely engineering gene regulatory networks (GRNs) using expression data; in contrast to pathways, GRNs are case specific, referencing multiple types of physical and biochemical interactions among genes and gene products (Madhamshettiwar et al., 2012), allowing more detailed investigations by not (or minimally) relying on prior knowledge. Previous investigations have attempted to extract novel biological information from GRNs, the majority focusing on the predictive value of (sub)networks and their potential use as biomarkers (Schadt, 2009; Emmert-Streib et al., 2014), which led to the discovery, for instance, of striking dissimilarities between networks of smokers with and without lung cancer (Wang et al., 2011). More than molecular snapshots of a specific phenotype, GRNs are important instruments to investigate the interface genotype–environment (i.e., diet, drugs, and chemical exposure). Early studies with *Saccharomyces cerevisiae* have shown that network interactions undergo critical changes after challenging with a DNA damaging agent, leading to extensive network rewiring (at least 70% out of 80,000 tested genetic interactions) (Bandyopadhyay et al., 2010). Other recent investigations employing low-throughput gene expression data further indicated that gene–gene interactions, although compound specific to some extent, show similar patterns resembling toxic properties across different chemicals – and incorporating interaction data into classification algorithms increase prediction accuracy (Yamane et al., 2016).

Recently, our research group has developed and validated a tool for inferring GRNs from HT gene expression data during chemical exposure, taking possible time and dose dependencies into account (Hendrickx et al., 2016). By using ordinary differential equations (ODEs), this method establishes a causal link between external perturbations and gene–gene interactions within a particular biological process, in addition to identifying potentially novel interactions. In this study, we aimed to extract mechanistic information from chemical-induced toxicity by

reversely engineering GRNs using HT gene expression data. For this, we compare GRNs inferred from groups of chemicals with distinct adverse effects, namely carcinogenicity and drug-induced liver injury (DILI), to those generated by exposure to non-adverse compounds. Through the reconstruction of gene–gene interactions from four stress-related pathways (TP53, ER, NRF2, and NF-κB), we aimed to gather causal evidence for dose dependent, dynamic, and potentially novel biological information related to chemical exposure.

MATERIALS AND METHODS

Dose–Time Network Identification (DTNI) Method

Dose–Time Network Identification (DTNI) is a method for inferring network interactions among genes through ODEs that relate changes in gene expression over time and dose and an external perturbation. DTNI requires measurements from multiple doses and time points – not necessarily sampled at equal intervals – and expression values of a reduced gene set (preferentially less than 100 genes). DTNI can be applied to single chemicals but also allows the use of group-wise approaches – in which a consensus network is inferred for multiple compounds. A detailed description of the method, its validation, and script availability for MATLAB is described elsewhere (Hendrickx et al., 2016).

Chemical Selection

In order to link network changes to chemical effects, we targeted compounds with well-known (adverse) effects in humans. We opted for three groups, two with different mechanisms of toxicity – carcinogenicity and DILI – and one with no weight of evidence for human carcinogenicity or DILI. Since DTNI requires datasets with a minimum of three doses and three time points, we used the Japanese database TG-GATEs (Toxicogenomics Project-Genomics Assisted Toxicity Evaluation system[1]). TG-GATEs contains microarray data of hundreds of compounds, generated *in vitro* (human and rat) and *in vivo* (rat), tested at multiple doses and time points. We therefore mined TG-GATEs for compounds matching the criteria of full availability of sets (i.e., both replicates and all doses/time points) and specific classification regarding carcinogenicity or DILI. To also avoid methodological biases, we aimed to create groups with fairly equal number of compounds; based on these constraints, the carcinogenic group comprised two chemicals [azathioprine (AZA) and cyclophosphamide (CYC)], while DILI (diclofenac, propylthiouracil, and nitrofurantoin) and non-carcinogenic/non-DILI (diazepam, omeprazole, and aspirin) groups contained three chemicals each (**Table 1**). **Table 1** contains detailed information on chemicals and evidence for their inclusion in their respective groups.

---

[1]toxico.nibiohn.go.jp/english/

TABLE 1 | Description of chemical groups used to infer gene regulatory networks (GRNs).

Chemical abstracts service (CAS)	Group	Evidence for inclusion
Azathioprine (446-86-6)	Carcinogenic	Classified as carcinogenic to humans by international agency for research on cancer (IARC)
Cyclophosphamide (6055-19-2)		
Diclofenac (15307-86-5)	DILI	Use is associated with risk of acute liver injury[1]
Nitrofurantoin (67-20-9)		
Propylthiouracil (51-52-5)		
Aspirin (50-78-2)	Non-DILI/non-carcinogenic	Clinical cases of acute liver injury are very rare[1]
Diazepam (439-14-5		
Omeprazole (73590-58-6)		Not classifiable or not classified as to its carcinogenicity by IARC

[1]*Information obtained from LiverTox database (livertox.nih.gov).*

Preprocessing of *In Vitro* Microarray Datasets

Raw files from each chemical set were downloaded from TG-GATEs[2] and preprocessed (background correction, log2-base transformation, and normalization) through R scripts provided on ArrayAnalysis (arrayanalysis.org) (Eijssen et al., 2013). Probes were annotated using customCDF version 19 with Entrez identifiers. To obtain differentially expressed genes (DEGs), we used the R package LIMMA to perform moderated t-test comparing mean intensities from exposed and time-matched controls in all three doses tested. Detailed information from datasets used in this study (accession numbers from microarrays and respective compounds/dose/time points) is available on Supplementary Data 1.

Selection of Biological Networks to Be Assessed by DTNI

Given our goal to identify common features underlying (non-)toxic mechanisms across chemicals, and DTNI's requirement for reduced gene sets, we opted for a group-wise approach for evaluating how carcinogenic, DILI, and non-carcinogenic/non-DILI compounds may (differentially) affect networks involved in toxicological responses. Our selection involves known stress pathways indicative of DNA damage (TP53), oxidative stress (NRF2), endoplasmic reticulum (ER) stress, and inflammation (NF-κB). From KEGG, we retrieved gene components from pathways ER (166 genes, of which 158 were present in normalized sets), TP53 (69 genes, of which 68 were present in normalized sets), and NF-κB (94 genes, of which 89 were present in normalized sets) – entries hsa04141, hsa04115, and hsa04064, respectively. NRF2 genes (143, of which 126 were present in normalized sets) were obtained from WikiPathways (entry WP2884) (the list of genes for each pathway can be found in Supplementary Data 1).

To gain insights into the activity of these pathways in the groups tested, we performed an additional pathway analysis with

all DEGs (FDR < 0.05 without fold change thresholds) from each chemical using the database ConsensusPathDB (CPDB) and its overrepresentation analysis tool (q-value < 0.05) (Kamburov et al., 2013).

Network Inference of Selected Networks through DTNI

For DTNI, we used scripts developed for use on MATLAB (Hendrickx et al., 2016). For this, we used log2-transformed ratios from all chemicals within each group as input file. Then, we performed leave-one-out cross validation (LOOCV) by excluding data from one compound at a time before performing a new DTNI. In all cases, parameters from DTNI were left as default, with a threshold for interaction strength (p-value) set to 0.05. The final network for each pathway within a chemical group was obtained after determining the intersection among the DTNIs (i.e., all chemicals and LOOCVs). The consensus network for NF-κB in the DILI group, for instance, was generated after overlapping the results from DTNI with all three compounds plus LOOCVs (with a total of four different runs). Cytoscape was used to generate and visualize the networks.

Comparison of GRNs across Groups of Chemicals: Biological Significance

To assess differences across networks generated by DTNI, we considered four aspects of the inferred GRNs: direct perturbations to the pathway being analyzed, number of interactions, genes involved, and overall direction of expression (up- or downregulation). Furthermore, while starting nodes (i.e., nodes with only outgoing interactions) may be used as basis to investigate potential molecular initiating events (MIEs), target genes may offer a clue to potential downstream effects. Therefore, those two categories of genes were investigated in more detail within the GRNs.

Furthermore, to validate the interactions found by our method, we used the "induced network modules" tool available on the database CPDB. For this, we used gene lists from each pathway as input. We allowed for intermediary nodes and limited our search to only high-confidence protein interactions in addition to genetic, biochemical, and gene regulatory ones. By comparing our inferred GRNs to those obtained from CPDB, we were able to detect indirect, direct, and potentially novel interactions. A predicted interaction was labeled "direct" when the same edge was present in CPDB and "indirect" when a third node mediated an interaction between two nodes. Indirect interactions may appear due to reactions that happen faster than the timescale measured (e.g., minutes instead of hours) (Hendrickx et al., 2016). Potentially novel interactions were identified as direct interactions predicted by DTNI and not present in CPDB.

---
[2]toxico.nibiohn.go.jp/

RESULTS

Inferred GRNs Vary across Groups of Chemicals with Different Toxicities: Direct Perturbations and Number of Interactions

GRNs reconstructed from gene expression data showed strong variation across carcinogenic, DILI, and non-carcinogenic/non-DILI groups. The number of significant edges (p-value < 0.05) found for each pathway in each chemical group is summarized in **Table 2**. DILI and carcinogenic compounds affected all networks tested, with the former frequently resulting in the highest number of gene interactions. In contrast, non-carcinogenic/non-DILI compounds only affected NF-κB genes. When generating the networks, we observed that those from DILI are more connected in comparison with the other groups. TP53 and ER resulted in sets of unconnected modules in the carcinogenic group. An overview of such network layouts is represented in Supplementary Data 1.

The aforementioned results contrast with those from pathway analysis, which showed that all groups of chemicals affected biological pathways associated with P53 and NRF2 (**Table 3** and Supplementary Data 1). Protein processing in ER was a hit for both Carcinogenic and DILI groups, while NF-κB was significant for non-DILI/non-carcinogenic and DILI groups.

Gene Disturbances in GRNs across Chemical Groups: Target Genes and Starting Nodes

Although input gene lists were identical, the resulting inferred networks showed great variation across chemical groups. To enable visualization and direct comparison, pathways were represented as the union of all interactions predicted in each chemical group, in which edge colors express different chemical groups (red for carcinogenic, blue for DILI, and green for non-DILI/non-carcinogenic) (**Figures 1–4**).

First, we evaluated similarities in interactions from inferred networks, and observed that only NF-κB and NRF2 showed an overlap of two-node interactions in different groups. A negative interaction between *CARD14* and *BCL10* was detected in NF-κB for both DILI and non-DILI/non-carcinogenic groups. For

TABLE 3 | Number of distinct biological processes affected by groups of chemicals related to investigated pathways – results from overrepresentation analysis using differentially expressed genes.

Pathway	Carcinogenic	DILI	Non-carcinogenic/non-DILI
ER	1	1	0
NF-κB	0	2	2
NRF2	3	3	3
TP53	4	4	3

Carcinogenic: azathioprine and cyclophosphamide,
DILI: diclofenac, propylthiouracil, and nitrofurantoin,
Non-carcinogenic/non-DILI: diazepam, omeprazole, and aspirin.

NRF2, we found the negative interactions *GCLM-CYP4A11* and *ABCC5-SLC2A1* shared by DILI and carcinogenic groups.

Since only a few interactions were shared among all groups, we decided to investigate whether this low overlap was due to distinct starting nodes and/or target genes. These are indicated as different sections in the networks, with the upper part containing starting nodes and the lower part the target genes; the latter is further divided into clusters to highlight genes that are specific or shared by more than one chemical group. This arrangement allowed us to detect striking differences among groups. First, we observed that the only pathway that resulted in shared target genes among all three groups was NF-κB – *TRAF6*, a gene from the TNF receptor associated factor family. Also for NF-κB, some genes were also shared by DILI and non-DILI/non-carcinogenic group (*GADD45B* and *BCL10*), as well as DILI and carcinogenic (*CXCL12* and *CARD10*), and carcinogenic and non-DILI/non-carcinogenic (*TNFSF11*, *RIPK1* and *CD14*). NRF2 showed the highest overlap, with 12 genes shared between DILI and carcinogenic compounds: *BLVRB, SLC39A4, HGF, SLC39A3, CBR3, CYP4A11, GSTA1, HSPA1A, SLC2A1, MGST2, SLC5A8,* and *RXRA*. *GTSE1* was the only gene shared between DILI and carcinogenic chemicals for the TP53 pathway, while *EDEM2, DNJB1,* and *DNJA1,* from the ER pathway, were found shared by the aforementioned groups.

Overall, we also detected distinct starting nodes, frequently targeting specific clusters of chemical groups, suggesting that these potential MIEs may result in different responses within the same pathway.

Gene Disturbances in GRNs across Chemical Groups: Validation of Predicted Interactions and Direction of Expression

To validate the edges predicted by DTNI, we compared our results to databases sourcing protein, genetic, and gene regulatory interactions as well as biochemical reactions. The detailed list of edges analyzed is provided in Supplementary Data 1. At least 40 and 30% of all edges predicted for TP53 and NF-κB, respectively, were present in CPDB as direct or indirect interactions. The majority of edges in NRF2 and ER sets, on the other hand, were labeled as potentially novel.

Since direction of expression is also an important feature to understand activation/repression of downstream biological effects, we added expression values to network nodes

TABLE 2 | Number of edges obtained after applying DTNI to gene expression data from primary human hepatocytes exposed to chemicals with distinct toxicity.

Pathway	Carcinogenic	DILI	Non-carcinogenic/non-DILI
ER	22	57	0
NF-κB	20	39	39
NRF2	59	29	0
TP53	8	94	0

Carcinogenic: azathioprine and cyclophosphamide,
DILI: diclofenac, propylthiouracil, and nitrofurantoin,
Non-carcinogenic/non-DILI: diazepam, omeprazole, and aspirin.

FIGURE 1 | Gene regulatory network (GRN) inferred for members from the endoplasmic reticulum (ER) pathway. Red-colored edges: interactions predicted for carcinogenic group; Blue-colored edges: interactions predicted for drug-induced liver injury (DILI) group. Cluster in the upper section indicate starting nodes, while clusters in the lower section are comprised of target genes specific to each chemical group.

FIGURE 2 | Gene regulatory network (GRN) inferred for members from the NF-κB pathway. Red-colored edges: interactions predicted for carcinogenic group; Blue-colored edges: interactions predicted for drug-induced liver injury (DILI) group; Green-colored edges: interactions predicted for non-DILI/non-carcinogenic group. Cluster in the upper section indicate starting nodes, while clusters in the lower section are comprised of target genes specific to each chemical group.

representing the average log2-transformed ratio of each gene. For cross-group comparisons, we partitioned every node into two or three segments, each containing the average expression calculated for the highest dose and latest time point – representing the maximum response across all measurements (**Figures 1**–**4**). With that, we aimed to understand the regulation

FIGURE 3 | Gene regulatory network (GRN) inferred for members from the NRF2 pathway. Red-colored edges: interactions predicted for carcinogenic group; Blue-colored edges: interactions predicted for drug-induced liver injury (DILI) group. Cluster in the upper section indicate starting nodes, while clusters in the lower section are comprised of target genes specific to each chemical group.

of such processes during exposure. Also, by generating individual chemical group/network graphics interchange format files, we confirmed these biological interactions as time- and dose dependent, showing clear oscillatory and/or progressive gene expression profiles (Supplementary Data 2).

In general, we observed great differences in gene regulation across chemical groups, especially in group-specific clusters. Shared genes, on the other hand, usually showed the same direction of expression with variable intensities. Considering global expression, we observed that most genes from TP53 pathway were downregulated after exposure to DILI compounds; the same was observed for NF-κB inferred for non-DILI/non-carcinogenic group. Both therefore suggest repression of these processes. On the other hand, widespread activation of NRF2 was indicated by upregulation of most genes in both carcinogenic and DILI groups, although solute carrier genes *SLC2A2, SLC2A6, SLC2A9, SLC39A10*, found to be target genes only in DILI group, were repressed.

DISCUSSION

The main vision for modern toxicity testing proposes a shift from apical measurements (i.e., pathological modifications related to a disease state) in non-human *in vivo* models toward HT approaches *in vitro*. In the present study, we try to address this challenge with a systematic approach using DTNI, an *in silico* tool that models gene expression changes taking into account dose and time dependencies during chemical exposure.

To investigate the behavior of pathway activation in exposed models, we investigated the effects of chemicals with different adverse outcomes (acute organ injury and carcinogenicity) compared to drugs not implicated in these pathologies. We aimed to reconstruct pathways based on gene expression data; for this, we selected four pathways (NRF2, ER, TP53, and NF-κB) due to their established involvement in mechanisms of toxicity (Jennings, 2013; Jennings et al., 2013). Our first observation was that pathway hits did not relate to direct perturbations to these networks. For instance, even though TP53 and NRF2 were significant in all groups of chemicals, only DILI and carcinogenic compounds showed direct effect (i.e., activation) on these networks. On the other hand, direct perturbation to the NF-κB pathway was only observed for non-DILI/non-carcinogenic compounds. Interestingly, the inhibition of NF-κB, in line of our observation of overall repression of these genes, was confirmed to two components of this group, aspirin and diazepam. Therefore, it seems that in contrast to pathway hits, which may reflect indirect effects or common causes, establishing a causal, direct link between exposure and network perturbations may offer more accurate evidence for the mechanisms of action (Woo et al., 2015).

In addition to different levels of perturbation, we also noticed the distinguishing aspects of GRN composition among groups. This confirms previous findings on the dynamic traits of biological networks, where noticeable rewiring was observed between different disease phenotypes (Mani et al., 2008) and following chemical exposure (Bandyopadhyay et al., 2010), with both studies concluding that genetic interactions may be

FIGURE 4 | Gene regulatory network (GRN) inferred for members from the TP53 pathway. Red-colored edges: interactions predicted for carcinogenic group; Blue-colored edges: interactions predicted for drug-induced liver injury (DILI) group. Cluster in the upper section indicate starting nodes, while clusters in the lower section are comprised of target genes specific to each chemical group.

condition dependent. This was evident in our inferred GRNs, and we suggest that such dissimilarities may be due to distinct initial perturbations, which in turn lead to alternative routes within the same pathway. Our results show that edge paths traced from starting nodes, as well as clusters of target genes, are mostly group specific (**Figures 1–4**). This can be illustrated by the GRNs inferred for genes from the NF-κB pathway (**Figure 2**). Although all groups directly affected this process, GRNs indicate that while non-DILI/non-carcinogenic compounds targeted the inhibition of NF-KB1 and NF-KBIA, those from the DILI group resulted in activation of NF-KB2 and RELA. In addition to differences in direction of expression of these targets, studies have also shown that some NF-κB units may act independently from each other, controlling proliferation and immune responses (Ishikawa et al., 1997, 1998). Therefore, investigation of these branches of biological events may be crucial to differentiate adverse from non-adverse outcomes.

Besides these widespread differences, we also detected some similarities. GRNs inferred for genes from the NRF2 pathway resulted in the highest overlap of target genes with consistent direction of expression between carcinogenic and DILI groups (**Figure 3**). Among these genes, we identified hepatocyte growth factor (*HGF*), known to play a role in tumorigenesis and tissue regeneration (Huh et al., 2004). Recent studies have shown *HGF* to play a role in acute liver injury, by protecting against isoniazid- and rifampicin-oxidative liver damage (Enriquez-Cortina et al., 2013). Another interesting target shared by both groups was enzyme biliverdin reductase B (*BLVRB*), the gene coding, which converts biliverdin to bilirubin. Bilirubin, which has long been

regarded as a cytotoxic waste product of heme metabolism, was recently discovered to possess strong antioxidant activity (Stocker et al., 1987). *CBR3*, a gene coding for an enzyme involved in the biotransformation of carbonyl compounds, was shown to be involved in predisposition of toxic responses in doxorubicin-treated patients (Fan et al., 2008). Taken together, these findings may also suggest that, in an attempt to avoid extreme injury, cells try to compensate by eliciting potent responders; because these mechanisms also have deleterious effects, shifts in their equilibrium may be the tipping point between adaptive and adverse responses.

Interestingly, NRF2 seemed to be the only GRN denoting similar toxic effects in DILI and carcinogenic groups. Perturbations in ER and TP53 pathways by chemicals in the carcinogenic group were very limited (**Figures 1, 4**), comprised mostly by fragmented subnetworks. Since carcinogens AZA and CYC also share DNA-damaging properties, we expected a large amount of disturbances in GRNs from both pathways. In view of this, we hypothesize that due to the different mechanisms underlying their effects (CYC is an alkylating agent while AZA is an antimetabolite), common significant interactions could not be inferred from the available data. The fact that these chemicals have stronger effects on dividing cells may yet be another cause, since human hepatocytes show limited proliferative capacity when cultured (Ramboer et al., 2014).

More importantly, many of the interactions found in our study were labeled as potentially novel ones. During our validation step, we noticed that interactions annotated for the NRF2 pathway in particular were very limited in comparison with the other three

(only 672 against 1518, 1609, and 1539 in ER, NF-κB, and TP53, respectively). The fact that TP53 and NF-κB have been extensively studied over the years due to their clinical relevance may explain why they showed the highest number of direct or indirect positive hits (more than 40 and 30% of all inferred interactions, respectively). Therefore, our results indicate the need for further studies targeting the validation of such interactions as means to expand the repertoire of biological interactions and assess their relevance as potential markers of toxicity.

It should be pointed out some of the limitations of GRN methods, in particular ODE-based methods such as DTNI. Although ODE-based methodology describes time-series data well, its deterministic nature does not account for statistical fluctuations in concentrations and kinetic parameters that greatly influence biological systems (Palmer and Shearwin, 2009). The generation of a consensus network using multiple compounds implemented in DTNI partly solves the problem of data availability, but also constrains chemical selection to substances with somewhat similar effects whose interactions survive after steps of LOOCV. It may also result in increased computational times if a large number of chemicals and input genes are being assessed.

Nonetheless, our approach demonstrated some unprecedented mechanistic aspects of GRNs upon exposure to chemicals with different toxic potential. First, GRNs are usually condition dependent, indicating distinct molecular events depending on the type of exposure. To some extent, however, there are similar gene targets shared by GRNs inferred for toxic groups – but not present in compounds considered non-DILI/non-carcinogenic – that may point toward relevant molecular events indicative of toxicity. Finally, the fact that disturbances in these molecular targets evolve with increases in dose reinforces the value of DTNI as an asset in network-based, HT investigations. Anchoring these dose-dependent events to apical measurements may therefore reveal molecular signatures and clarify the tipping points leading to adverse outcomes.

AUTHOR CONTRIBUTIONS

TS analyzed the data and wrote the manuscript. JK and DJ wrote the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2017.00142/full#supplementary-material

REFERENCES

Bandyopadhyay, S., Mehta, M., Kuo, D., Sung, M.-K., Chuang, R., Jaehnig, E. J., et al. (2010). Rewiring of genetic networks in response to DNA damage. *Science* 330, 1385–1389. doi: 10.1126/science.1195618

Bhattacharya, S., Zhang, Q., Carmichael, P. L., Boekelheide, K., and Andersen, M. E. (2011). Toxicity testing in the 21st century: defining new risk assessment approaches based on perturbation of intracellular toxicity pathways. *PLOS ONE* 6:e20887. doi: 10.1371/journal.pone.0020887

Eijssen, L. M. T., Jaillard, M., Adriaens, M. E., Gaj, S., de Groot, P. J., Müller, M., et al. (2013). User-friendly solutions for microarray quality control and pre-processing on ArrayAnalysis.org. *Nucleic Acids Res.* 41, W71–W76. doi: 10.1093/nar/gkt293

Emmert-Streib, F., Dehmer, M., and Haibe-Kains, B. (2014). Gene regulatory networks and their applications: understanding biological and medical problems in terms of networks. *Front. Cell Dev. Biol.* 2:38. doi: 10.3389/fcell.2014.00038

Enriquez-Cortina, C., Almonte-Becerril, M., Clavijo-Cornejo, D., Palestino-Dominguez, M., Bello-Monroy, O., Nuno, N., et al. (2013). Hepatocyte growth factor protects against isoniazid/rifampicin-induced oxidative liver damage. *Toxicol. Sci.* 135, 26–36. doi: 10.1093/toxsci/kft134

Fan, L., Goh, B.-C., Wong, C.-I., Sukri, N., Lim, S.-E., Tan, S.-H., et al. (2008). Genotype of human carbonyl reductase CBR3 correlates with doxorubicin disposition and toxicity. *Pharmacogenet. Genomics* 18, 623–631. doi: 10.1097/FPC.0b013e328301a869

Hendrickx, D. M., Souza, T., Jennen, D. G. J., and Kleinjans, J. C. S. (2016). DTNI: a novel toxicogenomics data analysis tool for identifying the molecular mechanisms underlying the adverse effects of toxic compounds. *Arch. Toxicol.* 91, 2343–2352. doi: 10.1007/s00204-016-1922-5

Huh, C.-G., Factor, V. M., Sanchez, A., Uchida, K., Conner, E. A., and Thorgeirsson, S. S. (2004). Hepatocyte growth factor/c-met signaling pathway is required for efficient liver regeneration and repair. *Proc. Natl. Acad. Sci. U.S.A.* 101, 4477–4482. doi: 10.1073/pnas.0306068101

Ishikawa, H., Carrasco, D., Claudio, E., Ryseck, R. P., and Bravo, R. (1997). Gastric hyperplasia and increased proliferative responses of lymphocytes in mice lacking the COOH-terminal ankyrin domain of NF-kappaB2. *J. Exp. Med.* 186, 999–1014. doi: 10.1084/jem.186.7.999

Ishikawa, H., Claudio, E., Dambach, D., Raventós-Suárez, C., Ryan, C., and Bravo, R. (1998). Chronic inflammation and susceptibility to bacterial infections in mice lacking the polypeptide (p)105 precursor (NF-kappaB1) but expressing p50. *J. Exp. Med.* 187, 985–996. doi: 10.1084/jem.187.7.985

Jennings, P. (2013). Stress response pathways, toxicity pathways and adverse outcome pathways. *Arch. Toxicol.* 87, 13–14. doi: 10.1007/s00204-012-0974-4

Jennings, P., Limonciel, A., Felice, L., and Leonard, M. O. (2013). An overview of transcriptional regulation in response to toxicological insult. *Arch. Toxicol.* 87, 49–72. doi: 10.1007/s00204-012-0919-y

Kamburov, A., Stelzl, U., Lehrach, H., and Herwig, R. (2013). The ConsensusPathDB interaction database: 2013 Update. *Nucleic Acids Res.* 41, D793–D800. doi: 10.1093/nar/gks1055

Khatri, P., Sirota, M., Butte, A. J., Kim, S., and Woo, H. (2012). Ten years of pathway analysis: current approaches and outstanding challenges. *PLOS Comput. Biol.* 8:e1002375. doi: 10.1371/journal.pcbi.1002375

Madhamshettiwar, P. B., Maetschke, S. R., Davis, M. J., Reverter, A., Ragan, M. A., Hanahan, D., et al. (2012). Gene regulatory network inference: evaluation and application to ovarian cancer allows the prioritization of drug targets. *Genome Med.* 4, 41. doi: 10.1186/gm340

Mani, K. M., Lefebvre, C., Wang, K., Lim, W. K., Basso, K., Dalla-Favera, R., et al. (2008). A systems biology approach to prediction of oncogenes and molecular perturbation targets in B-cell lymphomas. *Mol. Syst. Biol.* 4, 169. doi: 10.1038/msb.2008.2

Middleton, A., Cooper, S., Cull, T., Stark, R., Adeleye, Y., Boekelheide, K., et al. (2017). Workshop report on case studies in cellular stress: defining adversity/adaptation tipping points. *Appl. Vitro Toxicol* 3, 199–210. doi: 10.1089/aivt.2017.0003

Palmer, A. C., and Shearwin, K. E. (2009). Guidance for data collection and computational modelling of regulatory networks. *Methods Mol. Biol.* 541, 337–354. doi: 10.1007/978-1-59745-243-4_15

Ramboer, E., De Craene, B., De Kock, J., Vanhaecke, T., Berx, G., Rogiers, V., et al. (2014). Strategies for immortalization of primary hepatocytes. *J. Hepatol.* 61, 925–943. doi: 10.1016/j.jhep.2014.05.046

Schadt, E. E. (2009). Molecular networks as sensors and drivers of common human diseases. *Nature* 461, 218–223. doi: 10.1038/nature 08454

Stocker, R., Yamamoto, Y., McDonagh, A. F., Glazer, A. N., and Ames, B. N. (1987). Bilirubin is an antioxidant of possible physiological importance. *Science* 235, 1043–1046. doi: 10.1126/science.3029864

Wang, Y.-C., Chen, B.-S., Parkin, D., Bray, F., Ferlay, J., Pisani, P., et al. (2011). A network-based biomarker approach for molecular investigation and diagnosis of lung cancer. *BMC Med. Genomics* 4:2. doi: 10.1186/1755-8794-4-2

Woo, J. H., Shimoni, Y., Yang, W. S., Subramaniam, P., Iyer, A., Nicoletti, P., et al. (2015). Elucidating compound mechanism of action by network perturbation analysis. *Cell* 162, 441–451. doi: 10.1016/j.cell.2015.05.056

Yamane, J., Aburatani, S., Imanishi, S., Akanuma, H., Nagano, R., Kato, T., et al. (2016). Prediction of developmental chemical toxicity based on gene networks

of human embryonic stem cells. *Nucleic Acids Res.* 44, 5515–5528. doi: 10.1093/nar/gkw450

Mode of Action Assignment of Chemicals Using Toxicogenomics: A Case Study with Oxidative Uncouplers

Alessa Hawliczek-Ignarski[1], Peter Cenijn[2], Juliette Legler[2†], Helmut Segner[1] and Jessica Legradi[2*]

[1] Centre for Fish and Wildlife Health, University of Bern, Bern, Switzerland, [2] Environment and Health, VU University Amsterdam, Amsterdam, Netherlands

A criterion frequently used to group chemicals in risk assessment is "mode of toxic action" (MoA). Routinely, structure-based approaches are used for the MoA categorization of chemicals, but they can produce conflicting results or fail to classify compounds. Biological activity-based approaches such as toxicogenomics which provide an unbiased overview of the transcriptomic changes after exposure to a compound may complement structure-based approaches in MoA assignment. Here, we investigate whether toxicogenomic profiles as generated after *in vitro* exposure of an established cell line (C3A hepatoma cells) are able to group together chemicals with an uncoupling MoA, and to distinguish the uncouplers from chemicals with other MoAs. In a first step, we examined whether chemicals sharing the same uncoupling of oxidative phosphorylation (OXPHOS) MoA produce similar toxicogenomic profiles and can be grouped together. In a next step, we tested whether the toxicogenomic profiles discriminate between OXPHOS and chemicals displaying a (polar) narcotic MoA. Experimentally, cells were exposed *in vitro* to equipotent concentrations of the test compounds and gene expression profiles were measured. The resulting toxicogenomic profiles assigned OXPHOS to one cluster and discriminated between the OXPHOS and the (polar) narcotics. In addition, the toxicogenomics data revealed that one and the same chemical can display multiple MoAs, which may help to explain conflicting results of MoA classification from structure-based approaches. The results strongly suggest the feasibility of MoA grouping of chemicals by using *in vitro* cell assay-based toxicogenomic profiles.

Keywords: OXPHOS, *in vitro*, toxicogenomics, microarray, mode of action, uncoupler, chemical risk assessment

INTRODUCTION

Chemical risk assessment has to deal with a large and ever growing number of chemical substances. At the same time, increasing regulatory demands that are designed to ensure the health and safety of both humans and ecosystems lead to increasing needs for chemical toxicity testing. In this situation, a compound-by-compound evaluation using traditional toxicity testing approaches is hardly feasible given the resources needed in terms of time, labor, and costs, and with respect to

ethical considerations, as it would necessitate the use of extensive numbers of experimental animals. Thus, toxicologists are faced with the challenge to develop novel approaches of chemical hazard assessment which reduce testing needs and animal usage while still generating robust safety data (Mantus, 2007; NRC, 2007; ECHA, 2011; Kavlock et al., 2012).

One possibility to increase efficiency of chemical hazard assessment is to categorize chemicals. In such a group-wise evaluation the unknown toxicity of untested compounds can be inferred from the known toxicity of tested compounds. A criterion that is frequently used to group chemicals is their "mode of toxic action" (MoA). MoA is a loosely defined term used in both human toxicology and ecotoxicology, In this article, MoA is referred to as key toxic processes (e.g., chemical binding to a nuclear receptor) (Escher and Hermens, 2002; Vonk et al., 2009; Carmichael et al., 2011; Kienzler et al., 2017). The question is how to recognize that two chemicals have a common MoA, and therefore can be grouped together? One possibility is to assign MoA based on molecular interactions between the chemical and the biological system which initiate specific toxicity pathways like binding to the estrogen receptors, or measuring specific cellular reactions that are indicative for specific effects such as chlorophyll interference in tests with algae (Hamadeh et al., 2002; Escher et al., 2005; Nendza and Wenzel, 2006; Woods et al., 2007; Pereira-Fernandes et al., 2014). Another approach is the grouping of chemicals on the basis of structural rules (Schüürmann, 1998; Verhaar et al., 2000; Enoch et al., 2008; Blackburn et al., 2011). The chemical similarity principle postulates that chemicals with similar structure are likely to be toxicologically similar because they act through similar MOAs (Martin et al., 2002; Enoch et al., 2008). As recently reviewed by Kienzler et al. (2017), various structure-based classification schemes have been developed to group chemicals according to their MoAs, including the Verhaar scheme which classifies chemicals on the basis of correlations between apical toxic endpoints such as lethality and chemical descriptors like octanol-water partition coefficient (kow) (Verhaar et al., 1992), or the ASTER scheme of the US EPA which utilizes, among others, fish behavioral responses for MoA classification (Barron et al., 2015). The results of the different existing classifications schemes may differ. For instance, when comparing the outcome of three MoA classification schemes for 3,448 chemicals, Kienzler et al. (2017) found that only 432 out of the 3,448 chemicals were classified by all three schemes, and only for 42% of these 432 chemicals, the three schemes produced concordant MoA classifications. This illustrates that the structure-based approaches for MoA classification, although being highly useful, also have inherent limitations. Therefore, it has been suggested to extend the structure-based approaches to biological descriptors (Richard, 2006).

The present study examines the potential of toxicogenomic information to augment MoA assignment. Toxicogenomics has been selected as an experimental approach as it enables unbiased, global profiling of the molecular biological response space (Richard, 2006; Edwards and Preston, 2008), without pre-selection of response parameters. For single chemicals, it has been shown that exposure results in chemical-specific, gene expression profiles (Hamadeh et al., 2002; Yang et al., 2007; Kiyosawa et al.,

2010; Janssens et al., 2011; Piña and Barata, 2011; Meganathan et al., 2012; Theunissen et al., 2012). The hypothesis tested in the present study is whether specific MoAs have common gene expression profiles that can be used to classify compounds and whether compounds with multiple MoAs can be classified and distinguished from those with only one MoA. As the MoA to be tested in the present study, we selected uncoupling of oxidative phosphorylation (OXPHOS). The biological mechanism of uncouplers is relatively well understood (Terada, 1990; Spycher et al., 2005): they act as protonophores, which dissipate the pH gradient across the proton-impermeable energy-transducing membranes of mitochondria. As a consequence, phosphorylation is uncoupled from the electron transport and from hydrogen transfer. Uncoupling is normally assessed monitoring changes in respiration rate or membrane potential of extracted liver mitochondria or cells (Brand and Nicholls, 2011). Here, we test a number of chemicals which have been classified as uncouplers on the basis of biological mechanistic studies and/or chemical descriptors, and evaluate how similar or dissimilar their toxicogenomic profiles are, and whether these profiles can be discriminated from profiles of chemicals with other MoAs. In order to comply with the aim of reducing animal experimentation in toxicity testing (Pereira-Fernandes et al., 2014), we performed the toxicogenomic studies in an *in-vitro* test system, the human liver HepG2/C3A cell line.

MATERIALS AND METHODS

Chemicals

All solvents were of pro analysis (p.a.) quality or better, and purchased from Dr Grogg Chemie AG (Stettlen-Deisswil, Switzerland) unless stated otherwise. Aniline (99%, CAS 62-53-3, Product 132934), Benzo(a)pyrene (BaP) ($\geq$ 96%, CAS 50-32-8, Product B1760), 4-Chloroaniline (Cl-aniline) (98%, CAS 106-47-8, Product C22415), Dimethyl sulfoxid (DMSO) (99.5%, CAS 67-68-5, Product D4540), Ethylenglycol (1,2-Ethanediol, analytical standard, CAS 107-21-1, Product 85978), Pentachlorophenol (PCP) (98%, CAS 71-23-8, Product 402893), 2,3,4,5-Tetrachlorophenol (TCP) (99.9%, CAS 4901-51-3, Product 442281), were purchased from SIGMA-ALDRICH (Buchs, Switzerland). Carbonyl cyanide 3-chlorophenylhydrazone (CCCP) (99+%, CAS 555-60-2, Product L06932) was purchased from Alfa Aesar (Karlsruhe, Germany). Carbonyl cyanide 4-(trifluoromethoxy)phenylhydrazone (FCCP) (>99%, CAS 370-86-5, Product 0453) was purchased from TOCRIS (Missouri, USA). 6OH-BDE90 was synthesized in the Department of Materials and Environmental Chemistry, Environmental Chemistry Unit, Stockholm University, Sweden (Marsh et al., 1999, 2003). If needed, stock solutions and dilutions of the test substances were prepared in DMSO as vehicle solvent and were added to the culture to give a final DMSO concentration of 0.1% at maximum.

Cell Culture Maintenance

C3A hepatoma cells were obtained from LGC Standards (HepG2/C3A, Hepatocellular Carcinoma, Human (*Homo sapiens*), ATCC-CRL-10741, Middlesex, UK). C3A is a clonal

TABLE 1 | Uncoupler of the oxidative phosphorylation candidate compound list.

Name	CAS	Paper		
1-Naphthol	90-15-3			Ru
2,3,4,5-tetrachlorophenol	4901-51-3	S	G	
2,3,4,5-Tetrachlorophenol	4901-51-3			Ru
2,3,4,6-tetrachlorophenol	58-90-2	S		Ru
2,3,5,6-tetrachlorophenol	935-95-5		G	
2,3,5,6-Tetrafluorophenol	769-39-1		G	
2,3-dinitrophenol	66-56-8		G	
2,4,5-trichlorophenol	95-95-4	S		
2,4,6-tribromophenol	118-79-6	S		
2,4,6-Trichlorophenol	88-06-2	S		
2,4,6-trinitrophenol	88-89-1	S	G	
2,4-dichloro-6-nitrophenol	609-89-2		G	
2,4-dichlorophenol	120-83-2	S		
2,4-dinitrophenol	51-28-5	S	G	Ru
2,5-dinitrophenol	329-71-5		G	
2,6-dibromo-4-nitrophenol	99-28-5	S	G	
2,6-dichloro-4-nitrophenol	618-80-4		G	
2,6-dichlorophenol	87-65-0			Ru
2,6-diiodo-4-nitrophenol	305-85-1		G	
2,6-dinitro-4-cresol	609-93-8		G	
2,6-dinitrophenol	573-56-8	S	G	Ru
2-sec-butyl-4,6-dinitrophenol	88-85-7	S		Ru
2-tert-butyl-4,6-dinitrophenol	1420-07-1	S		
3,4,5,6-tetrabrom-2-cresol	576-55-6		G	
3,4,5-Trichlorophenol	609-19-8	S		
3,4-dichlorophenol	95-77-2	S		
3,4-dinitrophenol	577-71-9	S	G	
3,5-dibromo-4-hydroxy-benzonitrile	1689-84-5	S		
3,5-dibromo-4-methylphenol	13979-81-2	S		
3,5-dichloro-4-hydroxybenzonitril	1891-95-8			Ru
3,5-dichlorophenol	591-35-5	S		
3,5-diiodo-4-hydroxybenzonitril	1689-83-4			Ru
3-chloro-N-[3-chloro-2,6-dinitro-4 (trifluoromethyl) phenyl]-5-(trifluoromethyl)-2-Pyridinamin	79622-59-6	S		
3-Trifluoromethyl-4-nitrophenol	88-30-2			Ru
4,5,6,7-tetrabromo-2-(trifluoromethyl)-benzimidazol	2338-30-9	S		
4,5,6,7-tetrachloro-2-(trifluoromethyl)-benzimidazol	2338-29-6	S		
4,6-dinitro-2-cresol (4,6-dinitro-o-cresol)	534-52-1		G	
4,6-dinitro-o-cresol	534-52-1	S	G	Ru
4-methyl-2,6-dinitrophenol	609-93-8	S		
4-Nitro-2-trifluoromethylbenzimidazole	14689-51-1	S		
4-Nitrobenzamide	619-80-7			Ru
4-Nitrophenol	100-02-7	S		
4-tert-butyl-2,6-dinitrophenol	4097-49-8	S		
5,6-dichloro-2-(pentafluoroethyl)-benzimidazol	102516-93-8	S		
5,6-dichloro-2-(trifluoromethyl)-benzimidazol	2338-25-2	S		
5-chloro-2-(2,4-dichlorophenoxy)-phenol	3380-34-5	S		
5-nitro-2-trifluoromethylbenzimidazole	327-19-5	S		

(Continued)

TABLE 1 | Continued

Name	CAS	Paper		
6-Chloro-2,4-dinitrophenol	946-31-6	S		
Carbonyl_cyanide_m-chlorophenylhydrazone	555-60-2	S		
Carbonyl-cyanide-p-methoxyphenylhydrazone	370-86-5	S		
Pentabromophenol	608-71-9		G	
Pentachlorophenol	87-86-5	S	G	Ru
Pentachloropyridine	2176-62-7			Ru
Pentafluorophenol	771-61-9		G	
p-phenylazophenol	1689-82-3			Ru

Each compound is a designated uncoupler of the oxidative phosphorylation based on structural alerts. The column "paper" refers to the corresponding published paper (Russom et al., 1997; Schüürmann et al., 2003; Spycher et al., 2008b).

derivative of the HepG2 cell line. Cells were cultured in 75-cm^2 vent cell culture flasks (TPP$^+$, Trasadingen, Switzerland) in Minimal Essential Medium with phenol red indicator (MEM, Gibco, Product 31095, Zug, Switzerland) adding 10% FCS (Fetal Calf Serum, Gibco, Product 10290, Zug, Switzerland), 1% NEAA (MEM Non-Essential Amino Acids, Gibco, Product 11140-035, Zug, Switzerland), 1 mM Sodium Pyruvat (Sigma-Aldrich, CAS 113-24-6, Product P5280, Buchs, Switzerland), and 1% Penicilin-Streptomycin (Penn-Strep) (Sigma-Aldrich, Product P4333, Buchs, Switzerland). Incubation temperature was set at 37°C. Cells were subcultivated (split 1:2) twice weekly. The subculture procedure is described elsewhere (Marsh et al., 1999, 2003; Whitmore et al., 2000). 0.05% trypsin-EDTA was used to detach cells from the flask (Trypsin, Gibco, Product 15400-054, Zug, Switzerland).

For toxicity testing, C3A cells were grown in 75 cm^2 culture flasks until ∼80% confluency. Cells were then subcultured as quoted above until the uptake of trypsinised cells in media. C3A cells were diluted to reach a final density of 250 000 cells/ml. Two hundred microliters of cell suspension were seeded in the inner 60 wells of a black 96-well plate with clear bottom (Greiner bio-one, Product 655906, Reinach, Switzerland). The outer rows were filled with 200 µl 1x PBS (phosphate buffered saline, Sigma-Aldrich, Buchs, Switzerland). Cells were pre-incubated at 37°C for 24 h. After pre-incubation cells were treated with the test chemicals. Therefore, the test substances were dissolved in DMSO and added to culture medium at a final concentration of 0.1% DMSO to reach the desired test concentration in the test-medium.

Tetramethylrhodaminemethylester (TMRM)-Assay

Tetramethylrhodaminemethylester (TMRM) is a positively charged dye, exhibiting a red color. In cells with intact membrane potential, TMRM accumulates in the active mitochondria, due to their relative negative charge. When the membrane potential collapses, TMRM is no longer retained in the mitochondria and the observed fluorescence decreases. Culture medium was aspirated from the cells and 200 µl exposure medium

TABLE 2 | Compounds tested in the different bioassyas (TMRM, TPP$^+$ and Cytotoxicity), [a]biological proven uncoupler, [b]QSAR designated uncoupler.

Compound	MoA	logK$_{OW}$	TMRM assay EC$_{50}$ in µM	TPP$^+$ assay LOEC in µM	Cytotoxicity EC$_{50}$ in µM
CCCP	Uncoupler[a,b]	3.4	5.5	0.01	63.5
FCCP	Uncoupler[a,b]	3.7	1.2	0.5	26.9
TCP	Uncoupler[a,b]	4.45	9.3	0.5	25.7
PCP	Uncoupler[a,b]/narcotic	5.1	117.8	0.5	224.8
6OH-BDE90	Suspected uncoupler	6.8	n.t.	1.3	3.5
Ethylenglycol	Polar narcotic	−1.36	–	–	220
Cl-aniline	Narcotic	1.83	amb	–	9,100
AnilineAniline	Narcotic	0.95	amb	–	25,800
BaP	Receptor mediated		n.t.	n.t.	n.t.

n.t., not tested; amb., ambiguous results; –, no effects.

were gently added. After 2 h incubation, exposure medium was aspirated and TMRM medium was gently added to the wells. The TMRM medium was prepared by adding TMRM [Tetramethylrhodaminemethylester (perchlorate), 99%, CAS 115532-50-8, Product 88065, Anawa, Zürich, Switzerland] at a final concentration of 125 nM to serum- and phenol red-free culture medium (MEM without phenol red, Product 51200, Gibco, Zug, Switzerland). The cells were incubated in TRMR medium for 20 min. Afterwards, TMRM medium was aspirated, and 100 µl of phenol red free medium was added. Fluorescence was measured instantly using the EnSpire® microplate reader (Perkin Elmer, Schwerzenbach, Switzerland) at 540 nm excitation wavelength and 575 nm emission. Solvent controls (0.1% DMSO), positive controls (2.5 µM FCCP) and medium controls were included in all experiments. Per test concentration six well were measured and each experiment was repeated at least three times.

Triphenylphosophonium (TPP$^+$) Assay

The TTP+ assay is measuring changes in the inner mitochondrial membrane potential ($\Delta\Psi$) and mitochondrial respiration (O$_2$) in isolated rat liver mitochondrial with help of tetraphenylphosphonium (TPP$^+$) and oxygen sensitive electrodes. The amount of free TPP$^+$ in the reaction medium is a relative measurement of $\Delta\Psi$. The experimental animals were sacrified according to ethical procedures and the rat livers were immediately collected and kept in cold 0.9% NaCl$_{(aq)}$. The tissue was homogenized in isolation medium (0.25 M sucrose, 10 mM HEPES, 3 mM EGTA, 0.2% BSA) and then centrifuged for 6 min at 3,000 rpm 750 g at 4°C (one liver in 50 ml isolation medium). The supernatant was transferred to a clean tube and was centrifuged for 10 min at 10,000 g. Supernatant was removed, and the pellet washed with wash medium (0.25 M sucrose, 5 mM HEPES, 0.1% BSA) and centrifuged for 10 min at 10,000 g. The pellet was resuspended in a small volume of wash medium. For simultaneous measurement of respiration and inner mitochondrial membrane potential ($\Delta\Psi$), the mitochondria were incubated at 25°C in reaction medium (85 mM KCl, 20 mM HEPES, 5 mN KH2PO4, 2.3 mM MGCl2, 25 mM creatine, 25 mM phosphocreatine) in a closed and stirred perspex vessel equipped with an oxygen electrode and a TPP$^+$-sensitive electrode (Kamo et al., 1979). Coupling between respiration (O$_2$), inner membrane potential ($\Delta\psi$) and ATP (Adenosine triphosphate) production in the isolated mitochondria was demonstrated by addition of succinate (1 mM) and ATP (10 mM) to initiate oxidative phosphorylation. Rotenone was also added to block complex I (NADH-UQ reductase) of the electron transport chain (ETC), which allowed us to study inhibitory effects on the ETC as well as protonophoric uncoupling. All test compounds (dissolved in DMSO) were injected into the vessel after the membrane potential was stable (i.e., when state 4-respiration had been reached). For each compound, different concentrations were tested. Due to the complexity of the test $n = 1$ for the LOEC determination.

Cytotoxicity assay

The cytotoxic action of the test compounds on C3A cells was measured using the resaruzine dye which assesses cell viability. Therefore, the amount of resazurine (non-toxic, non-fluorescent redox active blue dye) that is reduced to resorufin (non-toxic, highly fluorescent, pink dye) by metabolic activity in the cells was measured fluorometrically (Abu-Amero and Bosley, 2005). Therefore, culture medium was aspirated from the cells and 200 µl exposure medium was gently added. Total volume reached 200 µl. After 22 h of exposure, 10 µl of 440 µM resazurine (resazurine sodium salt, CAS 62758-13-8, Product R7017, Sigma-Aldrich, Buchs, Switzerland) was added to the wells. Fluorescence (excitation wavelength 530 nm and emission wavelength at 590 nm) was measured immediately (T$_0$) using EnSpire® microplate reader. After 2 h, fluorescence was measured again (T$_2$). Background fluorescence was subtracted by applying the equation T$_{2h}$-T$_{0h}$ for each well. Exposure was carried out in six replicates per concentration. Finally, the T$_{2h}$-T$_{0h}$ fluorescence of each replicate was averaged and the standard deviation calculated. Solvent control (0.1% DMSO) and positive control (0.03% H$_2$O$_2$), as well as a negative control (medium) were incorporated in all experiments. Per test concentration six well were measured and each experiment was repeated at least three times.
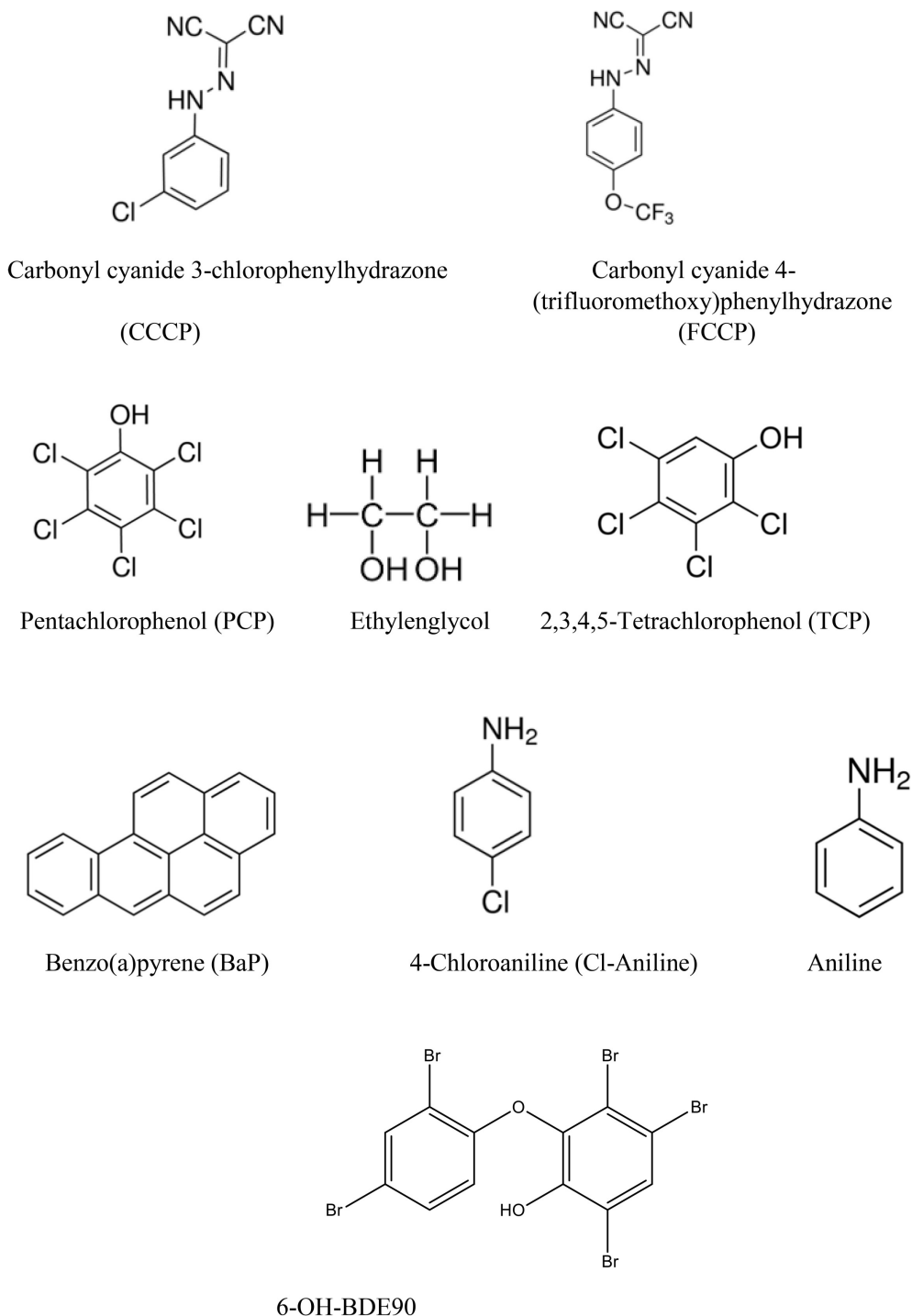
FIGURE 1 | Chemical structures of all the compounds studied.

Microarray Experiment

C3A cells were exposed for 2 h to substances as described above for the TMRM assay. For microarray analysis, six replicate wells per concentration were pooled. DMSO (0.1%) was used as solvent control. The concentrations used for the exposure were:

Microarray 1: calculated EC_{50} values from the 2 h TMRM assays, BaP 10 μM.

Microarray 2: calculated EC_{25} values from the 24 h cytotoxicity assays, i.e., CCCP: 22.8 μM; FCCP 15.7 μM; TCP: 13.8 μM; PCP: 101.7 μM; 6OH-BDE90: 4.2 μM;

FIGURE 2 | Dose-response curves for single compounds in the TPP$^+$ assay. Measurement of respiration rate (V) vs. mitochondria membrane potential (MMP) **(Left)**. Each point represents the measurements after compound was added to the vessel. After addition of compound the membrane potential decreases (non-disrupted MMP ~180 mv) and the oxygen consumption increases (uncoupling). At higher concentrations, the oxygen consumption starts to decrease indicating inhibition. Measurement of respiration rate (V) vs. concentration of compound **(Right)**. After addition of compound the oxygen consumption increases (uncoupling). At higher concentrations, the oxygen consumption starts to decrease indicating inhibition.

FIGURE 3 | Dose-response curves for single compounds in the cytotoxicity assay **(Left)** and the TMRM assay **(Right)**. For each experiment six wells per test concentration were averaged and each experiment was repeated three times. Data points present the mean over all measurements and the bars the standard deviation. TMRM graphs for Aniline and Cl-aniline where ambiguous, instead of a decrease an increase of the TMRM signal could be observed. 6OH-BDE90 was not measured in the TMRM assay.

Aniline: 44.66 mM; Cl-aniline: 6.54 mM; Ethylenglycol: 558.8 mM.

Total RNA was extracted using RNeasy Mini Kit (QUIAGEN, Product 74104, Hombrechtikon, Switzerland) according to manufacturers' instructions. DNA was digested using the RNeasy Mini Kit optional on-column DNase digestion. Quality of extracted RNA was checked via the Nanodrop 1000 spectraphotomoeter V3.7, as well as running the samples on a 1.5% agarose gel.

Total RNA was reverse transcribed to cDNA using 1 µg RNA Template. The reaction mix per sample had a total volume of 25 µl. The reaction mix contained the following products purchased from Promega AG (product number in brackets) (Promega AG, Dübendorf, Switzerland): 0.5 µl RNasin® Plus RNAse Inhibitor (N2611), 1.25 µl 10 mM dNTP (U1201, U1211, U1221, U1231), 1 µl M-MLV Reverse transcriptase (RNase H Minus, Point Mutant) (M3682), 5 µl of 5x Buffer (M3682), 1 µl random primer (C1181), X µl DEPC-water to reach the final volume of 25 µl. After reverse transcription, samples were diluted 10x to prevent inhibition of the PCR reaction by the cDNA synthesis buffer. The diluted cDNA was stored at −20°C until further usage.

For Microarray 1: Each treatment consisted of two biological replicates, the control consisted of three biological replicates. The RNA 6000 NanoChip kit (Agilent Technologies, number 506791511) was used with the Agilent 2100 Bioanalyzer (Agilent technologies) for analysis of total RNA samples. Samples with a RIN (RNA integrity number) above 7 were considered appropriate for consequent microarray testing. For Microarray analysis, Agilent SurePrint G3 Human Gene

Expression Microarrays (8x60K) were used in combination with a one-color based hybridization protocol (OneColor RNA Spike-In Mix, Agilent Technologies, number 51885282). All steps were carried out according to the manufacturer's instructions. Fluorescent signal intensities were detected with Scan Control A.8.4.1 Software (Agilent Technologies) on the Agilent DNA Microarray Scanner and extracted from the images using Feature Extraction 10.7.3.1 Software (Agilent Technologies). A quality check as described by the manufacturer was performed (e.g., SpikeIn controls).

For Microarray 2: In order to analyze more compounds in a comparative way we changed that microarray approach as following. Each treatment and controls consisted of three biological replicates. The RNA 6000 NanoChip kit (Agilent Technologies, number 506791511) was used with the Agilent 2100 Bioanalyzer for analysis of total RNA samples following the manual instructions. Samples with a RIN (RNA integrity number) above 7 were considered appropriate for consequent microarray testing. For Microarray analysis Agilent SurePrint G3 Human Gene Expression Microarrays (8x60K) were used (GPL14550). All replicates of the treatments were labeled with Cy5 and all corresponding solvent controls with Cy3. After hybridization, the microarrays were washed using Gene

TABLE 3 | Range of the test concentrations for Ethylenglycol, Cl-aniline, and Aniline which did not induce any effects on the oxygen consumption or membrane potential (data not shown).

Ethylenglycol	tested from 6.3 µM up to 260.5 mM
Cl-aniline	tested from 6.3 µM up to 3.8 mM
Aniline	tested from 3.4 mM up to 85 mM

TABLE 4 | Correlation coefficient between treatments for microarray experiment 1.

	BAP	CCCP	FCCP	TCP
A				
BAP	1.00	−0.13	−0.25	−0.14
CCCP	−0.13	1.00	0.83	0.78
FCCP	−0.25	0.83	1.00	0.72
TCP	−0.14	0.78	0.72	1.00
B				
BAP	1.00	−0.05	−0.07	0.07
CCCP	−0.05	1.00	0.90	0.81
FCCP	−0.07	0.90	1.00	0.75
TCP	0.07	0.81	0.75	1.00

(A) For the 5000 genes with the highest variation. **(B)** For the most regulated genes with a fold change of >2 or <0.5 (|M| > 1). Correlation coefficients above 0.7 are marked with gray background.
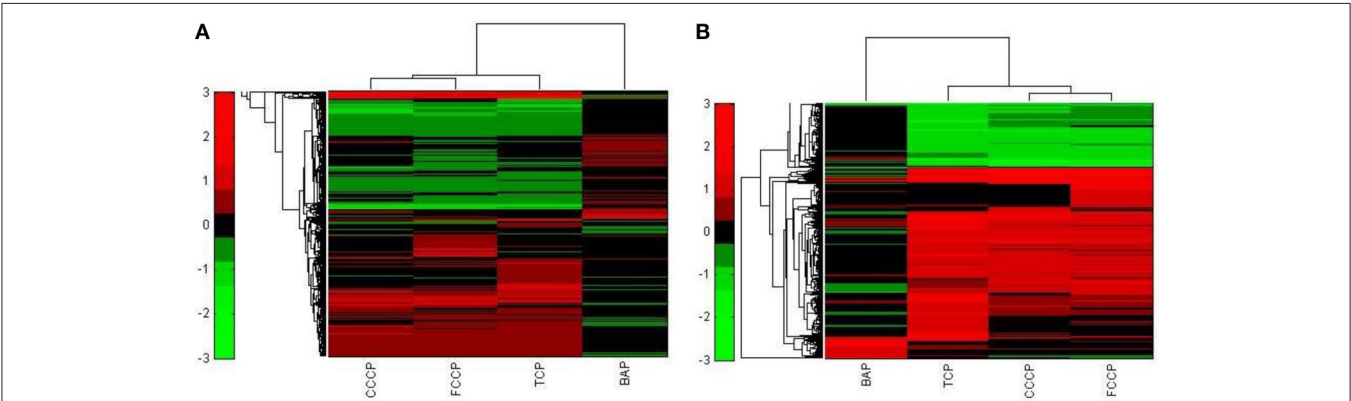
FIGURE 4 | Hierarchical cluster analysis of C3A cells exposed to 3 compounds (FCCP, CCCP, and TCP) with a known uncoupling MoA and BaP). **(A)** The hierarchical cluster analysis is based on the 5000 most varying genes. **(B)** The hierarchical cluster analysis is based on the significantly regulated genes (|M| > 1).

FIGURE 5 | Principle component analysis (PCA) of the 1st microarray experiment **(Left)** and the 2nd microarray experiment **(Right)**. Clear differences between uncoupler and non-uncoupler can be observed.

FIGURE 6 | Venn diagram showing the overlap of genes that were regulated with a fold change of >2 or <0.5 ($|M| > 1$) between treatments of microarray experiment 1. Only five genes overlap between all treatments.

Expression Wash Buffer Kit (Agilent Technologies), and scanned using the Scan Control Software A8.5 with the feature extraction 10.X Software (Agilent technologies) on the DNA microarray scanner G2505C (Agilent Technologies). More detailed description of the design of the microarrays (platform) is available from Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo/) under accession number GPL13607. All microarray data from this study have been deposited in NCBI's Gene Expression Omnibus under the accession number GSE75784. A quality check as described by the manufacturer was performed (e.g., SpikeIn controls).

Microarray Data Analysis

The microarray data was normalized, filtered, and biological replicates averaged. The complete microarray analysis was done using MatlabR2013a.

For the first microarray study a general quality check of the raw array data was performed (spot integrity, signal distribution). The data was normalized using quantile-normalization. Control spots as well as not uniform spots were removed. The solvent control was averaged and the M value (log_2FoldChange) calculated. Duplicated spots on the array were averaged before exposure replicates were averaged using the median. This data set was used for further analysis as described in the results.

For the second microarray study a general quality check of the raw data was performed. M values (log_2FoldChange) were calculated and LOEWESS and centering normalization applied. Non-uniform and control spots were removed from the data set before duplicated spots on the array were averaged (Median). Then exposure replicates were averaged using median. To identify significantly regulated genes, genes expressed $|M|>2$ were selected and a t-test performed. Adjusted p-values were calculated using the Bonferroni-Hochberg correction and a cut-off of 0.1 applied. The Pearson correlation coefficient between normalized and averaged replicates was calculated using MaltlabR2013a.

Data Analysis (Cytotoxicity and TMRM)

For the TMRM assay as well as for the cytotoxicity assay, the fluorescence was assessed as described above and dose-response curves were fitted using GraphPad Prism 5 (GraphPad Software Inc., Ca, USA). For dose-response curves, a sigmoidal curve with the formula $Y = A + (D-A/(x/EC50)^b)$ was used, where A is the minimum effect, D is the maximum effect, x is the chemical concentration, b is the slope was used. Confidence intervals were set at 95%. This formula was used to calculate the EC_{25} and EC_{50}. Significant differences in response were calculated using the T-test in GraphPad. P-Values were ≤ 0.05.

RESULTS AND DISCUSSION

The aim of the present study was to explore the suitability of an *in vitro*-based toxicogenomic approach for MoA assignment of chemicals. The MoA under consideration in the present study was uncoupling of OXPHOS. It is known that uncoupling can
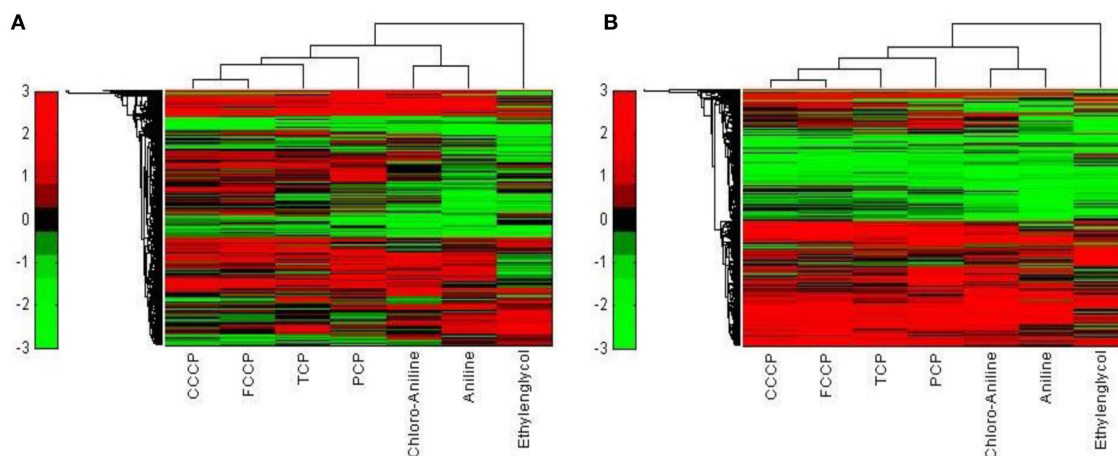
FIGURE 7 | Hierarchical cluster analysis of C3A cells exposed to seven compounds (Ethylenglycol, Cl-aniline, Aniline, PCP, CCCP, FCCP, and TCP). **(A)** The hierarchical cluster analysis is based on the 5000 most varying genes. **(B)** The hierarchical cluster analysis is based on the significantly regulated genes ($|M| > 2$ and $p_{adj} < 0.1$).
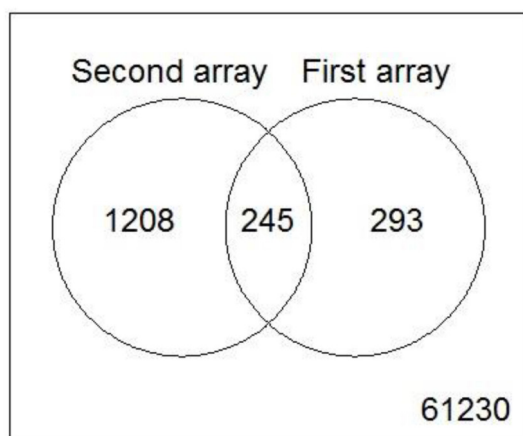
FIGURE 8 | Venn diagram showing the overlap of genes from the core uncoupler set that were regulated in both microarray experiments.

lead to fast transcriptional changes related to the physiological remodeling due to the disruption of ATP generation (Hänninen et al., 2010), and this mechanism may be reflected in the gene signatures of the uncoupler-exposed C3A cells.

Selection of Chemicals with Uncoupling MoA

The selection of chemicals with uncoupling MoA for the experiments of the present study was carried out on the one hand on the basis of biological mechanistic studies and, on the other hand, on the basis of studies applying physico-chemical descriptors. For FCCP and CCCP, several reports describe an interference with proton transfer across the mitochondrial membranes (LeBlanc, 1971; McLaughlin and Dilger, 1980; Terada, 1981; Lim et al., 2001; Stöckl et al., 2007). Since also

studies using structural rules identified FCCP and CCCP as uncouplers (Spycher et al., 2005, 2008a), these two compounds were selected as positive controls. In addition, we compiled a list of chemicals that were suggested to have an uncoupling MoA on the basis of structural alerts and physico-chemical descriptors (Schüürmann et al., 2003; Spycher et al., 2005, 2008a,b) which, however, were not proven to act as uncouplers in biological uncoupling assays. From the candidate compounds listed in **Table 1**, TCP and PCP were selected as compounds with uncoupling MoA mainly based on structural rules. Studies with black lipid bilayer also suggest uncoupling for PCP and TCP (McLaughlin and Dilger, 1980) and experiments with chromophores for TCP (Escher et al., 1996). Finally, chemicals with MoA's other than uncoupling were included as negative controls: BaP as a chemical with a specific MoA (agonist of the arylhydrocarbon receptor), Cl-aniline and aniline as polar narcotics, and ethylenglycol as a narcotic or baseline toxicant (**Table 2**). The structural information about the applied compounds are shown in **Figure 1**.

Biological Assessment of the Uncoupling Activity of the Test Compounds

Two different bioassays were applied to verify if the selected compounds act as biological uncouplers. The hypothesis was that FCCP, CCCP, TCP, and PCP should display protonophoric activity, whereas no such activity should be observed for BaP, the anilines and ethylenglycol. Two bioassays were used for measuring the protonophoric activity: the TMRM assay which is sensitive to alterations of membrane potential (Ehrenberg et al., 1988) was applied to our test system, the mammalian C3A cell line. Additionally, we performed the TPP$^+$ assay which is commonly used to test for uncoupling. The TPP$^+$ assay provides a direct measurement of mitochondrial respiration and membrane potential in isolated rat mitochondria (Lichtshtein et al., 1979). The results are shown in **Table 2**. Both assays showed

TABLE 5 | Correlation coefficient between treatments for microarray experiment 2.

	Ethylenglycol	6OH-BDE90	Cl-aniline	Aniline	CCCP	PCP	FCCP	TCP
A								
Ethylenglycol	1.00	0.40	0.21	0.40	0.25	0.08	0.21	0.27
6OH-BDE90	0.40	1.00	0.57	0.66	0.46	0.60	0.50	0.68
Cl-aniline	0.21	0.57	1.00	0.74	0.54	0.70	0.50	0.43
Aniline	0.40	0.66	0.74	1.00	0.40	0.53	0.39	0.41
CCCP	0.25	0.46	0.54	0.40	1.00	0.62	0.91	0.71
PCP	0.08	0.60	0.70	0.53	0.62	1.00	0.64	0.66
FCCP	0.21	0.50	0.50	0.39	0.91	0.64	1.00	0.75
TCP	0.27	0.68	0.43	0.41	0.71	0.66	0.75	1.00
B								
Ethylenglycol	1.00	0.48	0.38	0.50	0.44	0.26	0.41	0.42
6OH-BDE90	0.48	1.00	0.71	0.72	0.63	0.74	0.67	0.76
Cl-aniline	0.38	0.71	1.00	0.83	0.66	0.77	0.63	0.59
Aniline	0.50	0.72	0.83	1.00	0.57	0.66	0.56	0.54
CCCP	0.44	0.63	0.66	0.57	1.00	0.71	0.94	0.81
PCP	0.26	0.74	0.77	0.66	0.71	1.00	0.72	0.75
FCCP	0.41	0.67	0.63	0.56	0.94	0.72	1.00	0.83
TCP	0.42	0.76	0.59	0.54	0.81	0.75	0.83	1.00

(A) For the 5000 genes with the highest variation. *(B)* For the most significantly regulated genes. Correlation coefficients above 0.7 are marked with gray background.

TABLE 6 | Number of genes that are significant gene regulation ($|M| > 2$ and $p_{adj} < 0.1$) in microarray experiment 2.

Ethylenglycol	6OH-BDE90	Cl-aniline	Aniline	CCCP	PCP	FCCP	TCP
1,861	1,725	412	2,189	688	1,691	514	206

that FCCP and TCP have uncoupling activity. CCCP was clearly positive in the TMRM assay, while in the TPP^+ assay, first a decrease in oxygen consumption was observed then an increase indicating inhibition of OXPHOS together with uncoupling. The dose-response curves for the TPP^+ assay are shown in **Figure 2** and the curves for the TMRM assay in **Figure 3**.

In contrast, for PCP, the TPP^+ assay indicated an uncoupling activity at a very low concentration of $0.5\,\mu M$, while the TMRM only showed effects at $118\,\mu M$ with a higher variability between replicates. On the basis of structural descriptors, PCP is classified as uncoupler (Terada, 1981; McKim et al., 1987; Nendza and Müller, 2001; Schüürmann et al., 2003; Spycher, 2005; Spycher et al., 2005, 2008a). The differences of effect concentrations between our assays agrees well with what has been observed by Schüürmann et al. (1997) The authors identified uncoupling by an excess toxicity above the baseline level using 10 different biological test systems. None of the test systems was specific for uncoupling. In this study PCP was not identified as uncoupler. The authors hypothesized that the uncoupling activity of PCP exists but does not yield substantial contributions to its total acute toxicity as quantified by the biological test systems, shown in a very high baseline toxicity. The high lipophilicity of PCP (log Kow of 5.04) leads to relatively high doses and correspondingly high narcotic-type membrane perturbations in the organisms, so that the additional uncoupling activity may not lead to a

significant increase in overall toxicity. This means that although chemical descriptors classify PCP as uncoupler, the biological space only partly reflects this activity.

Surprisingly, the two anilines showed an ambiguous response in the TMRM bioassay at very high concentrations (mM). Instead of a decrease of the TMRM an increase of the signal could be observed. This might be due to an interference of the TMRM-dye and the compounds. The TPP^+ is more specific for uncoupling activity than the TMRM assay which just measures membrane integrity. Anilines are known to disrupt membranes which might also explain the response in the TMRM assay. Only the polar narcotic compound, ethylenglycol, was clearly negative in both assays. The concentration range tested for the anilines and ethylenglycol are presented in **Table 3**.

MoA Specific Gene Expression Profile at Effect Based Concentrations

A first and essential pre-requiste to enable comparison of toxigenomic profiles across compound sis the use of equipotent concentrations of the test chemicals. These concentrations were selected in the preliminary experiments and used for the following experiments. To test whether C3A cells display a toxicogenomic profile that is common for exposure to uncouplers. a proof of principle experiment was performed with the positive control compounds—FCCP, CCCP, and TCP and BaP as a non-uncoupling negative control. To ensure the comparability of the gene expression profiles, the test compounds were applied at equipotent concentrations (Shioda et al., 2006). The chemicals FCCP, CCCP, and TCP were tested at their EC_{50} values in the TMRM assay, i.e., the concentration which inhibited TMRM uptake by 50%. For BaP, the concentration of $10\,\mu M$ was used which is high enough to induce a significant biological

FIGURE 9 | Hierarchical cluster analysis of C3A cells exposed to eight compounds including 6OH-BDE90 (Ethylenglycol, Cl-aniline, Aniline, PCP, CCCP, FCCP, and TCP). **(A)** The hierarchical cluster analysis is based on the significantly regulated genes ($|M| > 2$ and $p_{adj} < 0.1$). **(B)** The hierarchical cluster analysis is based on the 5000 most varying genes.
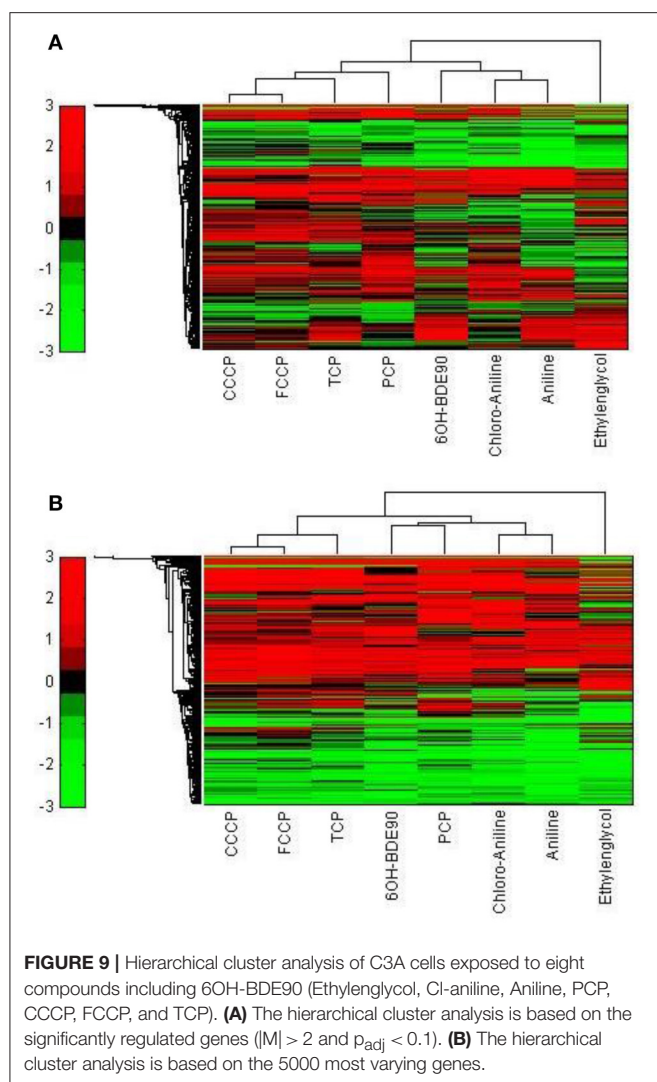
response of the BaP-specific MoA, i.e., cytochrome P4501A induction via the AhR, while it is low enough to avoid cytotoxicity (Hockley et al., 2009). The Cyp1a1 expression was the highest for the BaP exposed cells with an M value of +3. The M value equals the $\log_2$ fold change value. The gene expression profiles induced by the four test compounds were evaluated by means of hierarchical clustering. To this end, the 5000 genes with the highest signal variation over the whole data set were selected (**Figure 4A**). Each of the test chemicals displayed an individual gene expression profile, CCCP, FCCP, and TCP clustered close to each other, while the BaP profile was clearly separated. In an alternative analysis of the array data, we selected those genes that were regulated with a fold change of >2 or <0.5 ($|M| > 1$). Also with this analysis, a high degree of similarity of gene expression was evident for the uncouplers (FCCP, CCCP, and TCP) and a clear separation of the non-uncoupler, BaP (**Figure 4B**). Again, it can be seen that CCCP and FCCP are more closely related in comparison to TCP but all uncouplers are more similar to each other than to BaP. To further demonstrate the similarities between the uncouplers the correlation coefficient

between treatments is calculated. TCP, FCCP, and CCCP all have a correlation coefficient above 0.7, whereas BaP only shows a negative correlation with the other compounds below 0.25, for the 500 most varying genes and the most regulated genes (**Table 4**). The result of a PCA analysis are shown in **Figure 5**. Also in the PCA a clear separation from the uncoupler and BaP is observable. An overlap of expressed genes showed that only five genes are regulated by all treatments. Whereas, 125 genes are regulated by all of the uncouplers (**Figure 6**, Table S1). In summary, the gene expression profiles induced by these three compounds clustered closely together, and separated clearly from the "non-uncoupling negative control," BaP. Importantly, the cluster analysis was not done using genes selected on the basis of mechanistic knowledge, but it simply relied on those genes which showed the strongest variation or regulation. It is not only that similar genes are regulated by these substances, but a majority of the genes is regulated in the same direction indicating the existence of a MoA specific gene expression profile.

MoA Specific Gene Expression Profile at Cytotoxic Based Concentrations

While the former experiment used two strongly contrasting MoAs—uncoupling vs. AhR activation—in a next experiment we asked if the toxicogenomic profiles of compounds with more similar MoA can be distinguished. To this end, two aniline compounds and PCP were included which are suggested to have a polar narcotic MoA, as well as ethylenglycol, which has a narcotic MoA. Structure-based MoA assignments generate equivocal results concerning the separation between polar narcosis and uncoupling (Schüürmann et al., 1996), with some studies classifying the anilines and PCP as uncouplers, while other studies classifying them as polar narcotics (Russom et al., 1997; Schüürmann et al., 1997; Argese et al., 2001; Spycher et al., 2008a; Janssens et al., 2011; Dom et al., 2012). Since these test chemicals are not unequivocally classified as uncouplers, the TMRM assay could not be used for the selection of equipotent test concentrations. Therefore, a cytotoxicity test was applied to derive EC$_{25}$ equipotent concentrations (**Figure 3**). The initial analysis of the gene expression data was conducted using the 5000 most varying genes. The clustering results are shown in **Figure 7A**. Again, each of the compounds tested induced an individual gene expression profile but additionally similarities between the profiles existed. The most similar expression profiles were shown by CCCP, TCP, and FCCP. Thus, for the uncoupler set, the gene expression profiles confirmed the previous results and indicate that these three compounds cluster together. When comparing regulated genes of the uncoupler set between the first and second microarray experiment, a clear overlap could be seen (**Figure 8**). As expected, the higher concentrations of the second microarray experiment (EC$_{25}$ cytotoxicity) altered more genes than the EC$_{50}$ from the TMRM assay. TCP was applied in both experiments at similar concentrations therefore there was also a high overlap of altered genes. The heat map shows a clearly separated cluster of aniline and Cl-aniline vs. the core uncoupler cluster (CCCP, FCCP, and TCP). The third cluster exhibited by cluster analysis was the narcotic compound, ethylenglycol.

When focusing on the most significantly regulated genes, the core uncoupler set clustered together (**Figure 7B**). Again, PCP clustered together with the uncoupler set but is most distinct. The second cluster can be observed for the narcotics (Aniline and Cl-aniline) tested. The polar narcotic ethylenglycol is clearly separated from the other cluster. The results of the hierarchical clustering are also reflected in the correlation coefficient analysis (**Tables 5A,B**). The core uncoupler all correlate with a coefficient of above 0.7. The same holds for the anilines. Ethylenglycol does not correlate with any other treatment. The PCA analysis also shows that ethylenglycol is separated from the other treatments (**Figure 5B**). The number of significantly differently expressed genes for all treatments is shown in **Table 6**. Aniline, Ethylenglycol and PCP changed the most genes whereas TCP and Cl-aniline altered the least number of genes. This shows that the grouping of compounds is not based on number of regulated genes.

Toxiogenomic-Based MoA Assignment of a Chemical with Unknown MoA

In order to test the discrimination power of the toxicogenomic approach, we tested the brominated flame retardant metabolite 6OH-BDE90 for which structural criteria do not allow a MoA assignment, but an uncoupling MoA is suspected on the basis of bioassay data. 6OH-BDE90 clustered together with the narcotics, when the 5000 most varying genes were used (**Figure 9A**). When analyzed for significant gene regulation ($|M| > 2$ and $p_{adj} < 0.1$), 6OH-BDE90 clusters together with PCP forming a new cluster group next to the polar narcotics (**Figure 9B**).

Other studies confirm the equivocal behavior of PCP and 6OH-BDE90. Tests conducted by Legradi et al. (2014) including PCP, 6OH-BDE90, and FCCP show in regard to the TPP$^+$ assay which does not give a positive response for narcotic substances, PCP and FCCP exhibit a similar LOEC whereas 6OH-BDE90 has a much higher LOEC, presenting a similar potency of PCP and FCCP and a much lower potency for 6OH-BDE90. This finding is in line with our classification analysis based on the 5000 most varied genes (**Figure 9A**).

However, when comparing the effects of the three substances in the a TMRM assay, which also gives a positive response for uncoupling and narcotic substances, 6OH-BDE90 and PCP both alter the membrane potential in a similar concentration range whereas FCCP proves to be highly more potent (Legradi et al., 2014). In line with our classification analysis based on the significantly regulated genes, where 6OH-BDE90 and PCP from a new sub cluster together next to the polar narcotics (**Figure 9B**).

It appears that 6OH-BDE90 and PCP have the potential for both uncoupling and polar narcosis, depending on test conditions (concentration, duration) or test species (Sixt et al., 1995; Schüürmann et al., 1997; Legradi et al., 2014). The findings presented in this study suggest that the toxicogenomic approach is able to elucidate differences in MoA. In the case that compounds act via several highly similar MoA, toxicogenomics can reveal this and assign a compound to more than one MoA. Nevertheless, the array has the advantage to demonstrate in one single assay the diverse activity potentials of a compound and this information is a valuable complement to structure-based classification schemes.

In chemical risk assessment, MoA information is critical for proper hazard classification as compounds belonging to the same MoA class should show similar toxicity (Verhaar et al., 2000; Cronin and Livingstone, 2004; Nendza and Wenzel, 2006; Vonk et al., 2009). To date, the assignment of a chemical to a MoA is largely done on the basis of structural rules (Bradbury and Lipnick, 1990; Russom et al., 1997; Schüürmann, 1998). This study explored the potential of toxicogenomic profiles in MoA grouping of chemicals. Classification of chemicals and MoA identification was made purely with discriminative gene expression profiles and without any mechanistic profiling like pathway analysis. The results strongly suggest that transcript profiling indeed is able to identify MoA of chemicals and also is able to distinguish between MoA. Furthermore, as chemicals typically display several MoA, depending on the exposure concentration, duration, biological receptor etc., this ambiguity are well revealed by the array data, since high content methodologies such as toxicogenomics are appropriate tools to deal with such multi-dimensional properties. The present study exemplified this for the cases PCP and 6OH-BDE90 where structural rules yield equivocal results. The applicability of our approach for more MoA and larger chemical sets needs to be further investigated. Proper selection of reference compounds is essential and might be limited. Defining similarity and defining groups might also be more complicated when more MoA are included and advanced statistical analysis necessary. Exposure time could also be adapted to get more specific profiles. For certain MoA a treatment duration of 2 h might not be enough to induce a discriminative profile whereas for other MoA a longer exposure might be too long and mostly general stress response effects visible. Nevertheless, our results indicate that toxicgenomics profiling might be a useful approach for MoA assessment of chemicals.

AUTHOR CONTRIBUTIONS

AH-I and JeL performed the microarray experiments. AH-I did the cell experiments and JeL the TTP assay. JeL did the microarray analysis. PC did the analysis of the cell data. JuL and HS were involved in the design and managing of the experiments and the interpretation of the data. All authors contributed in the writing of the manuscript.

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fenvs.2017.00080/full#supplementary-material

REFERENCES

Abu-Amero, K. K., and Bosley, T. M. (2005). Detection of mitochondrial respiratory dysfunction in circulating lymphocytes using resazurin. *Arch. Pathol. Lab. Med.* 129, 1295–1298. doi: 10.1043/1543-2165(2005)129[1295:DOMRDI]2.0.CO;2

Argese, E., Bettiol, C., Agnoli, F., Zambon, A., Mazzola, M., and Ghirardini, A. V. (2001). Assessment of chloroaniline toxicity by the submitochondrial particle assay. *Environ. Toxicol Chem.* 20, 826–832. doi: 10.1002/etc.5620200418

Barron, M. G., Lilavois, C. R., and Martin, T. M. (2015). MOAtox: a comprehensive mode of action and acute aquatic toxicity database for predictive model development. *Aquat. Toxicol.* 161, 102–107. doi: 10.1016/j.aquatox.2015.02.001

Blackburn, K., Bjerke, D., Daston, G., Felter, S., Mahony, C., Naciff, J., et al. (2011). Case studies to test: a framework for using structural, reactivity, metabolic and physicochemical similarity to evaluate the suitability of analogs for SAR-based toxicological assessments. *Regul. Toxicol. Pharmacol.* 60, 120–135. doi: 10.1016/j.yrtph.2011.03.002

Bradbury, S. P., and Lipnick, R. L. (1990). Introduction : structural properties for determining mechanisms of toxic action. *Environ. Health Perspect.* 87, 181–182.

Brand, M. D., and Nicholls, D. G. (2011). Assessing mitochondrial dysfunction in cells. *Biochem. J.* 435, 297–312. doi: 10.1042/BJ20110162

Carmichael, N., Bausen, M., Boobis, A. R., Cohen, S. M., Embry, M., Fruijtier-Pölloth, C., et al. (2011). Using mode of action information to improve regulatory decision-making: an ECETOC/ILSI RF/HESI workshop overview. *Crit. Rev. Toxicol.* 41, 175–186. doi: 10.3109/10408444.2010.541225

Cronin, M. T. D., and Livingstone, D. J. (2004). *Predicting Chemical Toxicity and Fate.* Boca Raton, FL: CRC Press.

Dom, N., Vergauwen, L., Vandenbrouck, T., Jansen, M., Blust, R., and Knapen, D. (2012). Physiological and molecular effect assessment versus physico-chemistry based mode of action schemes: daphnia magna exposed to narcotics and polar narcotics. *Environ. Sci. Technol.* 46, 10–18. doi: 10.1021/es201095r

ECHA (2011). *The Use of Alternatives to Testing on Animals for the REACH Regulation.* ECHA.

Edwards, S. W., and Preston, R. J. (2008). Systems biology and mode of action based risk assessment. *Toxicol. Sci.* 106, 312–318. doi: 10.1093/toxsci/kfn190

Ehrenberg, B., Montana, V., Wei, M. D., Wuskell, J. P., and Loew, L. M. (1988). Membrane potential can be determined in individual cells from the nernstian distribution of cationic dyes. *Biophys. J.* 53, 785–794. doi: 10.1016/S0006-3495(88)83158-8

Enoch, S. J., Hewitt, M., Cronin, M. T., Azam, S., and Madden, J. C. (2008). Classification of chemicals according to mechanism of aquatic toxicity: an evaluation of the implementation of the Verhaar scheme in toxtree. *Chemosphere* 73, 243–248. doi: 10.1016/j.chemosphere.2008.06.052

Escher, B. I., Bramaz, N., Eggen, R. I. L., and Richter, M. (2005). *In vitro* assessment of modes of toxic action of pharmaceutical in aquatic life. *Environ. Sci. Technol.* 39, 3090–3100. doi: 10.1021/es048590e

Escher, B. I., and Hermens, J. L., (2002). Modes of action in ecotoxicology: their role in body burdens, species sensitivity, QSARs, and mixture effects. *Environ. Sci. Technol.* 36, 4201–4217. doi: 10.1021/es015848h

Escher, B. I., Snozzi, M., and Schwarzenbach, R. P. (1996). Uptake, speciation, and uncoupling activity of substituted phenols in energy transducing membranes. *Environ. Sci. Technol.* 30, 3071–3079. doi: 10.1021/es960153f

Hamadeh, H. K., Bushel, P. R., Jayadev, S., Martin, K., DiSorbo, O., Sieber, S., et al. (2002). Gene expression analysis reveals chemical-specific profiles. *Toxicol. Sci.* 67, 219–231. doi: 10.1093/toxsci/67.2.219

Hänninen, S. L., Ronkainen, J. J., Leskinen, H., and Tavi, P. (2010). Mitochondrial uncoupling downregulates calsequestrin expression and reduces SR Ca^{2+} stores in cardiomyocytes. *Cardiovasc. Res.* 88, 75–82. doi: 10.1093/cvr/cvq180

Hockley, S. L., Mathijs, K., Staal, Y. C., Brewer, D., Giddings, I., van Delft, J. H., et al. (2009). Interlaboratory and interplatform comparison of microarray gene expression analysis of HepG2 cells exposed to benzo(a)pyrene. *OMICS* 13, 115–125. doi: 10.1089/omi.2008.0060

Janssens, T. K., Giesen, D., Mariën, J., van Straalen, N. M., van Gestel, C. A., and Roelofs, D. (2011). Narcotic mechanisms of acute toxicity of chlorinated anilines in folsomia candida (Collembola) revealed by gene expression analysis. *Environ. Int.* 37, 929–939. doi: 10.1016/j.envint.2011.03.012

Kamo, N., Muratsugu, M., Hongoh, R., and Kobatake, Y. (1979). Membrane potential of mitochondria measured with an electrode sensitive to tetraphenyl phosphonium and relationship between proton electrochemical potential and phosphorylation potential in steady state. *J. Membr. Biol.* 49, 105–121. doi: 10.1007/BF01868720

Kavlock, R., Chandler, K., Houck, K., Hunter, S., Judson, R., Kleinstreuer, N., et al. (2012). Update on EPA's toxcast program: providing high throughput decision support tools for chemical risk management. *Chem. Res. Toxicol.* 25, 1287–1302. doi: 10.1021/tx3000939

Kienzler, A., Barron, M. G., Belanger, S. E., Beasley, A., and Embry, M. R. (2017). Mode of action (MOA) assignment classifications for ecotoxicology: an evaluation of approaches. *Environ. Sci. Technol.* 51, 10203–10211. doi: 10.1021/acs.est.7b02337

Kiyosawa, N., Manabe, S., Yamoto, T., and Sanbuissho, A. (2010). Practical application of toxicogenomics for profiling toxicant-induced biological perturbations. *Int. J. Mol. Sci.* 11, 3397–3412. doi: 10.3390/ijms11093397

LeBlanc, O. H. Jr. (1971). The effect of uncouplers of oxidative phosphorylation on lipid bilayer membranes: carbonylcyanide M-chlorophenylhydrazone. *J. Membr. Biol.* 4, 227–251. doi: 10.1007/BF02431973

Legradi, J., Dahlberg, A. K., Cenijn, P., Marsh, G., Asplund, L., Bergman, Å., et al. (2014). Disruption of oxidative phosphorylation (OXPHOS) by hydroxylated polybrominated diphenyl ethers (OH-PBDEs) present in the marine environment. *Environ. Sci. Technol.* 48, 14703–14711. doi: 10.1021/es5039744

Lichtshtein, D., Kaback, H. R., and Blume, A. J. (1979). Use of a lipophilic cation for determination of membrane potential in neuroblastoma-glioma hybrid cell suspensions. *Proc. Natl. Acad. Sci. U.S.A.* 76, 650–654. doi: 10.1073/pnas.76.2.650

Lim, M. L., Minamikawa, T., and Nagley, P. (2001). The protonophore CCCP induces mitochondrial permeability transition without cytochrome c release in human osteosarcoma cells. *FEBS Lett.* 503, 69–74. doi: 10.1016/S0014-5793(01)02693-X

Mantus, E. (2007). *Alttox.Org Toxicity Testing in the 21st Century.* Available online at: http://alttox.org/toxicity-testing-in-the-21st-century

Marsh, G., Hu, J., Jakobsson, E., Rahm, S., and Bergman, ?. (1999). Synthesis and characterization of 32 polybrominated diphenyl ethers. *Environ. Sci. Technol.* 33, 3033–3037. doi: 10.1021/es9902266

Marsh, G., Stenutz, R., and Bergman, Å. (2003). Synthesis of hydroxylated and methoxylated polybrominated diphenyl ethers – natural products and potential polybrominated diphenyl ether metabolites. *Eur. J. Org. Chem.* 2003, 2566–2576. doi: 10.1002/ejoc.200300081

Martin, Y. C., Kofron, J. L., and Traphagen, L. M. (2002). Do structurally similar molecules have similar biological activity ? Do structurally similar molecules have similar biological activity ? *J. Med. Chem.* 45, 4350–4358. doi: 10.1021/jm020155c

McKim, J. M., Bradbury, S. P., and Niemi, G. J. (1987). Fish acute toxicity syndromes and their use in the QSAR approach to hazard assessment. *Environ. Health Perspect.* 71, 171–186. doi: 10.1289/ehp.8771171

McLaughlin, S. G., and Dilger, J. P. (1980). Transport of protons across membranes by weak acids. *Physiol. Rev.* 60, 825–863.

Meganathan, K., Jagtap, S., Wagh, V., Winkler, J., Gaspar, J. A., Hildebrand, D., et al. (2012). Identification of thalidomide-specific transcriptomics and proteomics signatures during differentiation of human embryonic stem cells. *PLoS ONE* 7:e44228. doi: 10.1371/journal.pone.0044228

Nendza, M., and Müller, M. (2001). Discriminating toxicant classes by mode of action: 2. physico-chemical descriptors. *Quant. Struct. Activity Relationships* 19, 581–598. doi: 10.1002/1521-3838(200012)19:6<581::AID-QSAR581>3.0.CO;2-A

Nendza, M., and Wenzel, A. (2006). Discriminating toxicant classes by mode of action. 1. (Eco)toxicity profiles. *Environ. Sci. Pollut. Res. Int.* 13, 192–203. doi: 10.1065/espr2006.01.013

NRC (2007). *Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment.* Available online at: http://www.ncbi.nlm.nih.gov/books/NBK10219/

Pereira-Fernandes, A., Vanparys, C., Vergauwen, L., Knapen, D., Jorens, P. G., and Blust, R. (2014). Toxicogenomics in the 3T3-L1 cell line, a new approach for screening of obesogenic compounds. *Toxicol. Sci.* 140, 352–363. doi: 10.1093/toxsci/kfu092

Piña, B., and Barata, C. (2011). A genomic and ecotoxicological perspective of DNA array studies in aquatic environmental risk assessment. *Aquat. Toxicol.* 105, 40–49. doi: 10.1016/j.aquatox.2011.06.006

Richard, A. M. (2006). Future of toxicology-predictive toxicology: an expanded view of 'chemical toxicity'. *J. Med. Chem.* 19, 1257–1262. doi: 10.1021/tx060116u

Russom, C. L., Bradbury, S., Broderius, S. J., Hammermeister, D., Drummond, R. A., and Veith, G. D. (1997). Predicting modes of toxic action from chemical structure: acute toxicity in the fathead minnow (*Pimephales promelas*). *Environ. Toxicol. Chem.* 16, 948–967. doi: 10.1002/etc.5620160514

Schüürmann, G. (1998). "Ecotoxic modes of action of chemical substances," in *Ecotoxicology: Ecological Fundamentals, Chemical Exposure, and Biological Effects*, eds G. Schüürmann and B. Markert (John Wiley/Spektrum Akademischer Verlag), 665–749.

Schüürmann, G., Aptula, A. O., Kühne, R., and Ebert, R. U. (2003). Stepwise discrimination between four modes of toxic action of phenols in the *Tetrahymena Pyriformis* assay. *Chem. Res. Toxicol.* 16, 974–987. doi: 10.1021/tx0340504

Schüürmann, G., Segner, H., and Jung, K. (1997). Multivariate mode-of-action analysis of acute toxicity of phenols. *Aquat. Toxicol.* 38, 277–296. doi: 10.1016/S0166-445X(96)00842-9

Schüürmann, G., Somashekar, R. K., and Kristen, U. (1996). Structure–activity relationships for chloro- and nitrophenol toxicity in the pollen tube growth test. *Environ. Toxicol. Chem.* 15, 1702–1708. doi: 10.1002/etc.5620151008

Shioda, T., Chesnes, J., Coser, K. R., Zou, L., Hur, J., Dean, K. L., et al. (2006). Importance of dosage standardization for interpreting transcriptomal signature profiles: evidence from studies of xenoestrogens. *Proc. Natl. Acad. Sci. U.S.A.* 103, 12033–12038. doi: 10.1073/pnas.0605341103

Sixt, S., Altschuh, J., and Brueggemann, R. (1995). Quantitative structure-toxicity relationships for 80 chlorinated compounds using quantum chemical descriptors. *Chemosphere* 30, 2397–2414. doi: 10.1016/0045-6535(95)00111-K

Spycher, S., Escher, B. I., and Gasteiger, J. (2005). A quantitative structure–activity relationship model for the intrinsic activity of uncouplers of oxidative phosphorylation. *Chem. Res. Toxicol.* 18, 1858–1867. doi: 10.1021/tx050166j

Spycher, S. L. (2005). *Classification of Organic Compounds into Modes of Toxic Action*. Ph.D. dissertation, Friedrich-Alexander-Universität Erlangen-Nürnberg, Available online at: https://opus4.kobv.de/opus4-fau/frontdoor/index/index/docId/198

Spycher, S., Netzeva, T. I., Worth, A. P., and Escher, B. I. (2008a). Mode of action-based classification and prediction of activity of uncouplers for the screening of chemical inventories. *SAR QSAR Environ. Res.* 19, 433–463. doi: 10.1080/10629360802348803

Spycher, S., Smejtek, P., Netzeva, T. I., and Escher, B. I. (2008b). Toward a class-independent quantitative structure–activity relationship model for uncouplers of oxidative phosphorylation. *Chem. Res. Toxicol.* 21, 911–927. doi: 10.1021/tx700391f

Stöckl, P., Zankl, C., Hütter, E., Unterluggauer, H., Laun, P., Heeren, G., et al. (2007). Partial uncoupling of oxidative phosphorylation induces premature senescence in human fibroblasts and yeast mother cells. *Free Radic. Biol. Med.* 43, 947–958. doi: 10.1016/j.freeradbiomed.2007.06.005

Terada, H. (1981). The interaction of highly active uncouplers with mitochondria. *Biochim. Biophys. Acta Rev. Bioenerget.* 639, 225–242.

Terada, H. (1990). Uncouplers of oxidative phosphorylation. *Environ. Health Perspect.* 87, 213–218.

Theunissen, P. T., Robinson, J. F., Pennings, J. L., van Herwijnen, M. H., Kleinjans, J. C., and Piersma, A. H. (2012). Compound-specific effects of diverse neurodevelopmental toxicants on global gene expression in the neural embryonic stem cell test (ESTn). *Toxicol. Appl. Pharmacol.* 262, 330–340. doi: 10.1016/j.taap.2012.05.011

Verhaar, H. J. M., Van Leeuwen, C. J., and Hermens, J. L. M. (1992). Classifying environmental pollutants 1 : structure-activity relationships for prediction of aquatic toxicity 1. *Chemosphere* 25, 471–491. doi: 10.1016/0045-6535(92)90280-5

Verhaar, H. J., Solbé, J., Speksnijder, J., van Leeuwen, C. J., and Hermens, J. L. (2000). Classifying environmental pollutants: part 3. external validation of the classification system. *Chemosphere* 40, 875–883. doi: 10.1016/S0045-6535(99)00317-3

Vonk, J. A., Benigni, R., Hewitt, M., Nendza, M., Segner, H., van de Meent, D., et al. (2009). The use of mechanisms and modes of toxic action in integrated testing strategies: the report and recommendations of a workshop held as part of the European union OSIRIS integrated project. *Altern. Lab. Anim.* 37, 557–571.

Whitmore, D., Foulkes, N. S., and Sassone-Corsi, P. (2000). Light acts directly on organs and cells in culture to set the vertebrate circadian clock. *Nature* 404, 87–91. doi: 10.1038/35003589

Woods, C. G., Heuvel, J. P., and Rusyn, I. (2007). Genomic profiling in nuclear receptor-mediated toxicity. *Toxicol. Pathol.* 35, 474–494. doi: 10.1080/01926230701311351

Yang, L., Kemadjou, J. R., Zinsmeister, C., Bauer, M., Legradi, J., Müller, F., et al. (2007). Transcriptional profiling reveals barcode-like toxicogenomic responses in the zebrafish embryo. *Genome Biol.* 8:R227. doi: 10.1186/gb-2007-8-10-r227

Crosstalk between Receptor and Non-receptor Mediated Chemical Modes of Action in Rat Livers Converges through a Dysregulated Gene Expression Network at Tumor Suppressor Tp53

Karen M. Funderburk[1,2†], Scott S. Auerbach[3] and Pierre R. Bushel[2*]

[1] Department of Biology and Department of Mathematics & Statistics, College of Arts & Sciences, University of North Carolina at Greensboro, Greensboro, NC, United States, [2] Microarray and Genome Informatics Group, Biostatistics and Computational Biology Branch, National Institute of Environmental Health Sciences, Durham, NC, United States, [3] Toxicoinformatics Group, Biomolecular Screening Branch, National Institute of Environmental Health Sciences, Durham, NC, United States

Chemicals, toxicants, and environmental stressors mediate their biologic effect through specific modes of action (MOAs). These encompass key molecular events that lead to changes in the expression of genes within regulatory pathways. Elucidating shared biologic processes and overlapping gene networks will help to better understand the toxicologic effects on biological systems. In this study we used a weighted network analysis of gene expression data from the livers of male Sprague-Dawley rats exposed to chemicals that elicit their effects through receptor-mediated MOAs (aryl hydrocarbon receptor, orphan nuclear hormone receptor, or peroxisome proliferator-activated receptor-α) or non-receptor-mediated MOAs (cytotoxicity or DNA damage). Four gene networks were highly preserved and statistically significant in each of the two MOA classes. Three of the four networks contain genes that enrich for immunity and defense. However, many canonical pathways related to an immune response were activated from exposure to the non-receptor-mediated MOA chemicals and deactivated from exposure to the receptor-mediated MOA chemicals. The top gene network contains a module with 33 genes including tumor suppressor TP53 as the central hub which was slightly up-regulated in gene expression compared to control. Although, there is crosstalk between the two MOA classes of chemicals at the TP53 gene network, more than half of the genes are dysregulated in opposite directions. For example, Thromboxane A Synthase 1 (*Tbxas1*), a cytochrome P450 protein coding gene regulated by Tp53, is down-regulated by exposure to the receptor-mediated chemicals but up-regulated by the non-receptor-mediated chemicals. The regulation of gene expression by the chemicals

within MOA classes was consistent despite varying alanine transaminase and aspartate aminotransferase liver enzyme measurements. These results suggest that overlap in toxicologic pathways by chemicals with different MOAs provides common mechanisms for discordant regulation of gene expression within molecular networks.

INTRODUCTION

The environment that humans and other species are exposed to is a complex space that contains various biologic stressors (natural and manufactured) which can alter cellular processes and in some cases, result in disease and affliction (Wild, 2005; Rappaport and Smith, 2010). Toxicants elicit their toxicologic effect in the liver through an engagement of target macromolecules which leads to a cascade of events referred to as modes of action (MOAs; Casarett et al., 2001). Genomic signatures manifested from toxicant exposure reflect the gene regulation that orchestrates downstream signaling through a particular MOA (Nijman, 2015). Toxicants that act through different molecular initiating events possess distinct MOAs and therefore exhibit unique genomic signatures.

Several efforts have been undertaken to identify gene expression signatures in response to toxicant exposures and to classify chemicals according to their molecular fingerprints (Amin et al., 2002; Bushel et al., 2002; Hamadeh et al., 2002a,b; Kleinjans, 2014; Wei et al., 2014). There are several known cases of chemicals that exert their effect through a particular MOA and have overlaps in the gene expression regulatory networks that regulate the biological processes. For instance, nuclear receptor-mediated chemicals such as those that act through the aryl hydrocarbon receptor (AhR), the peroxisome proliferator-activated receptor (PPAR), or the constitutive androstane receptor and pregnane X receptor (CAR/PXR) have a high degree of agreement between the molecular pathways that are perturbed (Woods et al., 2007). However, little is known about the overlapping regulatory pathways between toxicants that exert their effect through different MOAs.

Reconstruction of gene regulatory networks from gene expression data has assisted in resolving connections between genes during static conditions or dynamically as conditions change over time, dose concentration and/or target tissue (Karlebach and Shamir, 2008). Comparing gene regulatory networks to identify overlaps in connected regions is a challenge. The weighted gene correlation network analysis (WGCNA) method is designed to resolve preserved co-expression gene network modules between two conditions (Zhang and Horvath, 2005; Yip and Horvath, 2007). The approach uses transformations of the correlation between co-expressed genes to reveal interconnectedness amongst gene network nodes and permutation procedures to identify statistically significant gene network modules that overlap between sample conditions (Langfelder et al., 2011). Recent utilization of WGCNA on rat liver gene expression data from drug toxicity studies revealed 415 gene network models that associate with mechanisms of liver pathogenesis (Sutherland et al., 2017). We used the WGCNA approach to reconstruct gene networks using microarray gene expression data from male rat livers and identify preserved modules between chemicals that exert their MOA through receptors (RM) vs. those that are non-receptor-mediated (NRM). We found that the most significant gene network contains 33 genes including tumor suppressor TP53 as the central hub and that the majority of the genes were regulated in opposite directions between the RM and NRM samples. Although there is crosstalk between the two MOA classes of chemicals at the Tp53 signaling pathway, more than half of the genes are dysregulated in opposite directions. The read across between gene networks of chemicals with different MOAs suggests flexibility in the regulatory components of molecular systems to utilize common gene networks to maximize diversity in biological responses.

MATERIAL AND METHODS

Chemicals and Modes of Action

Fifteen chemicals, each with a given dose and duration of exposure, were used for this study (**Table 1**). Sets of three chemicals share one of five MOAs. Three MOAs are associated with well-defined RM processes: peroxisome proliferator-activated receptor-α (PPARA), orphan nuclear hormone receptors (CAR/PXR), and aryl hydrocarbon receptor (AhR). The other two are NRM: DNA damage (DNA_damage) and cytotoxicity (Cytotox). The chemicals were administered orally or by intraperitoneal, intravenous or subcutaneous injection (5 ml/kg body weight). In order to ensure a maximal transcriptional response, 5-day maximum tolerated doses (MTD) of the chemicals were administered to the study animals. The MTD was determined in a 5-day dose range-finding study in which an MTD was determined as a 5–10% reduction in body weight relative to control.

Microarray Gene Expression Data

Total RNA extracted from the livers of male Sprague-Dawley rats exposed once daily for 3, 5, or 7 days in triplicate, depending on the chemical or vehicle control (saline, corn oil or carboxymethyl cellulose), were processed for microarray analysis as previously described (Wang et al., 2014). Animals were handled in accordance with The United States Department of Agriculture and Code of Federal Regulations Animal Welfare Act (9 CFR Parts 1, 2, and 3). Ethics committee approval was not required according to the local and national guidelines. Fragmented cRNA prepared from liver RNA was labeled and hybridized to the Affymetrix whole genome GeneChip® Rat Genome 230 2.0 Array comprised of 31,099 gene probe sets. Pixel intensity data was acquired by scanning of the arrays using the GeneChip® Scanner 3000 7G. CEL files were

TABLE 1 | Chemical exposures and modes of action.

MOA	Chemical	Dose (mg/kg body weight)	Duration (days)
Aryl hydrocarbon receptor (AhR)	3-Methylcholanthrene (3ME)	300	5
	Leflunomide (LEF)	60	5
	beta-Naphthoflavone (NAP)	1,500	5
Orphan nuclear hormone receptors (CAR/PXR)	Phenobarbital (PHE)	54	5
	Methimazole (MET)	100	3
	Econazole (ECO)	334	5
Cytotoxicity (Cytotox)	Chloroform (CHO)	600	5
	Thioacetamide (THI)	200	5
	Carbon tetrachloride (CAR)	1,175	7
DNA Damage (DNA_Damage)	Aflatoxin B1 (AFL)	0.3	5
	Ifosfamide (IFO)	143	3
	N-Nitrosodimethylamine (NIT)	10	5
Peroxisome proliferator-activated receptor alpha (PPARA)	Pirinixic acid (PIR)	364	5
	Bezafibrate (BEZ)	617	7
	Nafenopin (NAF)	338	5

generated using the GCOS software. The pixel intensity data was preprocessed using the robust multichip average (RMA) algorithm (Irizarry et al., 2003a,b) which includes background correction, quantile normalization, and summarization by the median polish approach and then log base 2 transformed. Due to a batch effect in the study design, the data was preprocessed further by mean centering on the route of administration of the chemicals. Next, we performed principal component analysis (PCA)-based gene probe filtering on the preprocessed data using the Bioconductor package "pvac," where the filtering is based on a score measuring consistency among probes within a probe set (Lu et al., 2011). The maximum value of the threshold for the score is set at 0.5, which corresponds to 50% of the total variation accounted for by the 1st principal component. Finally, the preprocessed data was converted to log base 2 ratios by subtracting the average of the controls from the treated samples matched according to nutritional status of the vehicle and route of administration (i.e., non-nutritional-intraperitoneal, intravenous or subcutaneous injection; nutritional-oral; non-nutritional-oral). The raw data is available in the Gene Expression Omnibus (GEO) database (Edgar et al., 2002; Barrett et al., 2013) under the accession number GSE47792.

Clinical Chemistry

Clinical chemistry evaluations of blood serum samples were performed using a Roche Cobas Fara chemistry analyzer (Roche Diagnostic Systems, Westwood, NJ, USA) to numerically measure enzyme levels and metabolic entities.

Statistical Modeling

The preprocessed log base 2 ratio microarray gene expression data comprised of 12,288 gene probe sets was analyzed with the following analysis of variance (ANOVA) model to identify gene probes that vary by MOA:

$$Y_{ijklm} = \mu + M_i + V_j + R_k + D(V^*R)_{jkl} + \varepsilon_{ijklm} \qquad (1)$$

where Y_{ijklm} represents the mth observation on the ith MOA (M), j^{th} vehicle (V), k^{th} route (R) and l^{th} study date (D). μ is the common effect for the whole study and ε_{ijklm} represents the random error. The errors ε_{ijklm} are assumed to be normally and independently distributed with mean 0 and standard deviation δ for all measurements. Significant gene probes that vary according to the MOAs were detected at a Benjamini–Hochberg false discovery rate (FDR) < 0.01.

Weighted Gene Correlation Network Analysis

The log base 2 ratio data of the 2,930 gene probe sets (2,405 genes) that vary significantly according to MOAs were averaged by replicate chemicals then divided into two data sets based on the manner in which the chemicals elicit their toxic effect: RM (AhR, CAR/PXR, and PPARA) and NRM (Cytotox and DNA_damage). A gene network was reconstructed for each data set using the WGCNA method (Zhang and Horvath, 2005; Yip and Horvath, 2007; Langfelder and Horvath, 2008). Briefly, a similarity matrix **S** is generated for each data set to determine how similar in expression genes are. Here, **S** is comprised of the Pearson correlation of the ith and jth gene probe sets (s_{ij}) within a data set. Then, **S** is transformed to an adjacency matrix **A** to ascertain groups of co-expressed genes. Here we used the following soft power adjacency function to generate **A**:

$$a_{ij} = \left| s_{ij} \right|^{\beta} \qquad (2)$$

where $\beta \geq 1$ is a user defined power parameter to control the thresholding of the grouping of the co-expressed genes. The higher the value of β, the fewer co-expressed genes are grouped together. We set $\beta = 10$. Finally, a determination is made if two nodes of co-expressed genes overlap. The topological overlap matrix (TOM) Ω measures two nodes interconnectedness and is computed as:

$$w_{ij} = \frac{\left| N_1(i) \cap N_1(j) \right| + a_{ij}}{\min \left\{ \left| N_1(i) \right|, \left| N_1(j) \right| \right\} + 1 - a_{ij}} \qquad (3)$$

where $N_1(i)$ denotes the set of direct neighbors of node i, $|\cdot|$ denotes the number of elements (i.e., the cardinality) and $\left| N_1(i) \cap N_1(j) \right|$ denotes the number of neighbor nodes that i and j have in common. Note that w_{ij} is bounded between 0 and 1: $w_{ij} = 0$ if nodes i and j are not connected and the two nodes do not share any neighbors; $w_{ij} = 1$ if there is a direct link between the two nodes and if one set of direct neighbors is a subset of the other. The topological dissimilarity measure is denoted as

$$d_{ij}^{w} = 1 - w_{ij}. \qquad (4)$$

Significance of preserved co-expressed genes network modules between RM and NRM exposures is determined by a permutation based composite Z statistic ($Z_{summary}$) defined as the mean of Z scores computed for density and connectivity measures (Yip and Horvath, 2007; Langfelder et al., 2011).

RESULTS

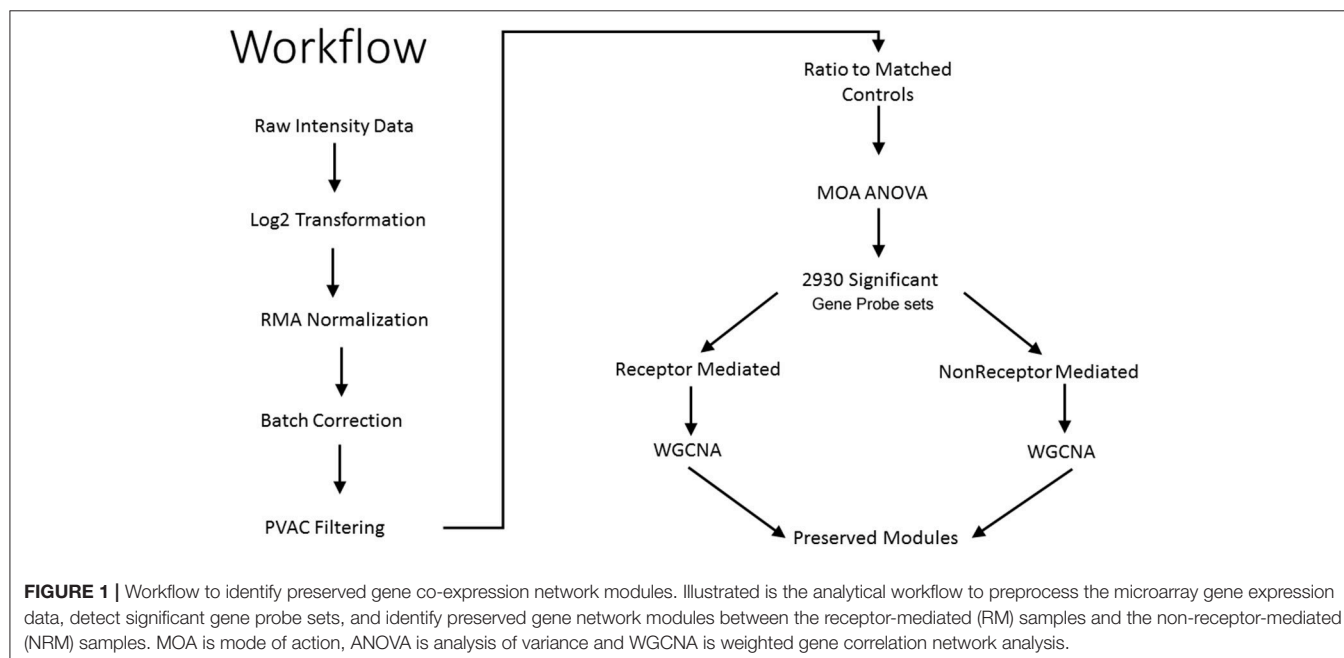## Chemicals Grouped by Mode of Action

To investigate the gene regulatory crosstalk between RM and NRM chemicals, we used the microarray gene expression data recently published from the livers of male Sprague-Dawley rats exposed in triplicate to various chemicals with different MOAs (Wang et al., 2014). The chemicals and their MOAs are listed in **Table 1** along with the doses and durations of exposure. Each

MOA consists of 3 chemicals. The five MOAs are mediated by aryl hydrocarbon receptor (Ahr), orphan nuclear hormone receptors (CAR/PXR), cytotoxicity (Cytotox), DNA damage (DNA_Damage) or peroxisome proliferator-activated receptor-α (PPARA). Clinical chemistry analysis of the samples revealed that alanine transaminase (ALT) and aspartate aminotransferase (AST) liver enzymes levels were substantially higher from the Cytotox and PPARA chemicals than the others indicative of more marked injury to the organ (**Table 2**).

TABLE 2 | Clinical chemistry of samples by mode of action.

Measurement	NN-IP	NN-OG	NU-OG	AhR	CAR/PXR	PPARA	Cytotox	DNA damage
ALT (U/L)	45.50	45.67	69.00	33.44	55.63	93.33	273.83	54.00
	11.94	7.55	13.81	9.79	13.33	41.73	106.02	9.70
AST (U/L)	80.50	76.50	84.80	69.56	72.44	155.67	456.00	98.50
	13.11	9.14	10.91	16.76	25.23	104.18	217.18	20.23
Albumin (g/L)	3.90	3.73	4.00	3.66	4.01	4.76	3.92	4.07
	0.17	0.33	0.11	0.47	0.27	0.24	0.18	0.08
BUN (mg/dL)	13.25	13.67	14.20	30.67	15.38	14.44	34.00	18.33
	1.32	2.80	1.47	37.01	3.94	3.24	31.86	2.16
Cholesterol (mg/dL)	73.00	68.67	70.40	117.33	76.67	68.22	66.83	67.33
	1.10	17.39	7.06	29.47	16.16	24.38	45.65	19.65
Creatine Phosphokinase (U/L)	571.00	192.33	269.20	178.78	345.56	506.67	521.67	333.83
	269.33	37.92	126.44	54.25	376.27	631.03	431.97	192.74
Glucose (mg/dL)	196.75	156.50	165.60	170.00	143.33	133.67	123.00	159.67
	20.56	29.32	5.75	28.39	14.05	10.67	11.72	10.19
Lactate Dehydrogenase (U/L)	175.00	145.67	128.20	137.00	141.22	356.67	270.50	145.50
	69.31	52.66	44.62	62.97	92.20	300.81	163.12	57.36
Total Bilirubin (mg/dL)	0.15	0.36	0.14	0.31	0.16	0.16	0.42	0.15
	0.04	0.21	0.05	0.11	0.05	0.09	0.24	0.04

Shown in the top line of each row is the mean of the measurement for all samples within a MOA. Shown in the bottom line of each row is the standard deviation of the mean.

FIGURE 1 | Workflow to identify preserved gene co-expression network modules. Illustrated is the analytical workflow to preprocess the microarray gene expression data, detect significant gene probe sets, and identify preserved gene network modules between the receptor-mediated (RM) samples and the non-receptor-mediated (NRM) samples. MOA is mode of action, ANOVA is analysis of variance and WGCNA is weighted gene correlation network analysis.

As shown in **Figure 1**, we use a bioinformatics workflow to process the gene expression data for statistical analysis. Following the preprocessing and filtering of the data, we modeled it with a MOA-ANOVA to identify 2,930 gene probe sets that vary statistically according to one or more of the MOAs. Using these dysregulated genes to project the samples into two-dimensional space by the amount of variability captured in principal components 1 and 2 (PC#1 and PC#2), we see that although the majority of the chemicals within each MOA grouped close to each

other, four chemicals (NIT, THI, ECO, and LEF) are separated from the other chemicals that are in their respective MOA (**Figure 2A**). The NIT samples are separated far from all other samples possibly because they were the only ones that exhibited a high level (moderate severity) of centrilobular necrosis of the liver from the exposure (Data not shown). The THI treated liver samples exhibited minimal centrilobular necrosis in all three replicates (Data not shown). This departure from the cohesiveness of the grouping of the chemicals within their MOA
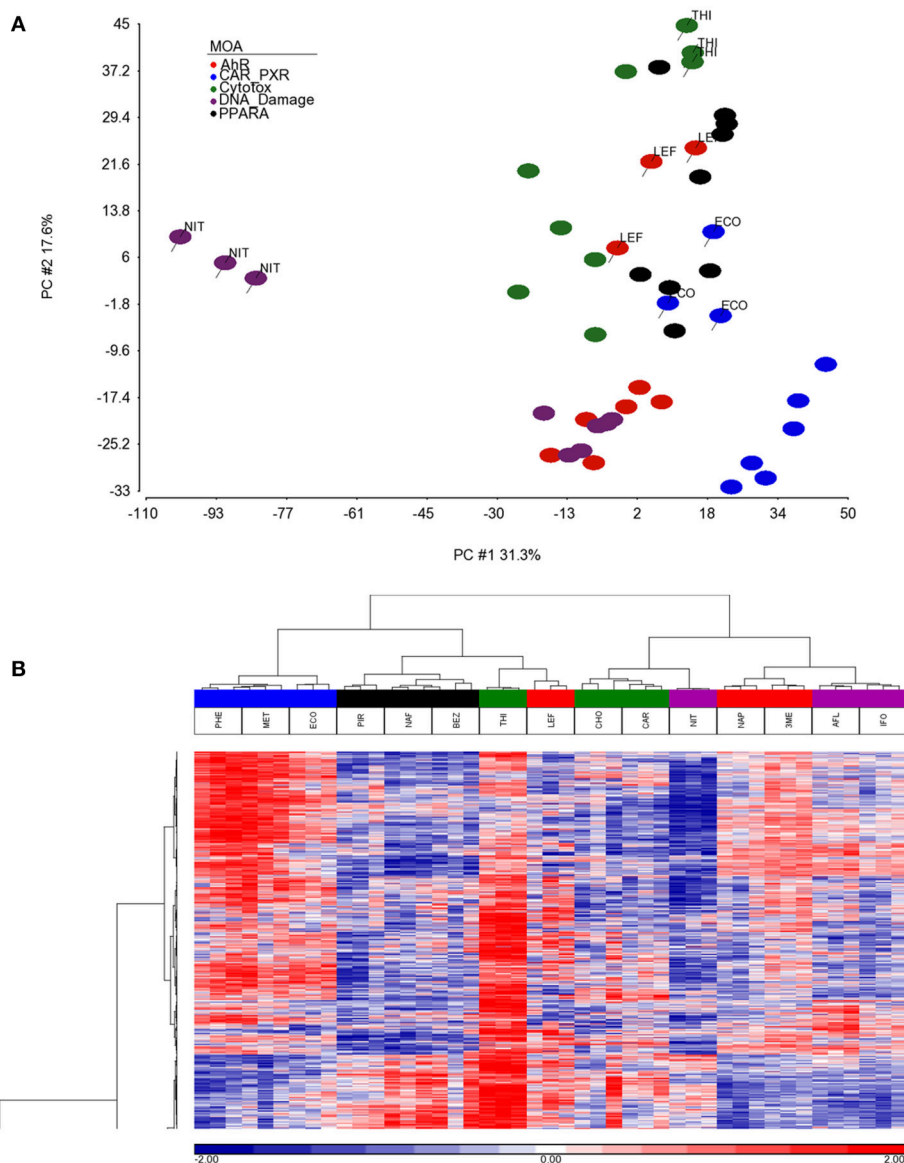
FIGURE 2 | Separation and clustering of samples exposed to the chemicals in triplicate. **(A)** Principal component analysis (PCA) separation of the MOA samples using the 2,930 significant gene probe sets that vary by MOA. The x-axis is PC#1 (31.3% variation captured), the y-axis is PC#2 (17.6% variation captured) and the colors represent the MOAs as shown in the figure legend. **(B)** Two-dimensional hierarchical, agglomerative clustering of the MOA samples using the 2,930 significant gene probe sets that vary by MOA. Clustering performed using Spearman rank as the similarity metric and the Ward minimum variance criterion for grouping. The x-axis is the MOA samples colored as described in the legend to **(A)**, the y-axis is the 2,930 significant gene probe sets. The data is the log base 2 ratio (treated sample to the average of the controls matched according to nutritional status of the vehicle and route of administration) and the scale on the bottom displays the color range for the log base 2 ratio values standardized to mean 0 and standard deviation of 1. Red denotes up-regulation, blue down-regulation, and white relatively no change.
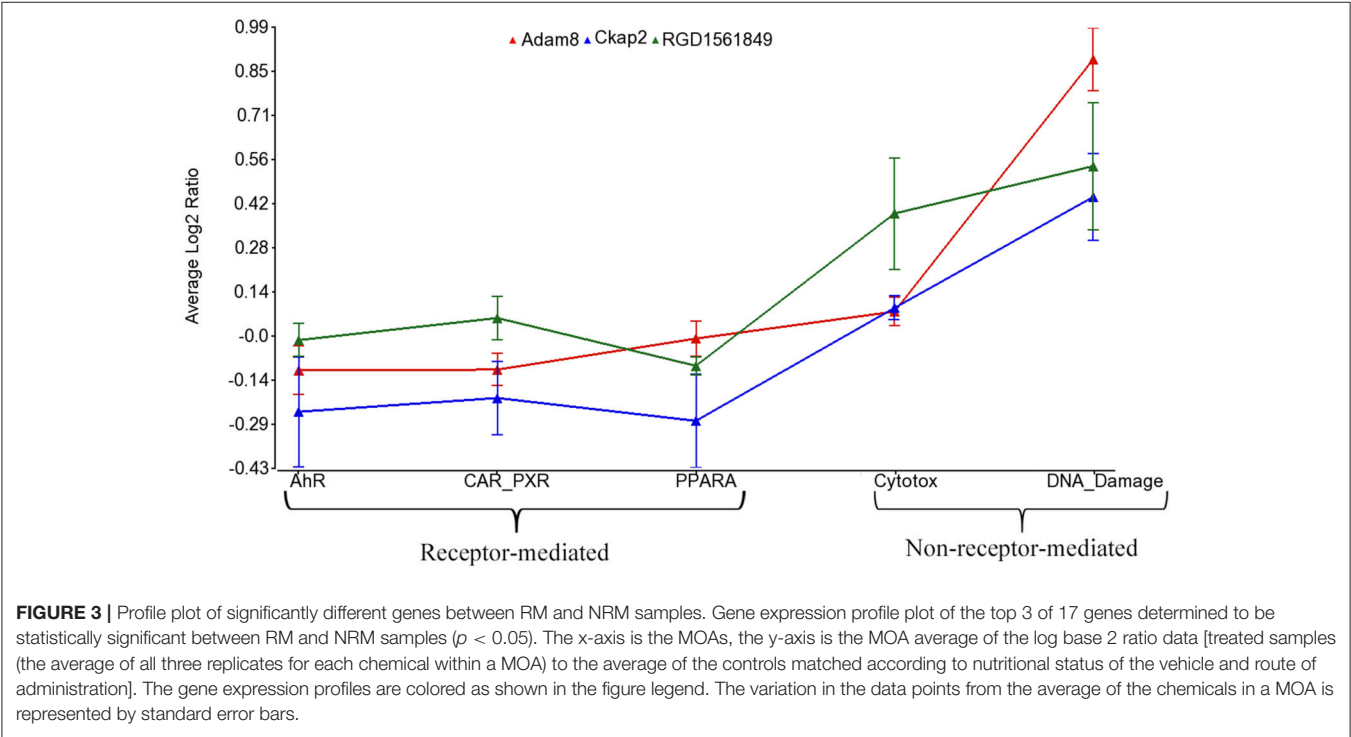
is also observed in the hierarchical clustering of chemicals by MOA into two branches of the dendrogram (**Figure 2B**). RM chemicals in MOAs CAR/PXR and PPARA cluster together and NRM chemicals in MOAs Cytotox and DNA_Damage cluster together. However, the Cytotox chemical THI clusters with the RM chemicals and the NAP and 3ME chemicals clusters with the

NRM chemicals. This suggests that although the chemicals share a MOA, the underlying gene expression changes elicited from the exposures can vary whether mediated by a receptor or not. Of interest is to determine if there are overlaps (i.e., read across) in gene expression between RM chemicals and NRM chemicals in the rat liver.

TABLE 3 | Differentially expressed genes between receptor-mediated and non-receptor-mediated MOAs.

ProbeID	Entrez gene	Gene symbol	Gene description	p-value	Fold change
1392754_at	499285	Adam8	ADAM metallopeptidase domain 8	3.47E-03	−1.47
1384068_at	306575	Ckap2	Cytoskeleton associated protein 2	1.35E-03	−1.42
1390317_at	500393	RGD1561849	Similar to RIKEN cDNA 3110035E14	6.86E-04	−1.40
1370902_at	286921 /// 296972	Akr1b10	Aldo-keto reductase family 1, member B10	1.35E-02	−1.39
1368271_a_at	79451	Fabp4	Fatty acid binding protein 4, adipocyte	2.01E-02	−1.38
1374775_at	291234	Mki67	Marker of proliferation Ki-67	9.28E-03	−1.38
1383747_at	361921	Ect2	Epithelial cell transforming 2	9.89E-03	−1.37
1379582_a_at	114494	Ccna2	Cyclin A2	4.45E-02	−1.35
1384449_at	100910797	LOC100910797	Uncharacterized	4.02E-05	−1.35
1398540_at	54289	Rgs1	Regulator of G-protein signaling 1	9.50E-03	−1.35
1393041_at	362519	Smc2	Structural maintenance of chromosomes 2	4.19E-03	−1.34
1370462_at	25460	Hmmr	Hyaluronan mediated motility receptor	4.16E-03	−1.33
1384231_at	364648	Shcbp1	Shc SH2-domain binding protein 1	5.34E-03	−1.33
1393848_at	362720 /// 100359539	Rrm2	Ribonucleotide reductase M2	1.21E-02	−1.32
1383578_at	499870 /// 100911267	Rad51	DNA repair protein RAD51 recombinase	8.33E-03	−1.32
1390659_at	25406	Cd44	CD44 molecule	2.94E-02	−1.32
1371074_a_at	–	–	–	3.32E-03	−1.30

Gene detected as statistically different in a one-way ANOVA model with the Fisher's least significant difference test between RM vs. NRM samples based on the average (triplicates per chemical) log2 ratio values of the 2,930 MOA-varying gene probe sets.

FIGURE 3 | Profile plot of significantly different genes between RM and NRM samples. Gene expression profile plot of the top 3 of 17 genes determined to be statistically significant between RM and NRM samples ($p < 0.05$). The x-axis is the MOAs, the y-axis is the MOA average of the log base 2 ratio data [treated samples (the average of all three replicates for each chemical within a MOA) to the average of the controls matched according to nutritional status of the vehicle and route of administration]. The gene expression profiles are colored as shown in the figure legend. The variation in the data points from the average of the chemicals in a MOA is represented by standard error bars.

Derivation of Gene Co-expression Networks

Using the 2,930 dysregulated gene probe sets, we grouped the Ahr, CAR/PXR, and PPARA chemicals into a RM class and the Cytotox and DNA_Damage chemicals into a NRM class. **Table 3** lists the genes detected as statistically significant between the two classes. The expression of all the genes from exposure to the RM chemicals are down-regulated in comparison to NRM chemicals. **Figure 3** shows a profile plot of the three genes (*Adam8*, *Ckap2*, and RGD1561849) that are down-regulated most.

FIGURE 4 | Cluster dendrograms and modules representative of gene co-expression networks. **(A)** RM cluster dendrogram. **(B)** NRM cluster dendrogram. The x-axes represent the modules identified relative to the RM clustering. The colors indicate the preserved modules. The significantly preserved modules are labeled by color. The y-axes show the heights where clusters are merged.

We then analyzed the 2,930 dysregulated gene probe sets with the WGCNA method to identify gene networks preserved between the two classes (**Figure 1**). Correlation between gene expression is measured by Pearson correlation and the determination of co-expression is accomplished by using an adjacency function. The interconnectedness of nodes of co-expressed genes in the network is assessed by a topology distance metric. **Figure 4A** depicts the RM co-expressed genes nodes as leaves in the dendrogram with the more similarly expressed genes grouped closer together. The colored modules represent the gene networks that were identified. The turquoise colored module is the largest of the 18 identified. **Figure 4B** illustrates the clustering of the NRM co-expressed genes nodes and the superimposing of the 18 RM modules. As can be seen by the diffuse overlapping of the modules in the two classes, the turquoise, magenta, red and blue colored ones are preserved most readily. This preservation is statistically assessed by a permutation test to derive of a summary Z score. The significant modules (Z_summary score > 10) are shown in **Table 4** with turquoise being the most significant. The gene network sizes are 540, 176, 325, and 131 gene probe sets for the turquoise, red, blue, and magenta modules respectively.

Pathways and Biological Processes Read Across Genes in Preserved Modules

To infer which pathways are over-represented by the genes in the preserved modules, we performed an enrichment test using the Protein ANalysis THrough Evolutionary Relationships (PANTHER) ontology database (Thomas et al., 2003). **Table 5** shows the significant biological processes that were enriched by the genes in each of the preserved modules. Immunity and defense was overwhelmingly significant (FDR < 10%) by the genes in three of the four modules. Using the Ingenuity Pathway Analysis (IPA) knowledgebase, we discovered that TP53 is a central hub of the 33 genes from the turquoise module that have connections (**Figure 5**). Interestingly, although the connections are the same between the RM and the NRM chemicals due to the preservation of the turquoise module, the expression of more

than half of the genes are dysregulated in opposite directions. Some of these genes code for proteins that are associated with cell division (FZR1, CDCA3), metabolism (TBXAS1), and DNA repair (PCLAF).

DISCUSSION

Exposure to chemicals can elicit pharmacologic effects if therapeutic, tailored accordingly and given at the right dose for an appropriate amount of time. In other cases, the exposure can have no detectable effect or can be toxic resulting in an adverse effect to biological systems. The molecular initiating events for many chemicals are well-studied. However, their MOAs remain to be determined. Having a better understanding of a chemical's MOA and the molecular consequences from their exposure will aid in determining points of potential crosstalk between regulatory pathways which may lead to unintended side effects if chemicals act synergistically.
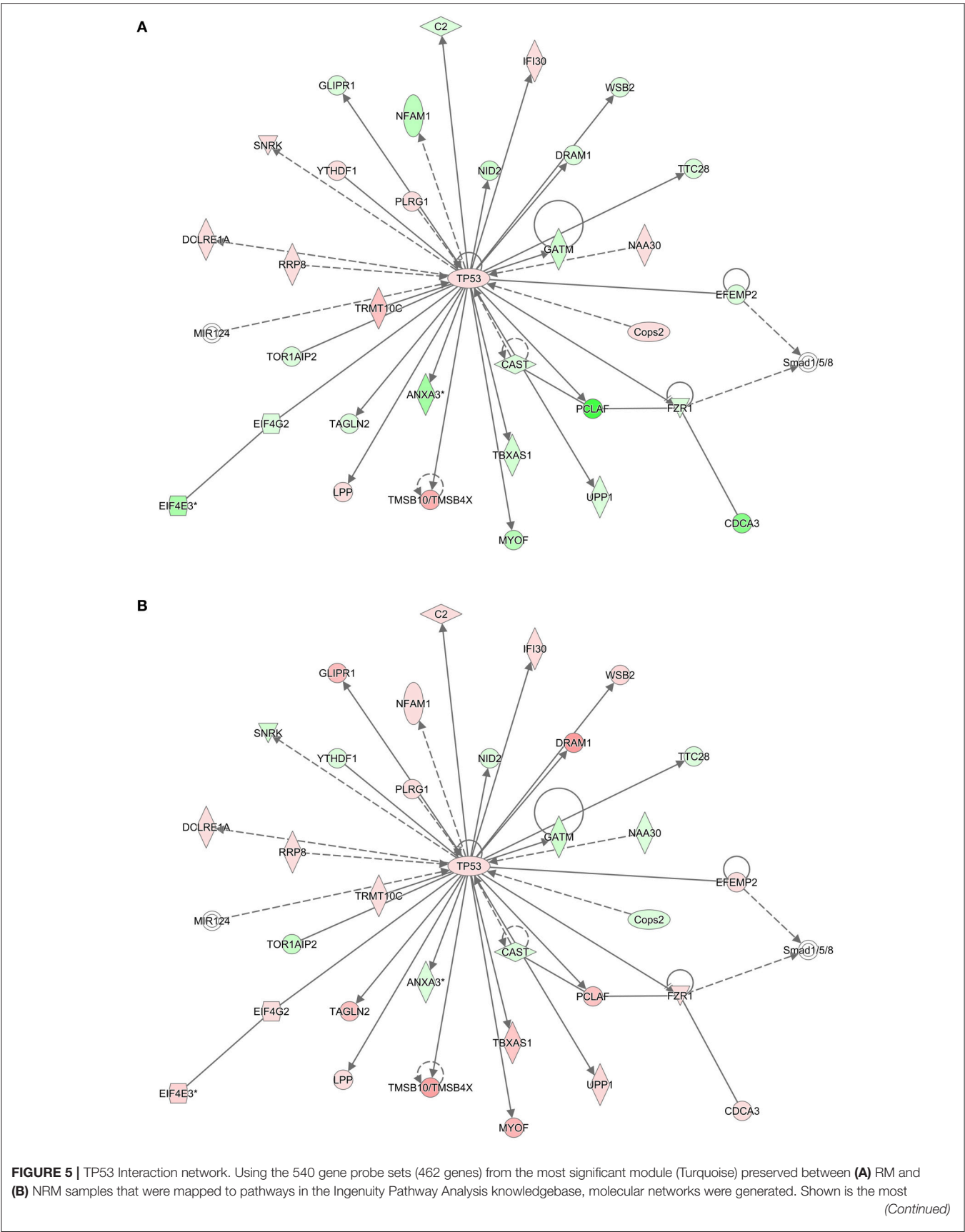
We used gene expression from the livers of male Sprague-Dawley rats exposed to a number of agents (**Table 1**) or vehicle control to identify overlapping gene networks between chemicals that are receptor-mediated (RM) and those that are non-receptor-mediated (NRM). The RM class of chemicals contained those that elicit their effect through either the peroxisome proliferator-activated receptor-α (PPARA), orphan nuclear hormone receptors (CAR/PXR) or aryl hydrocarbon receptor (AhR) while the NRM chemicals do so by cytotoxicity (Cytotox) or DNA damage (DNA_Damage). Each MOA contained 3 different chemicals with each chemical exposure in triplicate. Four gene network modules were preserved in a statistically significant manner between the two classes of chemicals (**Table 4**). The genes in three of the four networks over-represent immunity and defense biological processes (**Table 5**). DNA damage and cytotoxic chemicals are known to trigger an innate immune defense by eliciting parenchymal cell death and subsequent DAMP (Danger-associated molecular patterns) release (Srikrishna and Freeze, 2009). Notably scoring of the modules using the Nextbio Body Atlas (Data not shown) reveals that genes in all four modules are over expressed in inflammatory cells including the blood, suggesting that what may be being detected are transcripts from inflammatory infiltrates that manifest following tissue damage. Notably this is consistent with observations that co-regulation modules in the liver are related, in part to changes in cellularity (i.e., increases or decreases in certain cell types; Sutherland et al., 2017). Chemicals that act through a receptor have cascades of signaling that often attenuate cell death by inhibiting apoptosis

TABLE 4 | Preserved modules.

Module	Module size	Z summary
Turquoise	540	27.27
Red	176	16.93
Blue	325	15.00
Magenta	131	11.92

TABLE 5 | Pathway enrichment.

Module	Panther db ID	Biological process	Count	Fold enrichment	*p*-value	FDR
Turquoise	BP00148	Immunity and defense	68	2.38	1.48E-11	1.8E-08
Red	BP00148	Immunity and defense	54	1.72	8.37E-05	1.0E-01
Blue	BP00125	Intracellular protein traffic	33	1.66	4.93E-03	5.8E+00
Magenta	BP00148	Immunity and defense	52	1.66	2.73E-04	3.3E-01

FIGURE 5 | TP53 Interaction network. Using the 540 gene probe sets (462 genes) from the most significant module (Turquoise) preserved between **(A)** RM and **(B)** NRM samples that were mapped to pathways in the Ingenuity Pathway Analysis knowledgebase, molecular networks were generated. Shown is the most

(Continued)

(Mally and Chipman, 2002) therefore decreasing cell turnover. The decreased cell death may secondarily down-regulate baseline immune signaling as there is less cellular debris to clear (Rock and Kono, 2008). In addition, activation of PPAR-α and CAR/PXR have been demonstrated to down-regulate the expression of compliment and coagulation factors which may also be contributing to the decreased immune signaling seen with the RM chemicals (Kramer et al., 2003; Yadetie et al., 2003; Cariello et al., 2005; Rezen et al., 2009). Hence, it is plausible that the convergence point between these two groups of toxicologic agents occurs at a cellular level and cascades down into the molecular level where opposite effects on inflammatory signaling is observed. Although many gene expression signatures associated with toxicants likely represent cytotoxicity and cell damage, activation of an immune response is not just injury *per-se*, but is very much involved in repair and regeneration. The toxicant gene signatures likely reflect a genomic state in the liver during the process of the ensuing injury vs. the abating of it and beginnings of recovery and repair.

Here we show that with nine RM chemicals and six NRM chemicals, a converging point in one of the gene networks is at the tumor suppressor gene *TP53* (**Figure 5**). Tp53 in rats is a 391 amino acid containing phosphoprotein with an amino-terminal transactivation motif, DNA and zinc binding sites, a tetramerization domain and an unstructured basic domain at the carboxy-terminus. TP53 regulates the cell cycle, it plays a role in apoptosis and DNA repair, and functions as a tumor suppressor. TP53 in humans is highly mutated in cancers (Olivier et al., 2010) and has been explored extensively as a potential target for cancer therapeutics (Parrales and Iwakuma, 2015). In this study of the male rat livers exposed to the RM and NRM chemicals, *Tp53* gene expression is slightly up-regulated relative to control (but not statistically significant with a large enough fold change difference). This is not surprising as a small change in the expression of a transcription factor can dramatically impact the transcriptional regulation of its target genes (Niwa et al., 2000; Rizzino, 2008). In addition, per the IPA knowledgebase molecular network (**Figure 5**), TP53 interacts with 32 genes in the turquoise module; 21 genes with p53 binding sites and the others have molecular relationships such as protein-protein interaction or some form of biochemical modification. Of these 32 hub genes, the majority of them (*n* = 19) are dysregulated in opposite directions in RM vs. NRM. Some of these genes function in metabolism, cell division and DNA repair. This redundancy in the gene network circuitry is thought to be contrapuntal in

nature to provide organisms the flexibility to diversify function while conserving biologic resources (Komili and Silver, 2008). Examples are the coordinated gene expression regulation during seed development in *Arabidopsis thaliana* (Ruuska et al., 2002) and the crosstalk between Janus kinase-signal transducer and activator of transcription (JAK-STAT) and PPAR-α in COS-1 cells derived from monkey kidney tissue (Zhou and Waxman, 1999).

Although, the number of chemicals per RM and NRM MOAs limits the granularity in the reconstruction of the networks we ascertained as preserved between the two classes, the diversity in the types of chemicals, the varied structure activity groups, and broad therapeutic indications of the chemicals give credence to the biological interpretation of the molecular pathways in common but coordinately dysregulated. Furthermore, despite the incohesiveness of a few of the chemicals which did not cluster by gene expression with the other chemicals in their respective MOA (**Figure 2**), the bioinformatics processing of the data that we employed (**Figure 1**) was robust enough to elucidate molecular interaction networks that converge between the RM and NRM chemicals. However, caution in the interpretation of these results is prudent since we have not examined the entire scope of all the chemicals that fall into a given MOA and some chemicals are known to have multiple MOAs (Russom et al., 1997; Wenzel et al., 1997; Freidig et al., 1999). In addition, the comparison of RM and NRM MOAs is a simplistic one and each class does not cover all the RM or NRM MOAs, therefore is limited in the inference of the gene regulatory networks. Furthermore, the doses of the chemicals administered are of a single concentration and duration albeit the MTD and so essentially the preserved gene networks that we discovered are not dynamic in nature. It is important to emphasize that the gene modules described here are a starting point for MOA characterization and greater nuance will likely be required to characterize mechanistic processes associated with specific receptors (e.g., PPAR-α vs. AhR; LeBaron et al., 2014; Becker et al., 2015) or chemicals with mixed MOAs. An illustration of such nuance was shown in a recent study in which clear subgroups of chemicals in the RM class of compounds was observed (De Abrew et al., 2015). Further, it is important to note that careful consideration of the interpretive approach is necessary when evaluating MOA (Currie et al., 2014). In conclusion, our data and results provide a framework for investigators to follow-up on to possibly perturb individual components of biological pathways that read across between chemicals with different MOAs in order to better understand the consequences of environmental exposures.

AUTHOR CONTRIBUTIONS

KF performed most of the analyses. PB conceptualized the analysis strategy, directed the analyses, performed some analyses, interpreted the results and wrote the paper. SA provided the samples that the data were generated from, interpreted the results and provided biological/toxicological context.

ACKNOWLEDGMENTS

REFERENCES

Amin, R. P., Hamadeh, H. K., Bushel, P. R., Bennett, L., Afshari, C. A., and Paules, R. S. (2002). Genomic interrogation of mechanism(s) underlying cellular responses to toxicants. *Toxicology* 181–182, 555–563. doi: 10.1016/S0300-483X(02)00481-X

Barrett, T., Wilhite, S. E., Ledoux, P., Evangelista, C., Kim, I. F., Tomashevsky, M., et al. (2013). NCBI GEO: archive for functional genomics data sets–update. *Nucleic Acids Res.* 41, D991–D995. doi: 10.1093/nar/gks1193

Becker, R. A., Patlewicz, G., Simon, T. W., Rowlands, J. C., and Budinsky, R. A. (2015). The adverse outcome pathway for rodent liver tumor promotion by sustained activation of the aryl hydrocarbon receptor. *Regul. Toxicol. Pharmacol.* 73, 172–190. doi: 10.1016/j.yrtph.2015.06.015

Bushel, P. R., Hamadeh, H. K., Bennett, L., Green, J., Ableson, A., Misener, S., et al. (2002). Computational selection of distinct class- and subclass-specific gene expression signatures. *J. Biomed. Inform.* 35, 160–170. doi: 10.1016/S1532-0464(02)00525-7

Cariello, N. F., Romach, E. H., Colton, H. M., Ni, H., Yoon, L., Falls, J. G., et al. (2005). Gene expression profiling of the PPAR-alpha agonist ciprofibrate in the cynomolgus monkey liver. *Toxicol. Sci.* 88, 250–264. doi: 10.1093/toxsci/kfi273

Casarett, L. J., Doull, J., and Klaassen, C. D. (2001). *Casarett and Doull's Toxicology: The Basic Science of Poisons*. New York, NY: McGraw-Hill Medical Pub. Division.

Currie, R. A., Peffer, R. C., Goetz, A. K., Omiecinski, C. J., and Goodman, J. I. (2014). Phenobarbital and propiconazole toxicogenomic profiles in mice show major similarities consistent with the key role that constitutive androstane receptor (CAR) activation plays in their mode of action. *Toxicology* 321, 80–88. doi: 10.1016/j.tox.2014.03.003

De Abrew, K. N., Overmann, G. J., Adams, R. L., Tiesman, J. P., Dunavent, J., Shan, Y. K., et al. (2015). A novel transcriptomics based *in vitro* method to compare and predict hepatotoxicity based on mode of action. *Toxicology* 328, 29–39. doi: 10.1016/j.tox.2014.11.008

Edgar, R., Domrachev, M., and Lash, A. E. (2002). Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* 30, 207–210. doi: 10.1093/nar/30.1.207

Freidig, A. P., Verhaar, H. J. M., and Hermens, J. L. M. (1999). Comparing the potency of chemicals with multiple modes of action in aquatic toxicology: acute toxicity due to narcosis versus reactive toxicity of acrylic compounds. *Environ. Sci. Technol.* 33, 3038–3043. doi: 10.1021/es990251b

Hamadeh, H. K., Bushel, P. R., Jayadev, S., Disorbo, O., Bennett, L., Li, L., et al. (2002a). Prediction of compound signature using high density gene expression profiling. *Toxicol. Sci.* 67, 232–240. doi: 10.1093/toxsci/67.2.232

Hamadeh, H. K., Bushel, P. R., Jayadev, S., Martin, K., Disorbo, O., Sieber, S., et al. (2002b). Gene expression analysis reveals chemical-specific profiles. *Toxicol. Sci.* 67, 219–231. doi: 10.1093/toxsci/67.2.219

Irizarry, R. A., Bolstad, B. M., Collin, F., Cope, L. M., Hobbs, B., and Speed, T. P. (2003a). Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res.* 31:e15. doi: 10.1093/nar/gng015

Irizarry, R. A., Hobbs, B., Collin, F., Beazer-Barclay, Y. D., Antonellis, K. J., Scherf, U., et al. (2003b). Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* 4, 249–264. doi: 10.1093/biostatistics/4.2.249

Karlebach, G., and Shamir, R. (2008). Modelling and analysis of gene regulatory networks. *Nat. Rev. Mol. Cell Biol.* 9, 770–780. doi: 10.1038/nrm2503

Kleinjans, J. (2014). *Toxicogenomics-Based Cellular Models: Alternatives to Animal Testing for Safety Assessment*. Amsterdam: Elsevier/AP.

Komili, S., and Silver, P. A. (2008). Coupling and coordination in gene expression processes: a systems biology view. *Nat. Rev. Genet.* 9, 38–48. doi: 10.1038/nrg2223

Kramer, J. A., Blomme, E. A., Bunch, R. T., Davila, J. C., Jackson, C. J., Jones, P. F., et al. (2003). Transcription profiling distinguishes dose-dependent effects in the livers of rats treated with clofibrate. *Toxicol. Pathol.* 31, 417–431. doi: 10.1080/01926230390202353

Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 9:559. doi: 10.1186/1471-2105-9-559

Langfelder, P., Luo, R., Oldham, M. C., and Horvath, S. (2011). Is my network module preserved and reproducible? *PLoS Comput. Biol.* 7:e1001057. doi: 10.1371/journal.pcbi.1001057

LeBaron, M. J., Gollapudi, B. B., Terry, C., Billington, R., and Rasoulpour, R. J. (2014). Human relevance framework for rodent liver tumors induced by the insecticide sulfoxaflor. *Crit. Rev. Toxicol.* 44(Suppl. 2), 15–24. doi: 10.3109/10408444.2014.910751

Lu, J., Kerns, R. T., Peddada, S. D., and Bushel, P. R. (2011). Principal component analysis-based filtering improves detection for Affymetrix gene expression arrays. *Nucleic Acids Res.* 39:e86. doi: 10.1093/nar/gkr241

Mally, A., and Chipman, J. K. (2002). Non-genotoxic carcinogens: early effects on gap junctions, cell proliferation and apoptosis in the rat. *Toxicology* 180, 233–248. doi: 10.1016/S0300-483X(02)00393-1

Nijman, S. M. (2015). Functional genomics to uncover drug mechanism of action. *Nat. Chem. Biol.* 11, 942–948. doi: 10.1038/nchembio.1963

Niwa, H., Miyazaki, J., and Smith, A. G. (2000). Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. *Nat. Genet.* 24, 372–376. doi: 10.1038/74199

Olivier, M., Hollstein, M., and Hainaut, P. (2010). TP53 mutations in human cancers: origins, consequences, and clinical use. *Cold Spring Harb. Perspect. Biol.* 2:a001008. doi: 10.1101/cshperspect.a001008

Parrales, A., and Iwakuma, T. (2015). Targeting oncogenic mutant p53 for cancer therapy. *Front. Oncol.* 5:288. doi: 10.3389/fonc.2015.00288

Rappaport, S. M., and Smith, M. T. (2010). Epidemiology. Environment and disease risks. *Science* 330, 460–461. doi: 10.1126/science.1192603

Rezen, T., Tamasi, V., Lovgren-Sandblom, A., Bjorkhem, I., Meyer, U. A., and Rozman, D. (2009). Effect of CAR activation on selected metabolic pathways in normal and hyperlipidemic mouse livers. *BMC Genomics* 10:384. doi: 10.1186/1471-2164-10-384

Rizzino, A. (2008). Transcription factors that behave as master regulators during mammalian embryogenesis function as molecular rheostats. *Biochem. J.* 411, e5–e7. doi: 10.1042/BJ20080479

Rock, K. L., and Kono, H. (2008). The inflammatory response to cell death. *Annu. Rev. Pathol.* 3, 99–126. doi: 10.1146/annurev.pathmechdis.3.121806.151456

Russom, C. L., Bradbury, S. P., Broderius, S. J., Hammermeister, D. E., and Drummond, R. A. (1997). Predicting modes of toxic action from chemical structure: acute toxicity in the fathead minnow (*Pimephales promelas*). *Environ. Toxicol. Chem.* 16, 948–967. doi: 10.1002/etc.5620160514

Ruuska, S. A., Girke, T., Benning, C., and Ohlrogge, J. B. (2002). Contrapuntal networks of gene expression during Arabidopsis seed filling. *Plant Cell* 14, 1191–1206. doi: 10.1105/tpc.000877

Srikrishna, G., and Freeze, H. H. (2009). Endogenous damage-associated molecular pattern molecules at the crossroads of inflammation and cancer. *Neoplasia* 11, 615–628. doi: 10.1593/neo.09284

Sutherland, J. J., Webster, Y. W., Willy, J. A., Searfoss, G. H., Goldstein, K. M., Irizarry, A. R., et al. (2017). Toxicogenomic module associations with

pathogenesis: a network-based approach to understanding drug toxicity. *Pharmacogenomics J.* doi: 10.1038/tpj.2017.17

Thomas, P. D., Campbell, M. J., Kejariwal, A., Mi, H., Karlak, B., Daverman, R., et al. (2003). PANTHER: a library of protein families and subfamilies indexed by function. *Genome Res.* 13, 2129–2141. doi: 10.1101/gr.772403

Wang, C., Gong, B., Bushel, P. R., Thierry-Mieg, J., Thierry-Mieg, D., Xu, J., et al. (2014). The concordance between RNA-seq and microarray data depends on chemical treatment and transcript abundance. *Nat. Biotechnol.* 32, 926–932. doi: 10.1038/nbt.3001

Wei, X., Ai, J., Deng, Y., Guan, X., Johnson, D. R., Ang, C. Y., et al. (2014). Identification of biomarkers that distinguish chemical contaminants based on gene expression profiles. *BMC Genomics* 15:248. doi: 10.1186/1471-2164-15-248

Wenzel, A., Nendza, M., Hartmann, P., and Kanne, R. (1997). Testbattery for the assessment of aquatic toxicity. *Chemosphere* 35, 307–322. doi: 10.1016/S0045-6535(97)00157-4

Wild, C. P. (2005). Complementing the genome with an "exposome": the outstanding challenge of environmental exposure measurement in molecular epidemiology. *Cancer Epidemiol. Biomark. Prev.* 14, 1847–1850. doi: 10.1158/1055-9965.EPI-05-0456

Woods, C. G., Heuvel, J. P., and Rusyn, I. (2007). Genomic profiling in nuclear receptor-mediated toxicity. *Toxicol. Pathol.* 35, 474–494. doi: 10.1080/01926230701311351

Yadetie, F., Laegreid, A., Bakke, I., Kusnierczyk, W., Komorowski, J., Waldum, H. L., et al. (2003). Liver gene expression in rats in response to the peroxisome proliferator-activated receptor-alpha agonist ciprofibrate. *Physiol. Genomics* 15, 9–19. doi: 10.1152/physiolgenomics.00064.2003

Yip, A. M., and Horvath, S. (2007). Gene network interconnectedness and the generalized topological overlap measure. *BMC Bioinformatics* 8:22. doi: 10.1186/1471-2105-8-22

Zhang, B., and Horvath, S. (2005). A general framework for weighted gene co-expression network analysis. *Stat. Appl. Genet. Mol. Biol.* 4:17. doi: 10.2202/1544-6115.1128

Zhou, Y. C., and Waxman, D. J. (1999). Cross-talk between janus kinase-signal transducer and activator of transcription (JAK-STAT) and peroxisome proliferator-activated receptor-alpha (PPARalpha) signaling pathways. Growth hormone inhibition of pparalpha transcriptional activity mediated by stat5b. *J. Biol. Chem.* 274, 2672–2681. doi: 10.1074/jbc.274.5.2672

Zearalenone Exposure Enhanced the Expression of Tumorigenesis Genes in Donkey Granulosa Cells via the *PTEN/PI3K/AKT* Signaling Pathway

Guo-Liang Zhang[1,2], Jun-Lin Song[3], Chuan-Liang Ji[2], Yu-Long Feng[2], Jie Yu[2], Charles M. Nyachoti[4] and Gong-She Yang[1]*

[1] College of Animal Science and Technology, Northwest A&F University, Yangling, China, [2] National Engineering Research Center for Gelatin-based Traditional Chinese Medicine, Dong-E-E-Jiao Co., Ltd., Liaocheng, China, [3] Central Laboratory, Qingdao Agricultural University, Qingdao, China, [4] Department of Animal Science, University of Manitoba, Winnipeg, MB, Canada

Zearalenone (ZEA) is a natural contaminant existing in food and feed products that exhibits a negative effect on domestic animals' reproduction. Donkeys possess high economic value in China and are at risk of exposure to ZEA. However, few information is available on ZEA-induced toxicity and no report on toxicity in donkeys can be found in scientific literature. We investigated the biological effects of ZEA exposure on donkey granulosa cells (dGCs) by using RNA-seq analysis. ZEA at 10 and 30 μM were administered to GCs within 72 h of *in vitro* culture. ZEA at 10 μM significantly altered the tumorigenesis associated genes in dGCs. Exposure to 10 and 30 μM ZEA treatment significantly reduced mRNA expression of *PTEN*, *TGF*β, *ATM*, and *CDK2* genes, particularly, the ZEA treatment significantly increased the expression of *PI3K* and *AKT* genes. Furthermore, immunofluorescence, RT-qPCR, and Western blot analysis verified the gene expression of ZEA-exposed GCs. Collectively, these results demonstrated the deleterious effect of ZEA exposure on the induction of ovarian cancer related genes via the *PTEN/PI3K/AKT* signaling pathway in dGCs *in vitro*.

Keywords: donkey, granulosa cells, tumorigenesis, gene expression, RNA-seq

INTRODUCTION

Zearalenone (ZEA) is a mycotoxin produced by various Fusarium fungi (Bennett and Klich, 2003) that infects grains and maize worldwide. Similar to aflatoxins (AFs), ZEA is one of the most important and widespread trichothecenes that cause extensive and recurring economic damage in cereal grains and animal feedstuffs (Escriva et al., 2015). In domestic animals, ZEA causes porcine ovarian atrophy (Vanyi et al., 1994) and equine follicular hematomas (Cortinovis et al., 2013) and exhibits significant genotoxic potential and induces DNA damage (Zhang et al., 2017a) in experimental animals (mice and rats). Owing to its estrogenic activity, ZEA could cause reproductive disorders in a wide variety of species-specific organs in animals (Poor et al., 2015). Both low and high concentrations of ZEA can cause abortion and reproductive failure in livestock (Osweiler et al., 1990; Dacasto et al., 1995; Zwierzchowski et al., 2005). Moreover, ZEA is a potential carcinogen with a possible correlation to xenoestrogens and breast cancer risk (Yu et al., 2004).

The effects of ZEA and some of its metabolites on estrous mares have only been reported in a few cases. Juhasz et al. (2001) reported the effects of 10 day low dose ZEA exposure on the reproductive system of ovulating mares fed with 7 mg purified ZEA. No effect was observed on the length of the inter-ovulatory intervals, follicular phases, and ovarian follicular activity (the maximum size and number of the ovulatory follicles, growth rate, and the initial increase in the number of large follicles) (Juhasz et al., 2001). Another study showed that feeding mares with oats that were naturally contaminated with ZEA (12 mg/kg) had no relevant effects on the cycle length and the release of reproductive hormones (Cortinovis et al., 2013). However, the authors stated that mares fed with the naturally contaminated oats had a high incidence of follicular hematomas, which did not occur in the control (Cortinovis et al., 2013). Similar to the findings in estrous pig (Zhang et al., 2017b), ZEA and its derivatives induced *in vitro* apoptosis of the granulosa cells (GCs), which were collected from estrous mares' ovaries (Minervini et al., 2006). Hence, the authors suggested that ZEA could induce follicular atresia in domestic animals. These effects could be due to the direct interaction with ERs and 3-α/β-HSD enzymes present in the GCs and ovary, which are responsible for the synthesis or metabolism of the endogenous steroid hormones (Minervini et al., 2006).

The mechanism of ZEA toxicity is not fully understood, but ZEA is known to possess acute and chronic toxic effects in animals. Ovaries are female reproductive organs comprising follicles of varying sizes. The early stages of follicular growth depend on the development of the GCs and oocytes, which are in constant communication with each other. The development of one cell type influences the other's compartment. During follicle development, GCs replicate, secrete hormones, and support the growth of the oocyte (Hamel et al., 2008). Previous investigations demonstrated that ZEA may alter GC's function in swine (Ranzenigo et al., 2008; Zhang et al., 2017b). This study aims to evaluate the in vitro toxicity of 10 and 30 μM ZEA in donkey ovarian GCs through transcriptome analysis.

MATERIALS AND METHODS

Reagents

Zearalenone was purchased from Sigma Company (St. Louis, MO, United States). Stock solutions of ZEA were fixed by dissolving ZEA in dimethyl sulfoxide (DMSO). DMSO (D12345), fetal bovine serum (FBS; 10100147), M-199 medium (11150-059), penicillin and streptomycin were procured from Gibco Company (Carlsbad, CA, United States).

Animals

The mature donkeys' ovaries used in the experiment were obtained from the Dong E Donkey Production Company (Qingdao, Shandong, China). The ovaries were collected from the slaughterhouse of the company and maintained at 32–35°C for the isolation of GCs. All procedures of animal handling in this study were reviewed and approved by the Ethical Committee (Agreement No. 2017-18) of Qingdao Agricultural University.

Isolation and Culture of dGCs

Donkey GCs (dGCs) were aspirated from the antral follicles using a 10 ml syringe (Zhu et al., 2012). Standing for 15–18 min, the dGCs were centrifuged at 300 g for 5 min in accordance with previous report (Qin et al., 2015). Then the dGCs were cultured in DMEM medium (HyClone, SH30022.01, Beijing, China) supplement with 10% FBS (10099-141, Gibco, Australia) and 1% penicillin-streptomycin (Hyclone, SV30010) in incubator with 5% CO_2 at 37°C (Duda et al., 2011).

The primary GCs were passaged after culture 48–36 h. To avoid the stress of passage response, drug exposure was performed until 12 h later. The GCs were inoculation in a 6 cm petri dish (Corning, 430166, United States) at a density of 1×10^6 cells. ZEA was added to the cultured medium at final concentrations of the 10 or 30 μM, then the cells were incubated with ZEA for 72 h. The control and 10 μM ZEA group added the same dose of DMSO to 30 μM ZEA group for accuracy.

Immunofluorescence and Cell Counting

The GCs were collected and fixed in 4% paraformaldehyde for 2 h, then heated at 42°C for 2 h, finally attached onto a poly-lysine coated slide. For immunofluorescence, the sections were blocked with the BDT (10% goat serum in TBS, 3% BSA) for 35 min, and then incubated overnight with primary antibodies at 4°C (**Supplementary Table S1**). The sections were then incubated with Cy3/FITC-conjugated goat anti-rabbit secondary antibody (Beyotime, A0208, Nantong, China) at a dilution of 1:50 at 37°C for 1.5 h. Finally, the sections were incubated with Hoechst33342 (Beyotime, C1022) to visualize nuclei of GCs for 3 min at room temperature. The immunosignal was detected using a fluorescence microscope (Olympus, XB51, Japan), the images were captured and analyzed in accordance with cellSens Standard.

TUNEL Staining

TUNEL BrightRed Apoptosis Detection Kit was utilized to evaluate the GCs apoptosis (Vazyme, A11302, Nanjing, China). Briefly, the GCs were fixed with 4% paraformaldehyde for 2 h after ZEA treatment. Observation under fluorescence microscopy was carried out after TUNEL reaction following manufacturer's recommendations [60 min at 37°C without light, and DNA staining with Hoechst 33342 (Beyotime, C1022) incubation at room temperature for 3 min]. The TUNEL positive cells were detected under the fluorescence microscope (Olympus, XB51). To analyze TUNEL positive cell ratio, more than 2,000 cells were counted in each group, and each group included three biological replicates at least.

RNA Extraction, Reverse Transcription, and RNA-seq

Total RNA was extracted from the cultured GCs by the use of an RNAprep pure MicroKit in line with the manufacturer's instruction (Aidlab, RN07, Beijing, China). And the first-strand cDNA acquisition was utilized a cDNA Synthesis Kit (TransGen, AT311-03, Beijing, China) with reference to previous study (Pan et al., 2011). The reaction program was setup at: 42°C for 15 min,

85°C for 30 s, and 4°C for cooling. Then, RNA sequencing was performed with Hiseq 4000 platform by Novogene (Beijing, China). Three biological replicates were manipulated in each group. Primary RNA-seq data were uploaded to the SRA repository. The SRA index number was SRP139976. The RNA-seq data matrix used for DESeq2 analysis were uploaded to the GEO repository and the GEO accession number was GSE116950[1].

Data Preprocessing and Identification of Differentially Expressed Genes

R Bioconductor/DESeq2 package was applied to analyze the dGCs groups (control, 10 and 30 μM ZEA) to identify the difference of gene expression. DESeq2, the negative binomial was applied as the reference distribution and taken own normalization approach for raw counts in differential expression analysis. In other methods to avoid possible biases, the data of differential expression analysis has to be previously normalized (Benjamini et al., 2001; Luo and Brouwer, 2013). Because the design of sequencing contains biological replicates in each group, log2|fold change| is not setup as the filter condition. So $p_{adj} < 0.01$ was really considered as statistical significance.

KEGG Enrichment Analysis

R Bioconductor/clusterProfiler package was applied to analyze functional profiles (KEGG) of differentially expressed genes (DEGs) (Yu et al., 2012). Furthermore, R Bioconductor/Pathview package was applied to visualize KEGG enrichment results. According to the DEGs log2|fold change| value shows enrichment signaling pathway active status. The p-value adjusted used the Benjamini & Hochberg (BH) method (Luo and Brouwer, 2013). Still, the $p_{adj} < 0.05$ was considered as statistical significance.

Gene Set Enrichment Analysis

Gene set enrichment analysis (GSEA) does not need to specify a clearly differential expression of gene threshold, the algorithm performs GSEA based on overall trend of the gene raw read count. The file has to be made suitable for GSEA software analysis with R script (Subramanian et al., 2005). The data for GSEA analysis contained nine samples within three groups. The GSEA software option "Collapse dataset to gene symbols" parameters was "false" and option "Permutation type" parameters were "gene set." Other parameters were set as default as software manual references. The GSEA gene set with FDR q-value <0.05 were defined as significant difference.

Protein–Protein Interaction Network Construction and Modules Mining

Search Tool for the Retrieval of Interacting Genes/Proteins[2] (STRING) is a database of protein–protein interaction (PPI). This database contains the direct and physically related interactions between known and predicted protein and genes. The R

[1]https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE116950
[2]https://string-db.org/

Bioconductor/STRINGdb was applied for PPI of interested DEGs (Franceschini et al., 2013). The Cytoscape software was applied to visualize PPI results (Shannon et al., 2003).

Quantitative Real-Time PCR

Total RNA and cDNA reverse transcription was extracted as described above. Quantitative real-time PCR (RT-qPCR) was performed using Light-Cycler® 480 SYBR Green I Master Kit (Roche, Germany) on LightCycler 480 real-time PCR instrument. The RT-qPCR reaction was set as: 10 min at 95°C, followed by 45 cycles of 95°C for 10 s, 60°C for 30 s, 72°C for 30 s, and cooling step at 4°C. Quantitative RT-PCR primers are listed in **Supplementary Table S2** and the primer efficiency in the **Supplementary Figure S1**. The standard curve gene *GAPDH* in dGCs was used as the reference to normalize the related genes' mRNA expression. The difference multiples = $2^{-\Delta\Delta CT}$ method was employed for relative quantification of PCR. Each gene was expressed as the mean ± standard deviation (SD), which was calculated from independent biological replicates at least three times.

Western Blotting

Protein lysates isolated from dGCs were used for western blotting according to the standard protocol (Zhang et al., 2010; Chao et al., 2012). Proteins from each ZEA treatment were separated by 10% SDS-PAGE, then transferred onto the PVDF membranes. After, the membranes were incubated with primary antibodies (**Supplementary Table S1**) at 4°C overnight. Then the membranes were incubated with secondary antibodies (Beyotime, A0208) at 37°C for 2 h in TBST after rinsing three times with TBST. The final detection of related genes was carried out by AlphaImager® HP (ProteinSimple, 92-13824-00, United States). The band intensity was quantified using *GAPDH* as internal control and measured with IPWIN software.

Statistical Methods

Data are presented by mean ± SD. Different effects between the control and ZEA treatment groups of donkey was statistically determined by One-way ANOVA for multiple comparisons. All analyses were conducted using Graph-Pad Prism analysis software (San Diego, CA, United States). All experiments were repeated at least three times unless otherwise noted. Results were considered statistically significant at p-value <0.05.

RESULTS

The Apoptosis and Apoptosis-Related Gene Expression of dGCs Exposed to ZEA

Donkey granulosa cells were cultured *in vitro* and exposed to 10 or 30 μM ZEA for 72 h (**Figure 1A**). The percentages of TUNEL positive dGCs significantly increased as a result of exposure of ZEA (10 μM: 36.04 ± 1.52%; 30 μM: 44.30 ± 1.33%) compared to that of the control (0 μM: 9.83 ± 0.21%; $P < 0.01$; **Figure 1B**). As shown in **Figure 1C**, the ratios of *BAX/BCL2* mRNA

FIGURE 1 | Zearalenone exposure increasing apoptosis and inducing the apoptosis-related gene mRNA abundance in cultured GCs. GCs were stained blue with Hoechst33342 Solution. TUNEL assay was performed using immunostaining. **(A)** Immunofluorescent staining of TUNEL and Hoechst33342 of GCs. Bar indicates 50 μm. **(B)** The percentages of TUNEL positive GCs. **(C)** The mRNA levels of *BAX/BCL2* in cultured GCs exposed to ZEA, and the control determined by Q-PCR. The results are presented as mean ± SD. All experiments were repeated at least three times. *$P < 0.05$; **$P < 0.01$.

expression significantly increased in 10 and 30 μM ZEA exposed dGCs.

We performed RNA-seq to verify the effect of ZEA exposure on dGCs. Based on the criterion FDR <0.05, 14,506 DEGs was observed between the control and the ZEA-treated dGCs (**Figure 2A**). A total of 7,253 and 6,984 DEGs were noted between the control and 10 and 30 μM ZEA treatment groups, respectively. A total of 269 DEGs were observed between the 10 and 30 μM ZEA treatment groups. In this study, the DEGs between the control and ZEA treatment groups and among the treatment groups were obtained from: 0 μM vs. 10 or 30 μM and 10 vs. 30 μM ZEA group, respectively (**Figure 3A**).

We obtained three RNA-seq replicates for the 0, 10, and 30 μM ZEA-treated dGCs. The variation of the replicates from the control and ZEA treatment groups are shown in **Figure 2B**. A heat map was drawn from the DEG results (**Figure 2C**). In this study, we chose the DEGs from the 0 μM vs. 10 and 30 μM groups.

To explore the potential mechanism of ZEA exposure in dGCs, the STRING database was applied in annotating

functional interactions DEGs between the control and ZEA-treatment group. Also PPI networks was visualized by Cytoscape (**Supplementary Figures S2A,B**). The center nodes of the networks, some interesting genes, were observed, such as *PTEN*, *AKT*, *ATM*, *TGF*β, *PI3K*, *CCND2*, *HEY2*, *CDK2*, and *FDX1*. In addition, the cBioPortal was used to provide visualization, analysis of DEGs related to the ovarian cancer from large-scale cancer genomics data sets (**Supplementary Figures S2C, S3**).

DEGs Involved in KEGG Pathways

In order to attain functional insights of DEGs, the R package of clusterProfiler was carried out to establish extremely affected KEGG pathways. The Venn diagram were constructed by the results of DEGs (**Figure 3A**). A total of 5,863 DEGs containing in control and 30 μM ZEA groups were quantified (**Figure 3B**). The DEGs were significantly enriched in the *PI3K-AKT* signaling pathway (Count = 125, p_{adj} = 0.0049), endocytosis (Count = 91, p_{adj} = 0.049), regulation of actin cytoskeleton (Count = 90, p_{adj} = 0.0001), and proteoglycans in cancer (Count = 83, p_{adj} = 0.0009) in ZEA treated dGCs (**Figure 3C**).
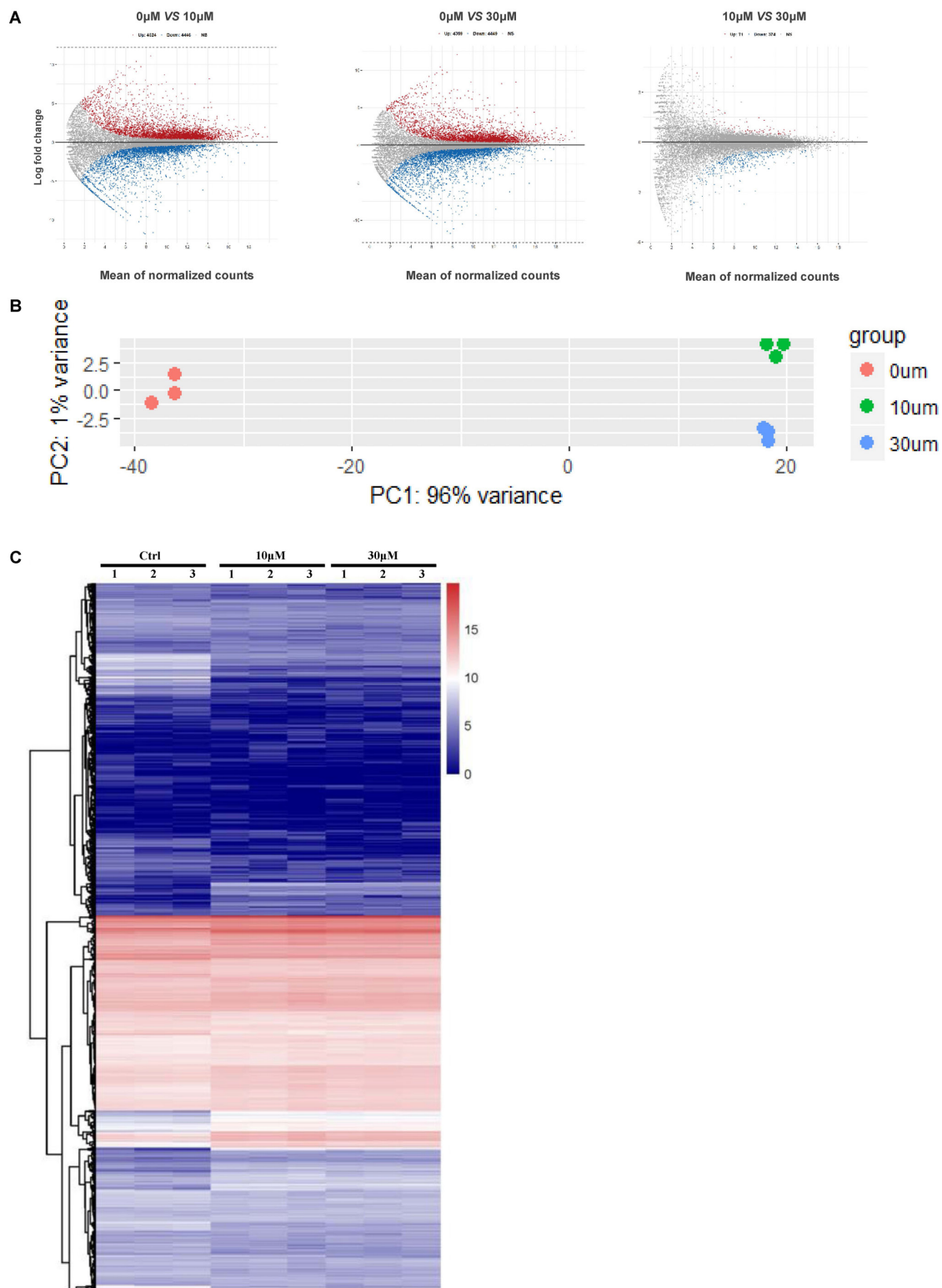
FIGURE 2 | The GCs gene expression profiling after ZEA treatment. **(A)** Scatterplot of gene expression after ZEA treatment. Control group vs. 10 μM ZEA-treatment group, control group vs. 30 μM ZEA-treatment group, 10 μM ZEA-treatment group vs. 30 μM ZEA-treatment group. *Red* and *blue plot* represents genes expressed differently. **(B)** The nine samples shown in the 2D plane spanned by the first three principal components. **(C)** Heatmap indicate the group difference of DEGs in the 10 and 30 μM ZEA-treated groups compared with the control group, and the repeatability within each group. The results are presented as mean ± SD. All experiments were repeated at three times.
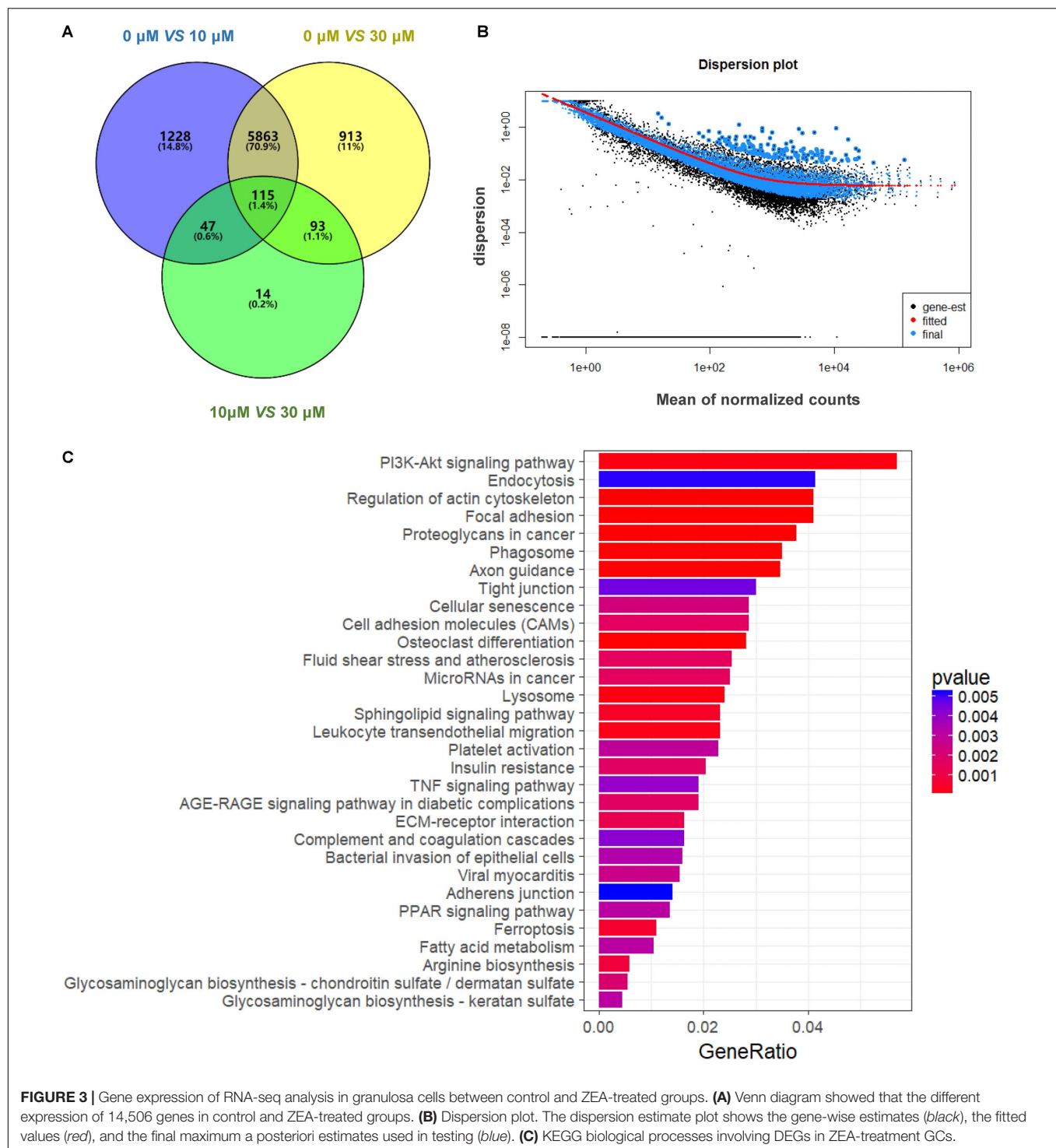
FIGURE 3 | Gene expression of RNA-seq analysis in granulosa cells between control and ZEA-treated groups. **(A)** Venn diagram showed that the different expression of 14,506 genes in control and ZEA-treated groups. **(B)** Dispersion plot. The dispersion estimate plot shows the gene-wise estimates (*black*), the fitted values (*red*), and the final maximum a posteriori estimates used in testing (*blue*). **(C)** KEGG biological processes involving DEGs in ZEA-treatment GCs.

The GSEA analysis showed that *OVARIAN_CANCER_LMP* and *TNF_SIGNALING_VIA_NFKB* gene set was enriched by 10 μM ZEA vs. the control of dGCs (**Figures 4A,B**), while *CANCER_HEAD_AND_NECK_VS_CERVICAL* and *BREAST_CANCER_16Q24_AMPLICON* gene set was enriched by the 30 μM ZEA vs. the control in dGCs (**Figures 4C,D**), respectively. The GSEA enrichment plot revealed a regulated tendency for most of enriched genes in ZEA-treatment compared with the control. The GSEA analysis also verified the KEGG analysis.

In addition, six genes were identified by performing the KEGG and GSEA pathway analysis in ZEA treatment, the *PI3K* and *AKT* genes were involved in *PI3K-AKT* signaling pathway of dGCs upregulating. DGCs in ZEA treatment, another four
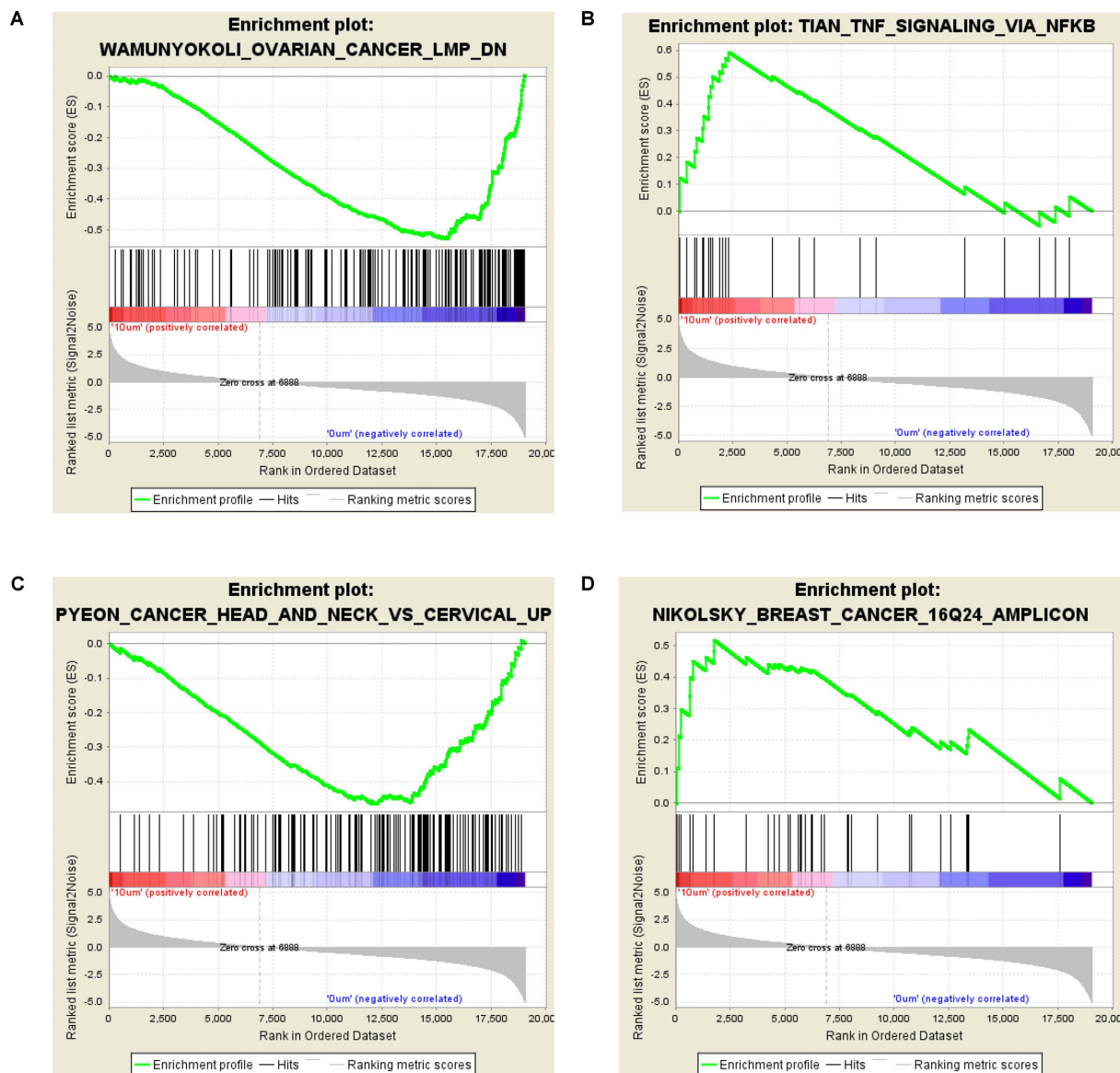
FIGURE 4 | Differentially expressed genes in GSEA analysis. **(A,B)** GSEA for GCs treated with 10 μM ZEA. **(C,D)** GSEA for GCs treated with 30 μM ZEA.

genes involved in anti-oncogene process downregulating were also identified, such as *PTEN*, *TGF*β, *CDK2,* and *ATM*. Finally, *CCND2, CDK6, TNF,* and *TP53* genes in treatment dGCs, were identified involving in cancer process either upregulation or downregulation.

Specific Impact of ZEA Exposure to dGCs

It is hypothesis that 10 and 30 μM ZEA exposure may affect the proliferation and carcinogenesis of dGCs (**Figure 4**). RT-qPCR was conducted to verify the hypothesis by evaluating the expression of different transcript together with enzymes in the pathway of dGCs between the control and ZEA treatments. RT-qPCR analyses indicated that 10 and 30 μM

ZEA exposure significantly downregulated mRNA abundance of *PTEN* while upregulated *PI3K* and *AKT* genes in dGCs ($P < 0.05$ or $P < 0.01$; **Figure 5A**). As illustrated in **Figure 5B**, ZEA-treated dGCs exhibited lower protein levels of *PTEN* but higher protein levels of *PI3K* and *AKT* compared with that of the control dGCs ($P < 0.05$ or $P < 0.01$). Moreover, 10 and 30 μM ZEA exposure significantly down-regulated the mRNA abundance and protein levels of *CDK2*, *TGF*β, and *ATM* genes than the control dGCs ($P < 0.05$ or $P < 0.01$; **Figures 6A,B**). Interestingly, the ZEA treatments significantly decreased the number of *PTEN* and *TGF*β but significantly increased the number of *PI3K* immunofluorescence positive dGCs, in comparison to the control (**Figures 7A,B**).
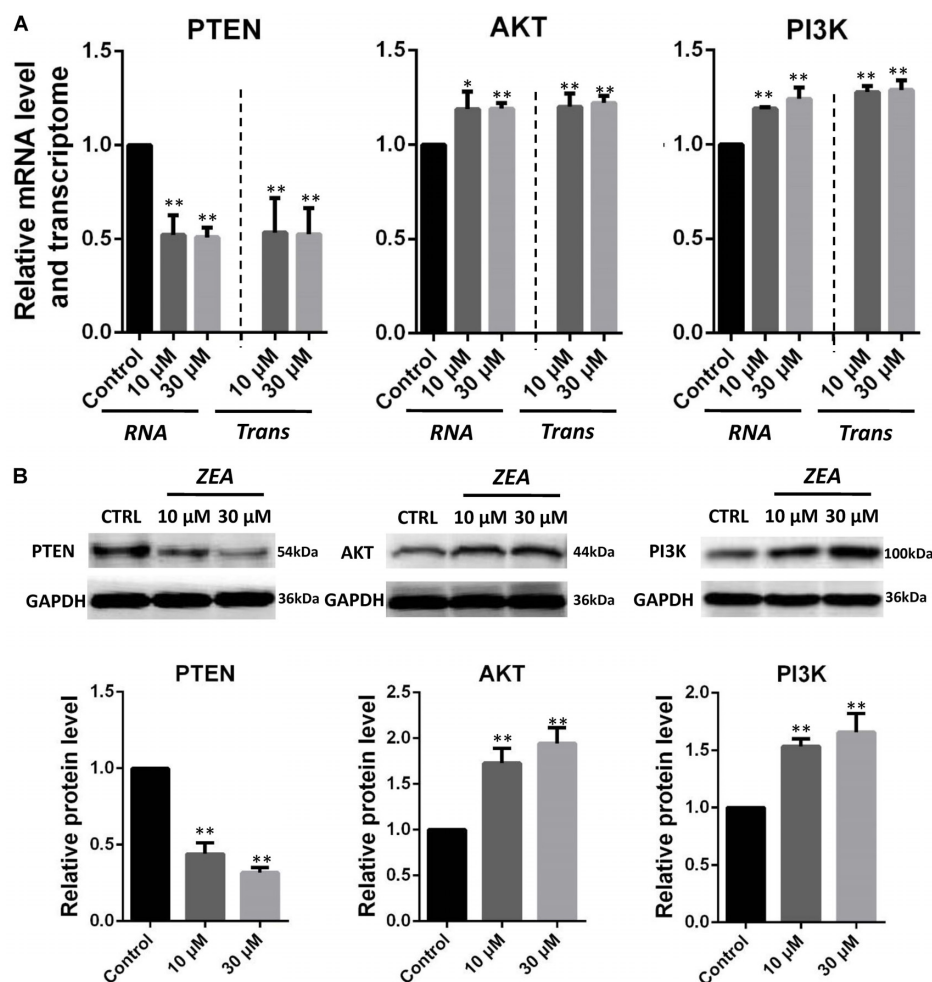
FIGURE 5 | Zearalenone exposure affecting mRNA and protein abundance of *PTEN*-related genes in cultured donkey GCs. **(A)** Quantitative RT-PCR for *PTEN*, *AKT*, and *PI3K* transcription factors. The mRNA levels of all genes were normalized to GCs *GAPDH* gene. **(B)** Protein levels of *PTEN/GAPDH*, *AKT/GAPDH*, and *PI3K/G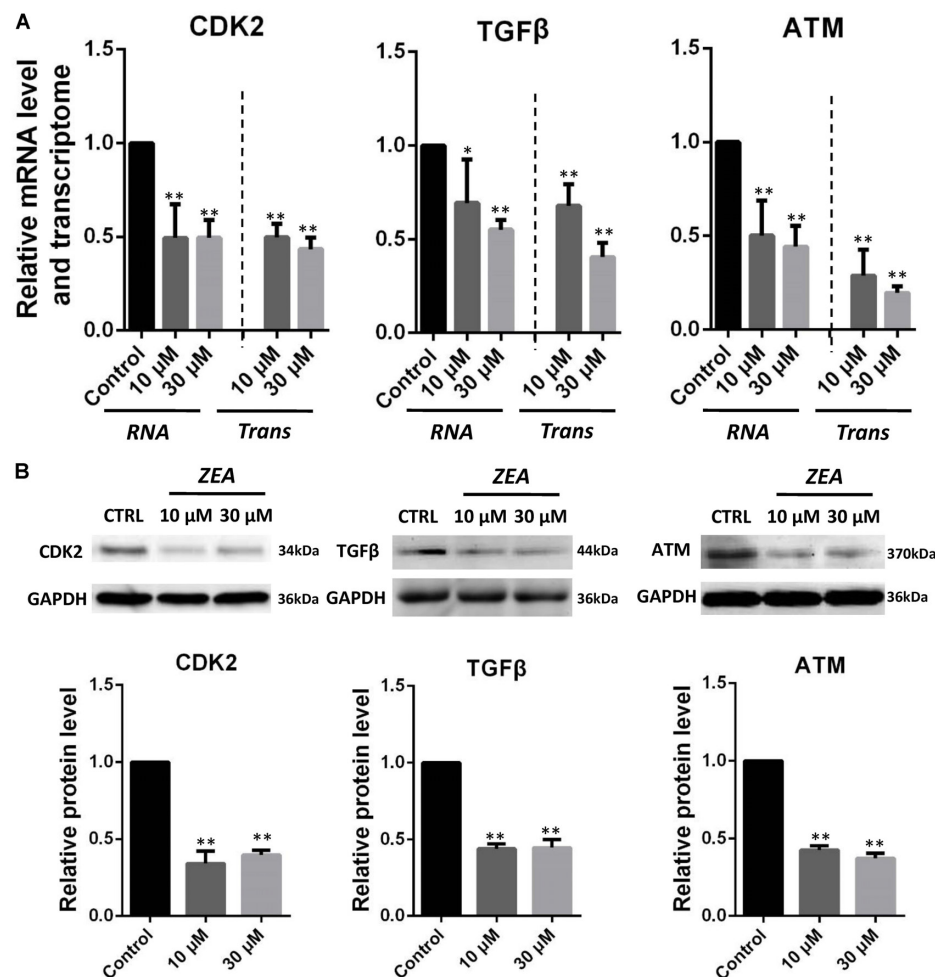APDH* by Western blotting. The protein levels were normalized to *GAPDH*. The results are presented as mean ± SD. All experiments were repeated at least three times. *$P < 0.05$; **$P < 0.01$.

DISCUSSION

Numerous studies have proven ZEA's cytotoxic effect on the reproductive (Kiang et al., 1978; Mehmood et al., 2000; Nikov et al., 2000), immune (Abbes et al., 2006a,b; Luongo et al., 2006), and endocrine (Mueller et al., 2004) systems, as well as on heredity (Kouadio et al., 2005). ZEA indirectly affects mammalian fertility by impairing the formation of primordial follicle (Zhang et al., 2017a), thereby causing changes in gene expression; it also induces DNA damage in ovarian GCs (Zhang et al., 2017b). Not much information is available on ZEA-induced alteration in dGC proliferation and development and impairment of mammalian fertility. This research is the first to describe the GC transcriptomes of donkeys.

The RNA-seq results showed that exposure to 10 μM ZEA significantly transformed the mRNA expression of thousands of genes in dGCs. In particular, *PTEN* genes were

downregulated, suggesting that ZEA exposure is involved with the cancer-related *PTEN* signaling pathway. The tumor suppressor gene, *PTEN*, plays an essential role in cell growth, survival, and tumor formation. Decrease in *PTEN* expression, either at the protein or mRNA level, has been associated with many primary malignancies, including ovarian cancer (Kechagioglou et al., 2014). Hence, the downregulation of *PTEN* is a hallmark of tumors. The *PI3K-AKT* pathway is one of the two major signaling pathways that have been identified as important in cancer. Through phosphorylation, *PI3K-AKT* inhibits the activity of proapoptotic members while activating anti-apoptotic members. The reduction in *PTEN* expression indirectly stimulates *PI3K-AKT* activity, thereby contributing to oncogenesis in mammals (Zhang et al., 2017c). Our studies showed that exposure to 10 μM ZEA significantly upregulated the *PI3K-AKT* gene expression in the dGCs, implying that exposure to low concentrations of ZEA (10 μM) might increase the donkey's risk for ovarian

FIGURE 6 | Zearalenone exposure affecting mRNA and protein abundance of tumorigenesis related genes in cultured GCs. **(A)** Quantitative RT-PCR for *CDK2*, *TGF*β, and *ATM* transcription factors. The mRNA levels of all genes were normalized to GCs *GAPDH* gene. **(B)** Protein levels of *CDK2/GAPDH TGF*β/*GAPDH*, and *ATM/GAPDH* by Western blotting. The protein levels were normalized to *GAPDH*. The results are presented as mean ± SD. All experiments were repeated at least three times. *$P < 0.05$; **$P < 0.01$.

cancer via the *PTEN/PI3K/AKT* pathway by suppressing the expression of antitumor genes or by activating the expression of cancer-causing genes (Zhang et al., 2017c). Moreover, it was reported that decreased *TGF*β-mediated signaling might predispose an individual to develop cancer (Pasche, 2001; Galliher et al., 2006; Jin et al., 2017). Previous studies have assessed the association between *TGF*β and risk for various forms of cancer, and several meta-analyses have demonstrated that *TGF*β is associated with risk for ovarian cancer (Akhurst and Derynck, 2001; Nilsson and Skinner, 2002; de la Cruz-Merino et al., 2009). Interestingly, *TGF*β expression in the ZEA treatment samples was significantly decreased, indicating an increased risk for ovarian cancer. The direct effect of *TGF*β is not solely responsible for influencing tumor behavior (Hagedorn et al., 2001; Tuxhorn et al., 2002). Lack of *TGF*β in fibroblasts can result in mammary gland tumor progression (Bhowmick et al., 2004; Matise et al., 2012).

Bioinformatics analysis confirmed that ataxia-telangiectasia mutated (*ATM*) and cyclin-dependent kinase 2 (*CDK2*) genes were affected by exposure to ZEA. A previous study showed that the *ATM* gene plays an essential role in DNA double-strand breaks (DSBs) repair (Kurz and Lees-Miller, 2004). These DSBs if left unchecked can result in the development of cancer (Weber and Ryan, 2015). Several studies showed that suppression of *ATM* is associated with a variety of tumors (Weber and Ryan, 2015). Here, we provide evidence that exposure to 10 μM ZEA significantly decreased *ATM* gene expression, suggesting a common mechanism of action in donkeys. Other studies revealed that cell cycle dysregulation, resulting in uncontrolled proliferation, is also a hallmark of cancer (Vijayaraghavan et al., 2017). The *CDK* family is composed of proteins associated with cell cycle regulation and is frequently mutated or overexpressed in ovarian cancer. Deregulation of the *CDK2/4/6* signaling pathway is among the most common aberrations found in ovarian cancer

FIGURE 7 | Immunofluorescence assay examining the expression of phosphor-*PTEN*, *TGF*β **(A)**, and *PI3K* **(B)** proteins. The fluorescence intensity and percentages of positive cells were analyzed, respectively. Bar indicates 50 μm. Data are presented as means ± SD. *$P < 0.05$; **$P < 0.01$.
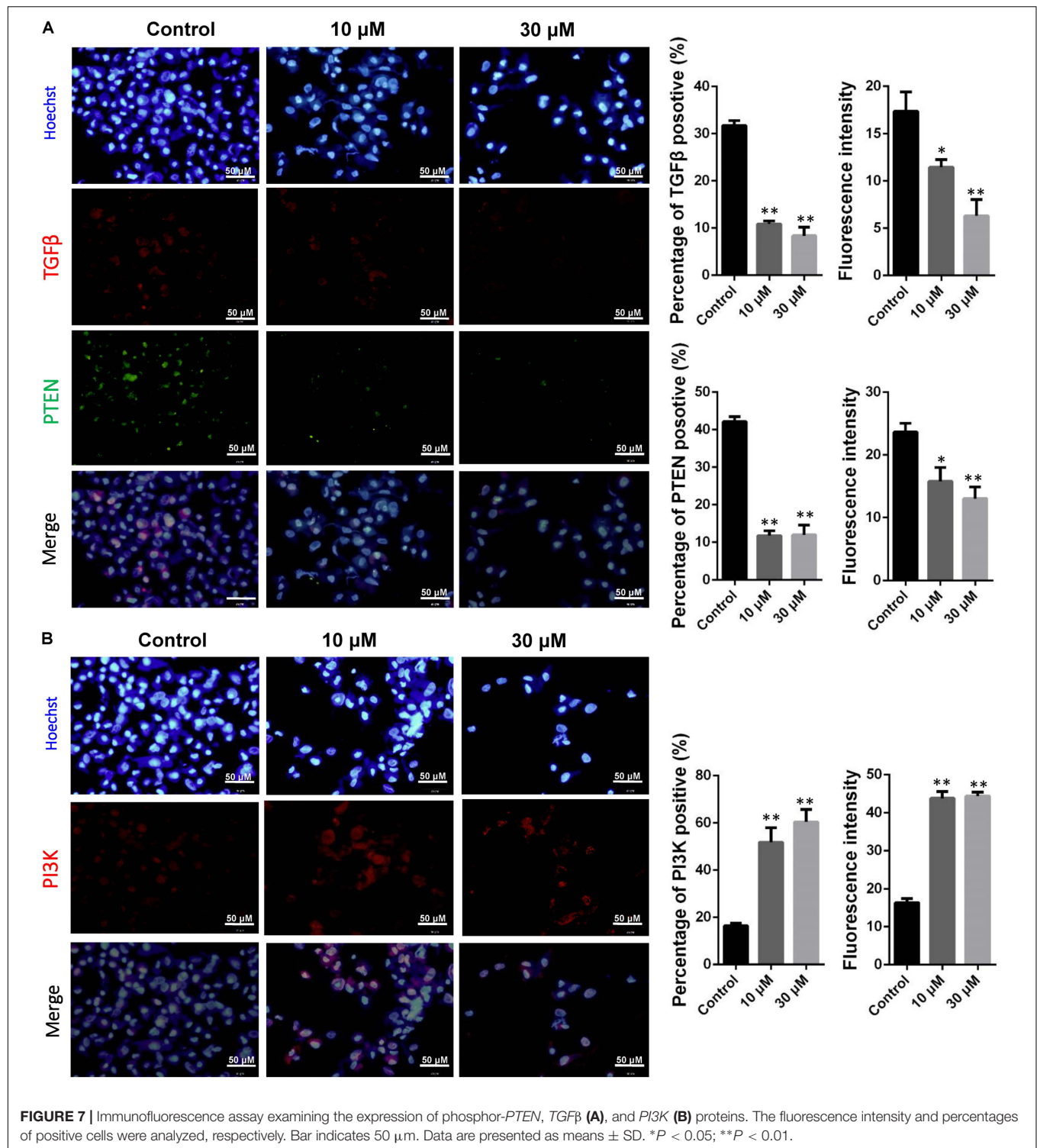
(D'Andrilli et al., 2004). Interestingly, exposure to 10 μM ZEA significantly downregulated the expression of *CDK2* genes, which regulate the cell cycle and are involved in ovarian cancer. Finally, exposure to ZEA increased the dGCs' apoptosis rate and elevated the expression of the *BAX/BCL2* genes.

CONCLUSION

This research is the first study to investigate ZEA-induced impairment of dGCs. ZEA (10 or 30 μM), a potentially carcinogenic substance, can directly cause tumorigenesis and in vitro apoptosis of dGCs. This study developed an innovative,

integrated, and low-cost approach to study GC exposure to ZEA. ZEA disrupts the endocrine and reproductive performance of domestic animals, and this study may help elucidate the mechanism of ZEA toxicity.

ETHICS STATEMENT

All donkeys were treated humanely during slaughter (No. 11002009000012, production license number: SCXK (Qingdao): 2012–0766, Dongeejiao, Qingdao, China). The procedures of animal handling in this study were reviewed and approved by the Ethical Committee of Qingdao Agricultural University (Protocol No. 2017-018).

AUTHOR CONTRIBUTIONS

G-LZ and G-SY designed the study and wrote the manuscript. J-LS, C-LJ, Y-LF, JY, and CN conducted the experiments and analyzed the data.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2018.00293/full#supplementary-material

FIGURE S1 | The standard curve and primer efficiency of all Q-PCR primers. **(A–F)** represent the standard curve and primer efficiency of *PTEN*, *AKT*, *PI3K*, *CDK2*, *TGF*β, and *ATM* genes.

FIGURE S2 | Protein–protein interaction (PPI) network based on the STRING database to annotate functional interactions between DEGs in control and 10 μM ZEA-treatment groups **(A)**, in control and 30 μM ZEA-treatment groups **(B)**. Cancer type detailed summary of DEGs related to the ovarian cancer **(C)**. The node degree ≥20 was selected as the threshold.

FIGURE S3 | Visualization of DEGs related to the ovarian cancer from large-scale cancer genomics data sets. The node degree ≥20 was selected as the threshold.

TABLE S1 | Primary antibodies.

TABLE S2 | Primers used for quantitative-PCR.

REFERENCES

Abbes, S., Ouanes, Z., ben Salah-Abbes, J., Houas, Z., Oueslati, R., Bacha, H., et al. (2006a). The protective effect of hydrated sodium calcium aluminosilicate against haematological, biochemical and pathological changes induced by zearalenone in mice. *Toxicon* 47, 567–574. doi: 10.1016/j.toxicon.2006.01.016

Abbes, S., Salah-Abbes, J. B., Ouanes, Z., Houas, Z., Othman, O., Bacha, H., et al. (2006b). Preventive role of phyllosilicate clay on the immunological and biochemical toxicity of zearalenone in Balb/c mice. *Int. Immunopharmacol.* 6, 1251–1258. doi: 10.1016/j.intimp.2006.03.012

Akhurst, R. J., and Derynck, R. (2001). TGF-beta signaling in cancer–a double-edged sword. *Trends Cell Biol.* 11, S44–S51. doi: 10.1016/S0962-8924(01)02130-4

Benjamini, Y., Drai, D., Elmer, G., Kafkafi, N., and Golani, I. (2001). Controlling the false discovery rate in behavior genetics research. *Behav. Brain Res.* 125, 279–284. doi: 10.1016/S0166-4328(01)00297-2

Bennett, J. W., and Klich, M. (2003). Mycotoxins. *Clin. Microbiol. Rev.* 16, 497–516. doi: 10.1128/CMR.16.3.497-516.2003

Bhowmick, N. A., Chytil, A., Plieth, D., Gorska, A. E., Dumont, N., Shappell, S., et al. (2004). TGF-beta signaling in fibroblasts modulates the oncogenic potential of adjacent epithelia. *Science* 303, 848–851. doi: 10.1126/science.1090922

Chao, H. H., Zhang, X. F., Chen, B., Pan, B., Zhang, L. J., Li, L., et al. (2012). Bisphenol a exposure modifies methylation of imprinted genes in mouse oocytes via the estrogen receptor signaling pathway. *Histochem. Cell Biol.* 137, 249–259. doi: 10.1007/s00418-011-0894-z

Cortinovis, C., Pizzo, F., Spicer, L. J., and Caloni, F. (2013). *Fusarium* mycotoxins: effects on reproductive function in domestic animals–a review. *Theriogenology* 80, 557–564. doi: 10.1016/j.theriogenology.2013.06.018

Dacasto, M., Rolando, P., Nachtmann, C., Ceppa, L., and Nebbia, C. (1995). Zearalenone mycotoxicosis in piglets suckling sows fed contaminated grain. *Vet. Hum. Toxicol.* 37, 359–361.

D'Andrilli, G., Kumar, C., Scambia, G., and Giordano, A. (2004). Cell cycle genes in ovarian cancer: steps toward earlier diagnosis and novel therapies. *Clin. Cancer Res.* 10, 8132–8141. doi: 10.1158/1078-0432.CCR-04-0886

de la Cruz-Merino, L., Henao-Carrasco, F., Garcia-Manrique, T., Fernandez-Salguero, P. M., and Codes-Manuel de Villena, M. (2009). Role of transforming growth factor beta in cancer microenvironment. *Clin. Transl. Oncol.* 11, 715–720. doi: 10.1007/s12094-009-0433-8

Duda, M., Knet, M., Tabarowski, Z., and Slomczynska, M. (2011). Luteal macrophage conditioned medium affects steroidogenesis in porcine granulosa cells. *Reprod. Biol.* 11, 117–134. doi: 10.1016/S1642-431X(12)60049-7

Escriva, L., Font, G., and Manyes, L. (2015). In vivo toxicity studies of *Fusarium* mycotoxins in the last decade: a review. *Food Chem. Toxicol.* 78, 185–206. doi: 10.1016/j.fct.2015.02.005

Franceschini, A., Szklarczyk, D., Frankild, S., Kuhn, M., Simonovic, M., Roth, A., et al. (2013). String v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res.* 41, D808–D815. doi: 10.1093/nar/gks1094

Galliher, A. J., Neil, J. R., and Schiemann, W. P. (2006). Role of transforming growth factor-beta in cancer progression. *Future Oncol.* 2, 743–763. doi: 10.2217/14796694.2.6.743

Hagedorn, H. G., Bachmeier, B. E., and Nerlich, A. G. (2001). Synthesis and degradation of basement membranes and extracellular matrix and their regulation by TGF-beta in invasive carcinomas (Review). *Int. J. Oncol.* 18, 669–681. doi: 10.3892/ijo.18.4.669

Hamel, M., Dufort, I., Robert, C., Gravel, C., Leveille, M. C., Leader, A., et al. (2008). Identification of differentially expressed markers in human follicular cells associated with competent oocytes. *Hum. Reprod.* 23, 1118–1127. doi: 10.1093/humrep/den048

Jin, Y., Chen, W., Yang, H., Yan, Z., Lai, Z., Feng, J., et al. (2017). *Scutellaria barbata* D. Don inhibits migration and invasion of colorectal cancer cells via suppression 435 of PI3K/AKT and TGF-beta/Smad signaling pathways. *Exp. Ther. Med.* 14, 5527–5534. doi: 10.3892/etm.2017.5242

Juhasz, J., Nagy, P., Kulcsar, M., Szigeti, G., Reiczigel, J., and Huszenicza, G. (2001). Effect of low-dose zearalenone exposure on luteal function, follicular activity and uterine oedema in cycling mares. *Acta Vet. Hung.* 49, 211–222. doi: 10.1556/004.49.2001.2.11

Kechagioglou, P., Papi, R. M., Provatopoulou, X., Kalogera, E., Papadimitriou, E., Grigoropoulos, P., et al. (2014). Tumor suppressor PTEN in breast cancer: heterozygosity, mutations and protein expression. *Anticancer Res.* 34, 1387–1400.

Kiang, D. T., Kennedy, B. J., Pathre, S. V., and Mirocha, C. J. (1978). Binding characteristics of zearalenone analogs to estrogen receptors. *Cancer Res.* 38, 3611–3615.

Kouadio, J. H., Mobio, T. A., Baudrimont, I., Moukha, S., Dano, S. D., and Creppy, E. E. (2005). Comparative study of cytotoxicity and oxidative stress induced by deoxynivalenol, zearalenone or fumonisin B1 in human intestinal cell line Caco-2. *Toxicology* 213, 56–65. doi: 10.1016/j.tox.2005.05.010

Kurz, E. U., and Lees-Miller, S. P. (2004). DNA damage-induced activation of ATM and ATM-dependent signaling pathways. *DNA Repair* 3, 889–900. doi: 10.1016/j.dnarep.2004.03.029

Luo, W., and Brouwer, C. (2013). Pathview: an R/Bioconductor package for pathway-based data integration and visualization. *Bioinformatics* 29, 1830–1831. doi: 10.1093/bioinformatics/btt285

Luongo, D., Severino, L., Bergamo, P., De Luna, R., Lucisano, A., and Rossi, M. (2006). Interactive effects of fumonisin B1 and alpha-zearalenol on proliferation and cytokine expression in jurkat T cells. *Toxicol. In Vitro* 20, 1403–1410. doi: 10.1016/j.tiv.2006.06.006

Matise, L. A., Palmer, T. D., Ashby, W. J., Nashabi, A., Chytil, A., Aakre, M., et al. (2012). Lack of transforming growth factor-beta signaling promotes collective cancer cell invasion through tumor-stromal crosstalk. *Breast Cancer Res.* 14:R98. doi: 10.1186/bcr3217

Mehmood, Z., Smith, A. G., Tucker, M. J., Chuzel, F., and Carmichael, N. G. (2000). The development of methods for assessing the in vivo oestrogen-like effects of xenobiotics in CD-1 mice. *Food Chem. Toxicol.* 38, 493–501. doi: 10.1016/S0278-6915(00)00022-3

Minervini, F., Giannoccaro, A., Fornelli, F., Dell'Aquila, M. E., Minoia, P., and Visconti, A. (2006). Influence of mycotoxin zearalenone and its derivatives (alpha and beta zearalenol) on apoptosis and proliferation of cultured granulosa cells from equine ovaries. *Reprod. Biol. Endocrinol.* 4:62. doi: 10.1186/1477-7827-4-62

Mueller, S. O., Simon, S., Chae, K., Metzler, M., and Korach, K. S. (2004). Phytoestrogens and their human metabolites show distinct agonistic and antagonistic properties on estrogen receptor alpha (ERalpha) and ERbeta in human cells. *Toxicol. Sci.* 80, 14–25. doi: 10.1093/toxsci/kfh147

Nikov, G. N., Hopkins, N. E., Boue, S., and Alworth, W. L. (2000). Interactions of dietary estrogens with human estrogen receptors and the effect on estrogen receptor-estrogen response element complex formation. *Environ. Health Perspect.* 108, 867–872. doi: 10.1289/ehp.00108867

Nilsson, E. E., and Skinner, M. K. (2002). Role of transforming growth factor beta in ovarian surface epithelium biology and ovarian cancer. *Reprod. Biomed. Online* 5, 254–258. doi: 10.1016/S1472-6483(10)61828-7

Osweiler, G. D., Stahr, H. M., and Beran, G. W. (1990). Relationship of mycotoxins to swine reproductive failure. *J. Vet. Diagn. Invest.* 2, 73–75. doi: 10.1177/104063879000200115

Pan, B., Chao, H., Chen, B., Zhang, L., Li, L., Sun, X., et al. (2011). DNA methylation of germ-cell-specific basic helix-loop-helix (HLH) transcription factors. *Mol. Hum. Reprod.* 17, 550–561. doi: 10.1093/molehr/gar017

Pasche, B. (2001). Role of transforming growth factor beta in cancer. *J. Cell. Physiol.* 186, 153–168. doi: 10.1002/1097-4652(200002)186:2<153::AID-JCP1016>3.0. CO;2-J

Poor, M., Kunsagi-Mate, S., Sali, N., Koszegi, T., Szente, L., and Peles-Lemli, B. (2015). Interactions of zearalenone with native and chemically modified cyclodextrins and their potential utilization. *J. Photochem. Photobiol. B* 151, 63–68. doi: 10.1016/j.jphotobiol.2015.07.009

Qin, X., Cao, M., Lai, F., Yang, F., Ge, W., Zhang, X., et al. (2015). Oxidative stress induced by zearalenone in porcine granulosa cells and its rescue by curcumin in vitro. *PLoS One* 10:e0127551. doi: 10.1371/journal.pone.0127551

Ranzenigo, G., Caloni, F., Cremonesi, F., Aad, P. Y., and Spicer, L. J. (2008). Effects of *Fusarium* mycotoxins on steroid production by porcine granulosa cells. *Anim. Reprod. Sci.* 107, 115–130. doi: 10.1016/j.anireprosci.2007.06.023

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504. doi: 10.1101/gr.1239303

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., et al. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U.S.A.* 102, 15545–15550. doi: 10.1073/pnas.0506580102

Tuxhorn, J. A., McAlhany, S. J., Yang, F., Dang, T. D., and Rowley, D. R. (2002). Inhibition of transforming growth factor-beta activity decreases angiogenesis in a human prostate cancer-reactive stroma xenograft model. *Cancer Res.* 62, 6021–6025.

Vanyi, A., Bata, A., Glavits, R., and Kovacs, F. (1994). Perinatal oestrogen syndrome in swine. *Acta Vet. Hung.* 42, 433–446.

Vijayaraghavan, S., Moulder, S., Keyomarsi, K., and Layman, R. M. (2017). Inhibiting CDK in cancer therapy: current evidence and future directions. *Target. Oncol.* 13, 21–38. doi: 10.1007/s11523-017-0541-2

Weber, A. M., and Ryan, A. J. (2015). ATM and ATR as therapeutic targets in cancer. *Pharmacol. Ther.* 149, 124–138. doi: 10.1016/j.pharmthera.2014.12.001

Yu, G., Wang, L. G., Han, Y., and He, Q. Y. (2012). Clusterprofiler: an R package for comparing biological themes among gene clusters. *OMICS* 16, 284–287. doi: 10.1089/omi.2011.0118

Yu, Z., Hu, D., and Li, Y. (2004). Effects of zearalenone on mRNA expression and activity of cytochrome P450 1A1 and 1B1 in MCF-7 cells. *Ecotoxicol. Environ. Saf.* 58, 187–193. doi: 10.1016/j.ecoenv.2004.03.001

Zhang, G. L., Sun, X. F., Feng, Y. Z., Li, B., Li, Y. P., Yang, F., et al. (2017a). Zearalenone exposure impairs ovarian primordial follicle formation via down-regulation of Lhx8 expression in vitro. *Toxicol. Appl. Pharmacol.* 317, 33–40. doi: 10.1016/j.taap.2017.01.004

Zhang, G. L., Zhang, R. Q., Sun, X. F., Cheng, S. F., Wang, Y. F., Ji, C. L., et al. (2017b). RNA-seq based gene expression analysis of ovarian granulosa cells exposed to zearalenone in vitro: significance to steroidogenesis. *Oncotarget* 8, 64001–64014. doi: 10.18632/oncotarget.19699

Zhang, Z., Chen, Q., Zhang, J., Wang, Y., Hu, X., Yin, S., et al. (2017c). Associations of genetic polymorphisms in PTEN/AKT/mTOR signaling pathway genes with cancer risk: a meta-analysis in Asian population. *Sci. Rep.* 7:17844. doi: 10.1038/s41598-017-17250-z

Zhang, P., Chao, H., Sun, X., Li, L., Shi, Q., and Shen, W. (2010). Murine folliculogenesis in vitro is stage-specifically regulated by insulin via the Akt signaling pathway. *Histochem. Cell Biol.* 134, 75–82. doi: 10.1007/s00418-010-0708-8

Zhu, L., Yuan, H., Guo, C., Lu, Y., Deng, S., Yang, Y., et al. (2012). Zearalenone induces apoptosis and necrosis in porcine granulosa cells via a caspase-3- and caspase-9-dependent mitochondrial signaling pathway. *J. Cell. Physiol.* 227, 1814–1820. doi: 10.1002/jcp.22906

Zwierzchowski, W., Przybylowicz, M., Obremski, K., Zielonka, L., Skorska-Wyszynska, E., Gajecka, M., et al. (2005). Level of zearalenone in blood serum and lesions in ovarian follicles of sexually immature gilts in the course of zearalenone micotoxicosis. *Pol. J. Vet. Sci.* 8, 209–218.

frontiers
in Genetics

Transcriptional Responses Reveal Similarities Between Preclinical Rat Liver Testing Systems

Zhichao Liu[1]*, Brian Delavan[1,2], Ruth Roberts[3,4] and Weida Tong[1]*

[1] Division of Bioinformatics and Biostatistics, National Center for Toxicological Research, U.S. Food and Drug Administration, Jefferson, AR, United States, [2] Department of Biosciences, University of Arkansas at Little Rock, Little Rock, AR, United States, [3] ApconiX, Alderley Edge, United Kingdom, [4] Department of Biosciences, University of Birmingham, Birmingham, United Kingdom

Toxicogenomics (TGx) is an important tool to gain an enhanced understanding of toxicity at the molecular level. Previously, we developed a pair ranking (PRank) method to assess *in vitro* to *in vivo* extrapolation (IVIVE) using toxicogenomic datasets from the Open Toxicogenomics Project-Genomics Assisted Toxicity Evaluation System (TG-GATEs) database. With this method, we investiagted three important questions that were not addressed in our previous study: (1) is a 1-day *in vivo* short-term assay able to replace the 28-day standard and expensive toxicological assay? (2) are some biological processes more conservative across different preclinical testing systems than others? and (3) do these preclinical testing systems have the similar resolution in differentiating drugs by their therapeutic uses? For question 1, a high similarity was noted (PRank score = 0.90), indicating the potential utility of shorter term *in vivo* studies to predict outcome in longer term and more expensive *in vivo* model systems. There was a moderate similarity between rat primary hepatocytes and *in vivo* repeat-dose studies (PRank score = 0.71) but a low similarity (PRank score = 0.56) between rat primary hepatocytes and *in vivo* single dose studies. To address question 2, we limited the analysis to gene sets relevant to specific toxicogenomic pathways and we found that pathways such as lipid metabolism were consistently over-represented in all three assay systems. For question 3, all three preclinical assay systems could distinguish compounds from different therapeutic categories. This suggests that any noted differences in assay systems was biological process-dependent and furthermore that all three systems have utility in assessing drug responses within a certain drug class. In conclusion, this comparison of three commonly used rat TGx systems provides useful information in utility and application of TGx assays.

Keywords: toxicogenomics, preclinical models, liver, gene expression, bioinformatics

INTRODUCTION

Toxicogenomics (TGx) combines toxicology with genomics or other high throughput molecular profiling technologies, offering a powerful method to study the underlying molecular mechanisms of toxicity (Nuwaysir et al., 1999; Aardema and MacGregor, 2002). Since, it was first described some 18 years ago (Nuwaysir et al., 1999), TGx has played an important role in various aspects

of toxicology including mechanistic studies, predictive toxicology and the development of safety biomarkers (Chen et al., 2012).

Toxicogenomic approaches can be broadly categorized into three purposes: predictive toxicology, risk assessment, and mechanistic studies (Suter et al., 2004; Chen et al., 2012). For example, Fielden et al. (2007) developed a short-term (5 day) repeated dose TGx assay in rat to predict non-genotoxic hepatocarcinogenicity with a sensitivity and specificity of 86 and 81%, respectively. Other studies have addressed various questions of applying TGx including optimal treatment duration and sample size for a better predictive performance (Liu et al., 2011; Gusenleitner et al., 2014; Matsumoto et al., 2015; Liu S. et al., 2017). TGx has also been used in semi-quantitative risk assessment such as defining points of departure and benchmark dosing (Yang et al., 2007; AbdulHameed et al., 2016; Chauhan et al., 2016; Dean et al., 2017; Farmahin et al., 2017; Kawamoto et al., 2017). Most widely used application of TGx approaches is to understand the molecular mechanisms of different toxicological endpoints (Ellinger-Ziegelbauer et al., 2008; Blomme et al., 2009; Rodrigues et al., 2016; Hendrickx et al., 2017; Rueda-Zárate et al., 2017). More recently, in addition to gene expression profiling, the study of microRNAs (Wang et al., 2009; Yang et al., 2012; Ward et al., 2014; Liu et al., 2016) and long non-coding RNAs (lncRNAs) (Aigner et al., 2016; Dempsey and Cui, 2017) are emerging as new technologies to be integrated into this field powered by next-generation sequencing technologies (Yu et al., 2014).

In drug development, TGx has been added as an endpoint to existing preclinical study designs to gain more information from these studies. For example, in studies of liver toxicity, preclinical assessment in rodents may use primary rat hepatocytes or may use single dose *in vivo* studies (24 h) or repeat dosing up to 28-days. Each of these test systems may serve a different purpose; *in vitro* studies using primary rat hepatocytes may be used for mechanistic and/or cytotoxicity assessments whereas single and repeat dose toxicity studies are used to determine tolerability and target organ toxicity. The addition of TGx to each of these study types has generated additional data of use in assessment of toxicological risk and mechanisms. Other researchers have compared different testing systems for analysis of such endpoints as identification of biomarkers (Kondo et al., 2009) and gene expression-induced by genotoxic carcinogens (Watanabe et al., 2007). However, a systematic comparison of the value of TGx data generated in the different test systems has not been fully assessed.

Unlike decades ago, there are now several large publicly available toxicogenomic datasets such as the Open Toxicogenomics Project-Genomics Assisted Toxicity Evaluation System (TG-GATEs) database (Uehara et al., 2010; Igarashi et al., 2015), DrugMatrix (Ganter et al., 2005) and PredTox (Suter et al., 2011), providing tremendous opportunities for comparing preclinical testing systems. For example, open TG-GATEs used four standard preclinical study designs to generate TGx data (Ippolito et al., 2015; Bell et al., 2016; Liu et al., 2016; Sutherland et al., 2016). Using TG-GATEs data, we developed a 'topic modeling' approach to explore the underlying relationships between different TGx assay systems (Lee et al., 2014, 2016)

and other toxicological assessments such as high throughput screening assay data from the Tox21 project (Lee et al., 2016).

In our previous study, we developed a Pair Ranking (PRank) method to assess the potential of *in vitro* to *in vivo* extrapolation (IVIVE) among three TGx assay systems (two *in vitro* assays using rat or human hepatocytes and a 28-day repeat-dose rat model) (Liu Z. et al., 2017). The study had an emphasis on assessing the IVIVE potential for different endpoints of drug-induced liver injury (DILI). It was concluded that the *in vitro* assay using primary rat hepatocytes and rat *in vivo* 28-day repeated dose models had high IVIVE potential for most DILI endpoints. However, several important questions remain for prediction of liver responses. Firstly, will a short-term *in vivo* assay (1-day experiment to detect acute response) correlate with a standard long-term *in vivo* repeated dose study (28-day study)? Secondly, are differences and similarities dependent upon biolgocial processes? Finally, can the different TGx assay systems distinguish compounds from different therapeutic categories?

In this study, we analyzed preclinical rat test system data from TG-GATEs comprising 131 compounds in three assays – (1) an *in vitro* study with rat primary hepatocytes (denoted as InVitro hereafter), (2) a rat *in vivo* single-dose treatment wih sample collection after 24 h (denoted as InVivo_S), and (3) a rat 28 day repeat-dose study (denoted as InVivo_R hereafter). Comparative analysis among these three assay systems were analyzed using PRank. Additonal useful comparisons were genererated by limiting the analyses firstly to compounds from certain therapeutic categories and secondly to gene sets representing specific toxicogenomic pathways.

MATERIALS AND METHODS

Toxicogenomics Datasets

The open TG-GATEs[1] was employed to investigate preclinical TGx assay systems in rats (Igarashi et al., 2015). Three rat toxicogenomic data sets from the TG-GATEs Phase I study were included covering 131 compounds from different therapeutic categories. The rat *in vitro* data had three concentrations (low, medium, and high) and three treatment durations (2, 4, and 24 h). The rat *in vivo* single dose also used three doses (low, medium, and high) and the samples were collected at four different timepoints after treatment (3, 6, 9, and 24 h). The *in vivo* repeated dose data was generated under the standard *in vivo* experiment design with three doses (low, medium, and high) and four treatment durations (3, 7, 14, and 28 days), where the rat liver tissue was isolated 24 h after treatment. In this study, we focused on the highest concentration/dose and longest treatment duration of 120 common compounds among the three assay systems for each assay system (the data used are available from **Supplementary Table S1**). Specifically, (1) "InVitro" is the data from *in vitro* assay with rat primary hepatocytes treated with the highest dose and the sample is collected 24 h after treatment, (2) "InVivo_S" is the data from rat *in vivo* single high dose and the sample is collected 24 h after treatment, and (3) "InVivo_R" is

---

[1]http://toxico.nibiohn.go.jp/english/

repeated dose daily with highest dose for 28 days. More details on concentration and dose definition are listed in our previous study (Liu Z. et al., 2017) and elsewhere (Igarashi et al., 2015).

Microarray Data Normalization and Differentially Expressed Genes (DEGs) Calculation

The microarray data from three rat TGx systems was processing using Factor Analysis for Robust Microarray Summarization (FARMS) (Hochreiter et al., 2006) with a custom CDF file from BRIANARRAY[2]. The details were as described previously (Hochreiter et al., 2006; Liu Z. et al., 2017). Replicate measurements were collapsed to one measurement per gene. The collapsed data can be downloaded from http://dokuwiki.bioinf.jku.at/doku.php/tgp_prepro. The downloaded data were transformed as MAT File Format as an input for further analysis. For each compound in each assay system, the fold change values were generated by comparing the treatment group vs. matched control group for each time and concentration/dose condition.

Therapeutic Categories

The Anatomical Therapeutic Chemical (ATC) classification system was used to group the compounds into different therapeutic classes. The ATC classification system has five levels of code to characterize a chemical/drug based on (1) the system/organ it acts on, (2) its therapeutic use, (3) its pharmacological functions, (4) its chemical properties, and (5) the chemical itself. In this study, the second-level of ATC codes indicating the main therapeutic group were used (see **Supplementary Table S1**).

Toxicity Pathways Related Gene Sets

The gene sets related to different toxicity pathways were extracted from the Comparative Toxicogenomics Database (CTD) (Davis et al., 2017), which aims to illustrate how environmental chemicals affect human health. Specifically, the gene and pathway relationship data were downloaded from http://ctdbase.org/downloads/. There are a total of 135,815 gene and pathway relationships. Due to the gene symbols (Entrez Gene IDs) in CTD database was based on homo sapiens, we mapped Entrez gene IDs from homo sapiens to Rattus norvegicus based on NCBI HomoloGene build 68[3]. We clustered the genes based on their related pathways and kept the pathways containing more than 200 genes for further analysis (see **Supplementary Table S2**).

Pair Ranking Method (PRank)

The Pair Ranking (PRank) method was used to compare the three rat TGx assay systems (Liu Z. et al., 2017). First, the pairwise compound similarity of any two compounds within an assay system was calculated using their biologically significant genes which were the top and down 200 ranked genes by their fold change values. The total number of 400 genes as

the compound signatures were used for similarity calculation. The Dice's coefficient was employed to measure the similarity between the transcriptional profiles of compounds, which were suggested by SEQC I study (Wang et al., 2014). In this study, the overlapped genes were counted by taking into consideration of their regulation direction and the Dice's coefficient were calculated by using the following equation,

$$Dice's\ coefficient = \frac{2(N_{i,j,up} + N_{i,j,down})}{400 + 400} = \frac{N_{i,j,up} + N_{i,j,down}}{400} \quad (1)$$

where, $N_{i,j,up}$ and $N_{i,j,down}$ denote the number of overlapped the up/down regulated genes between compound i and compound j, respectively. Then, each pair was ranked from high to low by the pairwise similarity. Lastly, the PRank score was calculated between any two assay systems by using receiver operating characteristic (ROC) curve and the area under the curve (AUC). For ROC-AUC analysis, we need to transform the ranked Dice's coefficient to binary values (positive and negative: 0/1). In this study, the Dice's coefficient value more than 0.4 was selected as cut-off, which is close to 95% quantile. The ROC-AUC analysis was conducted by using function *perfcurve.m* from MATLAB R2016a.

To investigate whether the compounds within a therapeutic category were more similar than across therapeutic categories, we used the following formula,

$$stability\ ratio = \frac{mean\ (\sum_{i=1}^{n} Dice\_inter)}{mean\ (\sum_{i=1}^{n} Dice\_across)} \quad (2)$$

where, n is the number of compound pairs. For inter therapeutic category, the compound pairwise similarity was generated by calculating the Dice's coefficients between any two compounds from the same category. For across therapeutic categories, the pairwise similarity was generated between compounds from the different therapeutic categories. Finally, we calculated the stability ratio between inter therapeutic and across therapeutic categories to investigate whether the assay system could distinguish one therapeutic category to another. If the stability is more than 1, it means that the similarity among compounds for inter therapeutic category is more than across therapeutic category, indicating the similarity based on toxicogenomic profiles is capable of distinguishing the compounds from one therapeutic category to another.

For compound pairwise similarity calculations using the gene sets from different toxicogenomic pathways, we followed the following procedures. First, we mapped each gene set derived from toxicogenomic pathways to rat genes represented by the microarrays used in open TG-GATEs. Then, we retained the overlapped genes with fold change more than 1.5 for each compound as individual signatures. Finally, we calculated Dice coefficients between any two compounds based on the generated signatures in each system.

Percentage of Overlapping Pathways (POP)

The concordance among the three assay systems were also assessed in the different KEGG pathways. Specifically, the 400

genes for each compound in each assay system was input to the Database for Annotation, Visualization and Integrated Discovery (DAVID) software to carry out KEGG pathway analysis (Huang et al., 2008). The pathways with a Benjamini–Hochberg adjusted *p*-value less than 0.05 were considered as statistically significant pathways. Then, the enriched KEGG pathways in each assay systems were ranked based on frequency of pathways perturbed by the compounds ($p \leq 0.05$). Finally, the POP represented the number of common pathways between any two assay systems divided by L, the number of pathways in each of subset of ranked pathway list. In this study, L was set from 5 to 60.

Chemical Structure Similarity

The chemical structure of 120 common compounds could be found from our previous study (Liu Z. et al., 2017). The Pipeline Pilot 8.0 (Accelrys, Biovia, and Dassault Systems) was used to calculate the compound pairwise similarity based on their functional class fingerprints (FCFPs) with a radius of FCFP-4. The compound pairwise chemical similarities were listed in **Supplementary Table S1**.
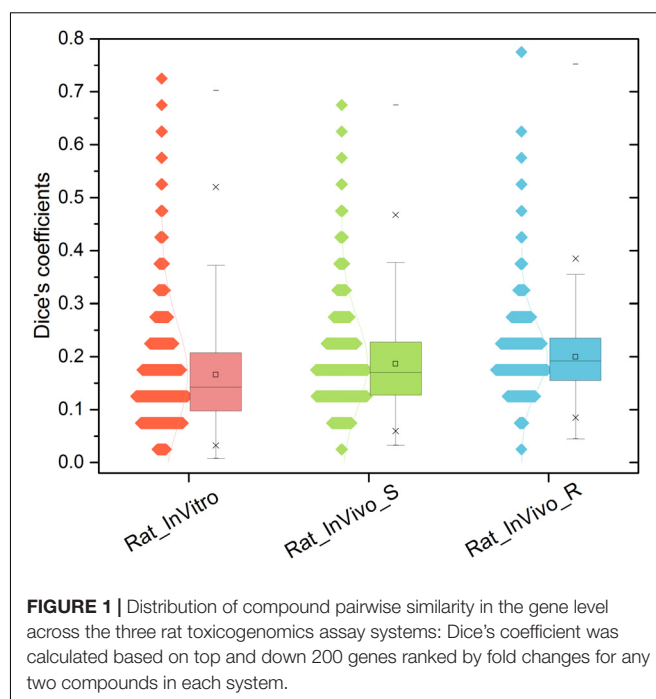
Code availability

The scripts and processed data in this study were available in https://github.com/iguana128/Frontier-source-codes.

## RESULTS

Detection Power of Three TGx Assay Systems

We first examined each assay's ability to differentiate drug pairs. **Figure 1** illustrates the pairwise similarity distribution for the three TGx assay systems. The average Dice's coefficients in the three assay systems were ranked as InVivo_R (Dice's coefficient = 0.200) > InVivo_S (Dice's coefficient = 0.187) > InVitro (Dice's coefficient = 0.166) (see **Supplementary Table S3**). The low Dice's coefficients indicated that all three TGx assay systems could differentiate one compound pair to another, where the InVivo_R assay seems to be less sensitive compared to other two assays.

Read-across have been widely applied to risk assessment based on chemical structure similarity (Vink et al., 2010; Rand-Weaver et al., 2013). Recently, the read-across concept has been expanded to integrate biological data profiles such as TGx and cell-based *in vitro* assays (Zhu et al., 2016). Here, the drug pairs in each assay system were compared with the compound pairwise chemical similarity (Dice coefficients > 0.2). It was illustrated that the correlation between assay systems and chemical space was low with the Pearson's correlation coefficients of 0.30, 0.20, and 0.21 for chemical space vs. InVitro, InVivo_S, and InVivo_R, respectively (**Supplementary Figure S1**). The difference between the chemical space and toxicogenomic space suggested that the read-cross can be improved by combining the information from both chemistry and toxicogenomics spaces.

FIGURE 1 | Distribution of compound pairwise similarity in the gene level across the three rat toxicogenomics assay systems: Dice's coefficient was calculated based on top and down 200 genes ranked by fold changes for any two compounds in each system.

Therapeutic Class Analysis

We further investigated whether the three TGx assay system could be utilized to discriminate different therapeutic categories. There was a total of 12 therapeutic categories with at least five compounds (**N02**-Analgesics; **M02**-Topical products for joint and muscular pain; **A10**-Drugs used in diabetes; **C10**-Lipid modifying agents; **N03**-Antiepileptics; **L01**-Antineoplastic agents; **M01**-Antiinflammatory and antirheumatic products; **C01**-Cardiac therapy; **N05**-Psycholeptics; **N06**-Psychoanaleptics; **J01**-Antibacterials for systemic use; **S01**-Ophthalmological*s*) (**Supplementary Figure S2**). For each therapeutic category and each assay system, the stability ratios were calculated by comparing the mean value between and across categories. Almost all the therapeutic categories in each assay system had a stability ratio of more than 1 (**Figure 2**), suggested that the assay systems could distinguish the different therapeutic categories from each other. Among 12 therapeutic categories, the high stability ratios of **C10**-Lipid modifying agents was observed in all three assay systems, indicating the lipid modifying agents including clofibrate, fenofibrate, gemfibrozil, nicotinic acid, simvastatin could be distinguished from compounds in other therapeutic categories in TGx assay systems. Furthermore, **J01**-Antibacterials for systemic use and **S01**-Ophthalmological*s* with stability ratios less than 1 in all three assay systems, showing the lower discrimination power of TGx assay systems for compounds from these two therapeutic categories. It could be seen that the high proportion of compounds were overlapped between the some therapeutic categories (**J01**-Antibacterials for systemic use and **S01**-Ophthalmologicals, and **M01**-Antiinflammatory and antirheumatic products and **M02**-Topical products for joint and muscular pain) due to the multiple therapeutic uses of these compounds, indicating the complexity of off-target space of
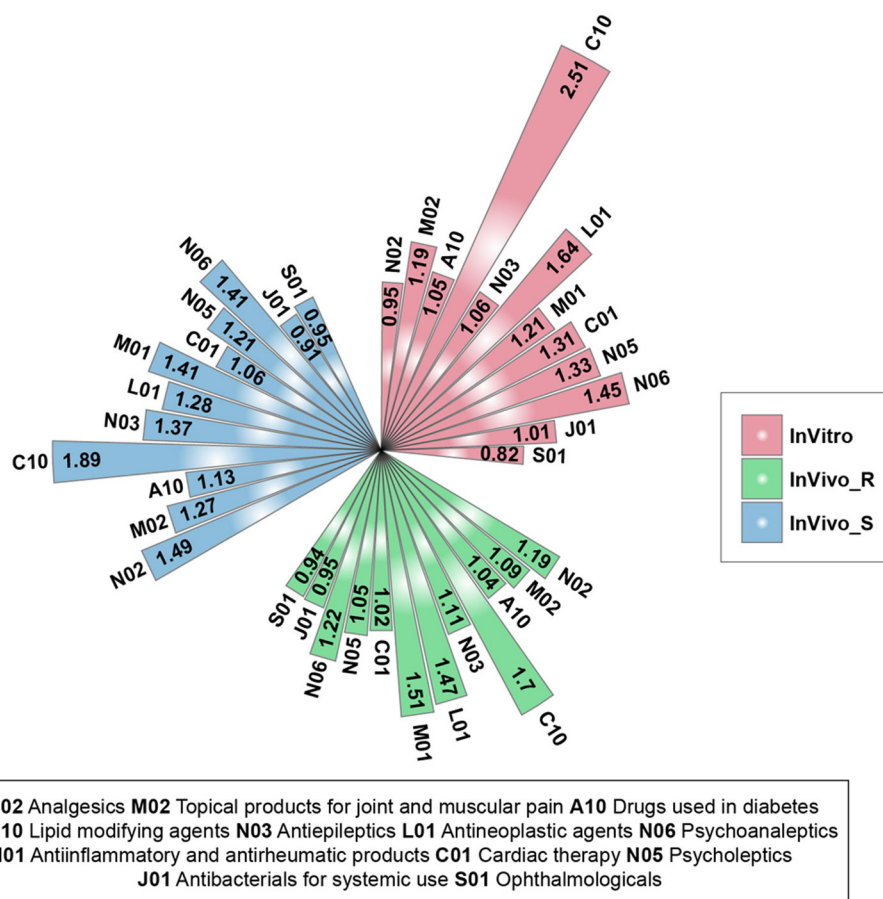
FIGURE 2 | Stability ratios for the 12 therapeutic categories in each assay system: for each assay system, the stability ratio was calculated based on the average of Dice's coefficient of inter and across therapeutic categories.

these compounds, which may partially explain the unsatisfactory discrimination power.

Concordance Among Three TGx Assay Systems

Figure 3A shows the concordance among three assay systems (InVitro, InVivo_S, and InVivo_R) based on the PRank scores. The highest concordance was noted for the InVivo_S (24 h) and InVivo_R (28-day) with a PRank score 0.90, suggesting the potential to replace long-term treatments with a 1-day experiment using a single dose treatment without loss of prediction. As reported in our previous study, the InVitro and InVivo_R also had a relatively high PRank score (0.71), suggesting a good IVIVE potential (Liu Z. et al., 2017). However, the concordance between InVitro and InVivo_S (PRank score = 0.56) was lower despite the same treatment duration in these two assay systems.

The concordance among the three assay systems was compared at the pathway level. Specifically, the percentage of overlapped pathways (POP) was calculated based on shared over-represented KEGG pathways (Fisher's exact test with adjusted p-value < 0.05) between any two assay systems. As illustrated in **Figure 3B**, the highest concordance was for the two *in vivo* systems (POP value = 0.875), followed by InVitro-InVivo_R (POP value = 0.750) and InVitro-InVivo_S (POP value = 0.563). Therefore, a similar pattern was found at both the gene and pathway level. Furthermore, pathways related to lipid metabolism such as *steroid hormone biosynthesis* and *fatty acid metabolism* were consistently over-represented in the three assay systems (**Table 1**).

Toxicity Pathway Analysis

We further investigated the concordance among the three assay systems when limiting the genes to specific toxicity pathways. The >135K gene-pathway relationships in CTD were employed, and a total of 106 toxicity pathways related genes sets with at least 200 genes for each were extracted (see **Supplementary Table S2**). **Figure 4A** depicts the concordance among the three assay systems in different toxicity pathway. The concordance among the three assay systems in the gene sets level was consistent with the finding in the whole gene/pathway level with a concordance ranking as InVivo_S-InVivo_R > InVitro-InVivo_R > InVitro-InVivo_S. We furthermore compared the top 15 common gene sets related pathways in the three comparisons based on the

FIGURE 3 | The concordance among the three rat toxicogenomics assay systems: **(A)** PRank methodology based on the top and down 200 genes based on fold change values. **(B)** The percentage of overlapping KEGG pathways based on over-represented KEGG pathways using Fisher's exact test with adjusted *p*-value less than 0.05.

TABLE 1 | The overlapping KEGG pathways among the three assay systems.

KEGG entry	Pathways names	Categories
rno00140	Steroid hormone biosynthesis	Lipid metabolism
rno00071	Fatty acid metabolism	Lipid metabolism
rno00330	Arginine and proline metabolism	Amino acid metabolism
rno00280	Valine, leucine, and isoleucine degradation	Amino acid metabolism
rno00480	Glutathione metabolism	Metabolism of other amino acids
rno00982	Drug metabolism	Xenobiotics biodegradation and metabolism
rno00980	Metabolism of xenobiotics by cytochrome P450	Xenobiotics biodegradation and metabolism
rno00830	Retinol metabolism	Metabolism of cofactors and vitamins
rno03320	PPAR signaling pathway	Endocrine system

FIGURE 4 | The concordance among the three rat toxicogenomics assay systems for gene sets related to different toxicogenomic pathways: **(A)** the stacked plots of PRank scores for different gene sets in the three assay systems; **(B)** a Venn diagram of the top 15 gene set-related toxicogenomic pathways based on the PRank score ranking in each rat assay systems.

PRank scores (**Figure 4B**). We found that two lipid related pathways, i.e., *Metabolism of lipids and lipoproteins* and *Fatty acid, triacylglycerol, and ketone body metabolism* were common in all three comparisons, which is also consistent with the finding in the whole gene/pathway level. Furthermore, the similar conclusion was also drawn based on the stability analysis, which suggested the *lipid modifying agents* (C10) were highly discriminated in all three TGx testing systems.

Confirmation Based on Multiple Time and Dose Points

The multiple time and dose combination design of TG-GATEs data sets provides a great opportunity to fully evaluate the pharmacokinetic and pharmacodynamic characteristics of chemical-induced toxicity and further facilitate early predictive biomarkers development for toxicity prediction and prevention. In the main part of this study, we comprehensively investigated the concordance among the three rat TGx assay systems at high

dose and longest duration condition. Moreover, we expanded the comparisons to the different time and dose combinations. **Figure 5** shows the concordance among three assay systems at different time and dose conditions based on proposed
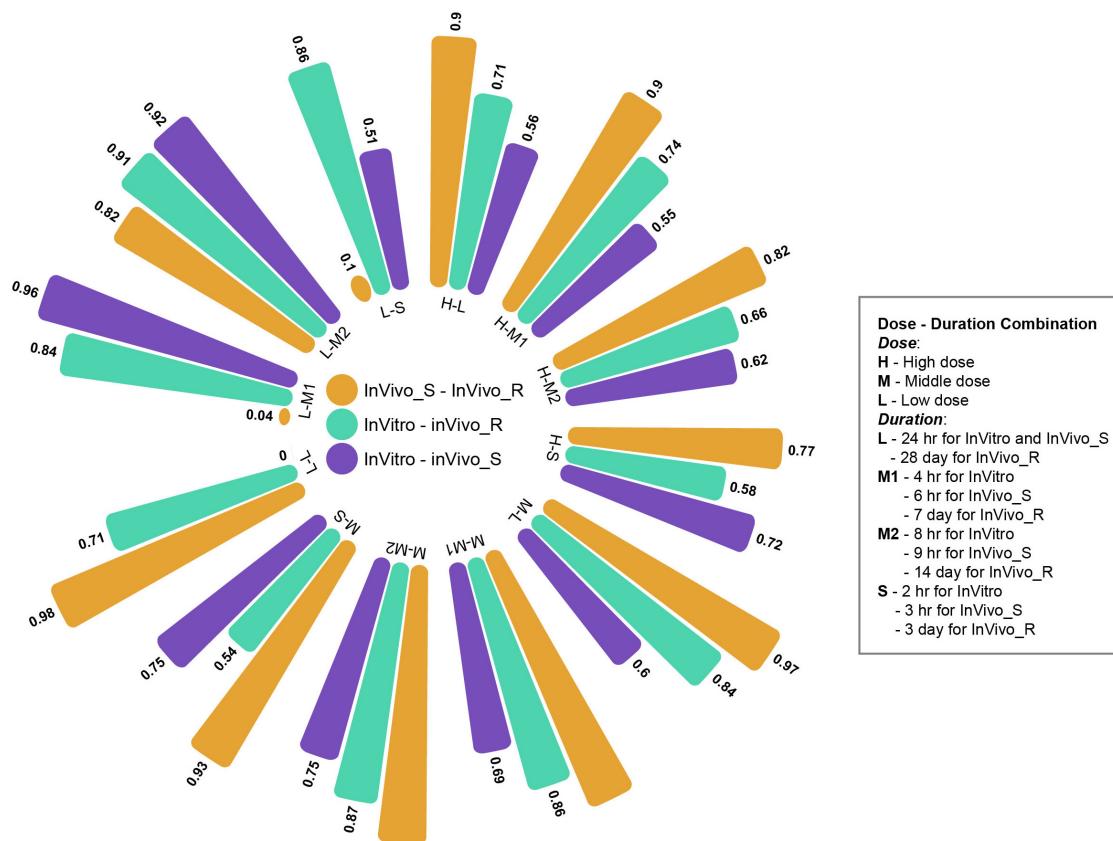
FIGURE 5 | Circle bar plot for the concordance among three rat toxicogenomics systems at different time and dose combinations: the concordance between assay systems were conducted under a total of 12 time/dose combination by using Prank strategy.

PRank method. The circle bar plot represented the PRank scores. The similar trends of PRank scores changes (InVivo_S-InVivo_R > InVitro-InVivo_R > InVitro-InVivo_S) could be observed in high and medium dose with long and middle treatment duration. However, the low dose and short treatment durations were not able to provide enough discrimination power to assess the concordance among three TGx testing systems. Furthermore, the N-way ANOVA analysis were used to estimate the resource of variances contributing to the concordance among the TGx testing assays by using MATLAB function *anovan.m*. It was indicated that dose was more dominated influential factor of the concordance between testing systems than treatment duration (see **Supplementary Table S4**).

DISCUSSION

Animal models are indispensable in drug development and risk assessment, although extrapolation from animal models to human responses remains a challenge (Shanks et al., 2009). A key focus of research into animal models is how they could better recapitulate the human toxicological and physiological environment and provide a more reliable and robust prediction of human toxicity. Cell-based *in vitro* assays and *in silico* approaches

have been proposed that could refine, reduce or even replace animal models (Hamburg, 2011; Goodman et al., 2015). In support of this, it is key to gain a better understanding on the similarities and differences between data generated in cell-based assay (*in vitro*) systems and animal (*in vivo*) models. Previously, we assessed similarities in TGx data between rat and human primary hepatocyte cultures and rat liver after 28 days of repeated dosing for a number of drugs and chemicals (Liu Z. et al., 2017). Here, we carried out a comparative analysis among three frequently-used rat TGx assay systems (InVitro, InVivo_S, and InVivo_R) using our previously described Pair Ranking (PRank) methodology.

The data indicated that there was a high concordance between the two *in vivo* assay systems (24 h and 28 days), indicating a potential to use a short-term *in vivo* assay for some endpoints saving time and money. Furthermore, the *in vitro* TGx data set had a relatively high similarity to the standard 28-day *in vivo* repeated dose experiment data, suggesting a good correlation of *in vitro* with longer term treatment *in vivo*. However, there was a poor concordance between *in vitro* and the *in vivo* single dose (24 h) treatment. This observation is at first surprising but one explanation could be that extraction of hepatocytes into cell culture followed by 24 h of treatment represents a level of chemical/environmental stress more equivalent to 28 days of

in vivo treatment compared with 24 h (single dose) *in vivo* where the liver may only just be responding to a new chemical stress. Specifically, gene activities associated with the survival cells of the hepatocytes reflect a level of the adaptation that resemble to these in the 28-day repeated dosing conditions.

All three TGx assay systems could distinguish compounds by therapeutic category. Among the 12 investigated therapeutic categories, the **C10**-*Lipid modifying agents* with highest stability ratios in all the three assay systems, indicating the high discrimination power. It is very interesting that, when the analyses were focused on specific pathways, several pathways such as lipid metabolism-related pathways were consistently over-represented in all three assay systems, the finding is consistent with the therapeutic categories, suggesting that similarity between the systems is to some extent dependent on different biological process and compounds under different therapeutic categories.

It is worthwhile to consider some additional studies to further our knowledge and confirm the findings from this study. Firstly, the current comparisons among the three TGx assay systems were based on the perturbation of gene expression within each of these assay systems. Although this could be the case, there is no certainty that these conclusions are applicable to other assay systems where there may be differences in intrinsic properties such as species or tissue type and extrinsic properties such as time of exposure and culture conditions. Therefore, we proposed more retrospective analyses of preclinical TGx data sets should be undertaken to provide a boarder and more comprehensive picture of how animal models and cell-based *in vitro* assay systems can be translated to predict human responses. Secondly, in this study we employed TG-GATEs datasets, currently the largest dataset in the TGx research arena. Despite this, there are still many classes of chemicals and drugs missing. Therefore, more comprehensive and larger scale TGx datasets could yield more robust conclusion. Thirdly, in the current study, transcriptomic profiles (gene expression) data were used. With the advance of technology, other data such as microRNAs profiles should be investigated since these may be considered more conserved in different species and organ systems (Mack, 2007). Finally, in the current study we focused on the top 400 differentially expressed genes (DEGs) to reveal the relationship between testing systems. In our previous study, we have discussed the influence of the number of selected genes to the similarity measure and concluded that the selected 400 genes could generate the stable similarity

ranking list in each assay system (Liu Z. et al., 2017). With that said, other methods and/or different lengths of DEGs applied should be consider to enhance the comprehensiveness of the investigation.

Toxicogenomics has been used widely to supplement risk assessment data, to elucidate underlying mechanisms of toxicology and to support predictive toxicology. One of the contentious questions in the toxicology field is whether animal models can provide sufficient predictive power for human toxicity. In this study, we investigated concordance among TGx data from three rat assay systems using a Pairwise Ranking strategy. The data generated provide an insight into the utility of these assay systems for drug safety evaluation and risk assessment.

AUTHOR CONTRIBUTIONS

Conceived and designed the experiments: ZL and WT. Analyzed the data and first version of the manuscript: ZL. Contributed reagents/materials/analysis tools: ZL and BD. Wrote the manuscript: ZL, RR, BD, and WT. All authors read and approved the final manuscript.

SUPPLEMENTARY MATERIAL

FIGURE S1 | Concordance of drug pairs between chemical space and assay systems: significant drug pairs in each assay system (Dice's coefficient > 0.2) were used to investigate the concordance with chemical space by using Pearson's correlation coefficients.

FIGURE S2 | Twelve therapeutic categories with at least five compounds based on the second-level of Anatomical Therapeutic Chemical (ATC) Classification System system.

TABLE S1 | Information of 120 common drugs among three rat toxicogenomics systems.

TABLE S2 | 106 toxicity pathways related genes sets with at least 200 genes from Comparative Toxicogenomics Database (CTD) database.

TABLE S3 | Pairwise Dice's coefficients in different time/dose/assay systems.

TABLE S4 | Impact of time and dose for correction between assay systems base on N-way ANOVA analysis.

REFERENCES

Aardema, M. J., and MacGregor, J. T. (2002). Toxicology and genetic toxicology in the new era of "toxicogenomics": impact of "-omics" technologies. *Mutat. Res.* 499, 13–25. doi: 10.1016/s0027-5107(01)00292-5

AbdulHameed, M. D. M., Ippolito, D. L., Stallings, J. D., and Wallqvist, A. (2016). Mining kidney toxicogenomic data by using gene co-expression modules. *BMC Genomics* 17:790. doi: 10.1186/s12864-016-3143-y

Aigner, A., Buesen, R., Gant, T., Gooderham, N., Greim, H., Hackermüller, J., et al. (2016). Advancing the use of noncoding RNA in regulatory toxicology:

Report of an ECETOC workshop. *Regul. Toxicol. Pharmacol.* 82, 127–139. doi: 10.1016/j.yrtph.2016.09.018

Bell, S. M., Angrish, M. M., Wood, C. E., and Edwards, S. W. (2016). Integrating publicly available data to generate computationally predicted adverse outcome pathways for fatty liver. *Toxicol. Sci.* 150, 510–520. doi: 10.1093/toxsci/kfw017

Blomme, E. A., Yang, Y., and Waring, J. F. (2009). Use of toxicogenomics to understand mechanisms of drug-induced hepatotoxicity during drug discovery and development. *Toxicol. Lett.* 186, 22–31. doi: 10.1016/j.toxlet.2008.09.017

Chauhan, V., Kuo, B., McNamee, J. P., Wilkins, R. C., and Yauk, C. L. (2016). Transcriptional benchmark dose modeling: exploring how advances in

chemical risk assessment may be applied to the radiation field. *Environ. Mol. Mutagen.* 57, 589–604. doi: 10.1002/em.22043

Chen, M., Zhang, M., Borlak, J., and Tong, W. (2012). A decade of toxicogenomic research and its contribution to toxicological science. *Toxicol. Sci.* 130, 217–228. doi: 10.1093/toxsci/kfs223

Davis, A. P., Grondin, C. J., Johnson, R. J., Sciaky, D., King, B. L., McMorran, R., et al. (2017). The comparative toxicogenomics database: update 2017. *Nucleic Acids Res.* 45, D972–D978. doi: 10.1093/nar/gkw838

Dean, J. L., Zhao, Q. J., Lambert, J. C., Hawkins, B. S., Thomas, R. S., and Wesselkamper, S. C. (2017). Editor's highlight: application of gene set enrichment analysis for identification of chemically induced, biologically relevant transcriptomic networks and potential utilization in human health risk assessment. *Toxicol. Sci.* 157, 85–99. doi: 10.1093/toxsci/kfx021

Dempsey, J. L., and Cui, J. Y. (2017). Long non-coding RNAs: a novel paradigm for toxicology. *Toxicol. Sci.* 155, 3–21. doi: 10.1093/toxsci/kfw203

Ellinger-Ziegelbauer, H., Gmuender, H., Bandenburg, A., and Ahr, H. J. (2008). Prediction of a carcinogenic potential of rat hepatocarcinogens using toxicogenomics analysis of short-term in vivo studies. *Mutat. Res.* 637, 23–39. doi: 10.1016/j.mrfmmm.2007.06.010

Farmahin, R., Williams, A., Kuo, B., Chepelev, N. L., Thomas, R. S., Barton-Maclaren, T. S., et al. (2017). Recommended approaches in the application of toxicogenomics to derive points of departure for chemical risk assessment. *Arch. Toxicol.* 91, 2045–2065. doi: 10.1007/s00204-016-1886-5

Fielden, M. R., Brennan, R., and Gollub, J. (2007). A gene expression biomarker provides early prediction and mechanistic assessment of hepatic tumor induction by nongenotoxic chemicals. *Toxicol. Sci.* 99, 90–100. doi: 10.1093/toxsci/kfm156

Ganter, B., Tugendreich, S., Pearson, C. I., Ayanoglu, E., Baumhueter, S., Bostian, K. A., et al. (2005). Development of a large-scale chemogenomics database to improve drug candidate selection and to understand mechanisms of chemical toxicity and action. *J. Biotechnol.* 119, 219–244. doi: 10.1016/j.jbiotec.2005.03.022

Goodman, J., Chandna, A., and Roe, K. (2015). Trends in animal use at US research facilities. *J. Med. Ethics* 41, 567–569. doi: 10.1136/medethics-2014-102404

Gusenleitner, D., Auerbach, S. S., Melia, T., Gómez, H. F., Sherr, D. H., and Monti, S. (2014). Genomic models of short-term exposure accurately predict long-term chemical carcinogenicity and identify putative mechanisms of action. *PLoS One* 9:e102579. doi: 10.1371/journal.pone.0102579

Hamburg, M. A. (2011). Advancing Regulatory Science. *Science* 331, 987–987. doi: 10.1126/science.1204432

Hendrickx, D. M., Souza, T., Jennen, D. G. J., and Kleinjans, J. C. S. (2017). DTNI: a novel toxicogenomics data analysis tool for identifying the molecular mechanisms underlying the adverse effects of toxic compounds. *Arch. Toxicol.* 91, 2343–2352. doi: 10.1007/s00204-016-1922-5

Hochreiter, S., Clevert, D.-A., and Obermayer, K. (2006). A new summarization method for affymetrix probe level data. *Bioinformatics* 22, 943–949. doi: 10.1093/bioinformatics/btl033

Huang, D. W., Sherman, B. T., and Lempicki, R. A. (2008). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* 4, 44–57. doi: 10.1038/nprot.2008.211

Igarashi, Y., Nakatsu, N., Yamashita, T., Ono, A., Ohno, Y., Urushidani, T., et al. (2015). Open TG-GATEs: a large-scale toxicogenomics database. *Nucleic Acids Res.* 43, D921–D927. doi: 10.1093/nar/gku955

Ippolito, D. L., AbdulHameed, M. D., Tawa, G. J., Baer, C. E., Permenter, M. G., McDyre, B. C., et al. (2015). Gene expression patterns associated with histopathology in toxic liver fibrosis. *Toxicol. Sci.* 149, 67–88. doi: 10.1093/toxsci/kfv214

Kawamoto, T., Ito, Y., Morita, O., and Honda, H. (2017). Mechanism-based risk assessment strategy for drug-induced cholestasis using the transcriptional benchmark dose derived by toxicogenomics. *J. Toxicol. Sci.* 42, 427–436. doi: 10.2131/jts.42.427

Kondo, C., Minowa, Y., Uehara, T., Okuno, Y., Nakatsu, N., Ono, A., et al. (2009). Identification of genomic biomarkers for concurrent diagnosis of drug-induced renal tubular injury using a large-scale toxicogenomics database. *Toxicology* 265, 15–26. doi: 10.1016/j.tox.2009.09.003

Lee, M., Liu, Z., Huang, R., and Tong, W. (2016). Application of dynamic topic models to toxicogenomics data. *BMC Bioinformatics* 17:368. doi: 10.1186/s12859-016-1225-0

Lee, M., Liu, Z., Kelly, R., and Tong, W. (2014). Of text and gene–using text mining methods to uncover hidden knowledge in toxicogenomics. *BMC Syst. Biol.* 8:93. doi: 10.1186/s12918-014-0093-3

Liu, S., Kawamoto, T., Morita, O., Yoshinari, K., and Honda, H. (2017). Discriminating between adaptive and carcinogenic liver hypertrophy in rat studies using logistic ridge regression analysis of toxicogenomic data: the mode of action and predictive models. *Toxicol. Appl. Pharmacol.* 318, 79–87. doi: 10.1016/j.taap.2017.01.006

Liu, Z., Fang, H., Borlak, J., Roberts, R., and Tong, W. (2017). In vitro to in vivo extrapolation for drug-induced liver injury using a pair ranking method. *ALTEX* 34, 399–407. doi: 10.14573/altex.1610201

Liu, Z., Kelly, R., Fang, H., Ding, D., and Tong, W. (2011). Comparative analysis of predictive models for nongenotoxic hepatocarcinogenicity using both toxicogenomics and quantitative structure–activity relationships. *Chem. Res. Toxicol.* 24, 1062–1070. doi: 10.1021/tx2000637

Liu, Z., Wang, Y., Borlak, J., and Tong, W. (2016). Mechanistically linked serum miRNAs distinguish between drug induced and fatty liver disease of different grades. *Sci. Rep.* 6:23709. doi: 10.1038/srep23709

Mack, G. S. (2007). MicroRNA gets down to business. *Nat. Biotechnol.* 25, 631–638. doi: 10.1038/nbt0607-631

Matsumoto, H., Saito, F., and Takeyoshi, M. (2015). Applicability of a gene expression based prediction method to SD and Wistar rats: an example of CARCINOscreen®. *J. Toxicol. Sci.* 40, 805–807. doi: 10.2131/jts.40.805

Nuwaysir, E. F., Bittner, M., Trent, J., Barrett, J. C., and Afshari, C. A. (1999). Microarrays and toxicology: the advent of toxicogenomics. *Mol. Carcinog.* 24, 153–159. doi: 10.1002/(sici)1098-2744(199903)24:3<153::aid-mc1<3.0.co;2-p

Rand-Weaver, M., Margiotta-Casaluci, L., Patel, A., Panter, G. H., Owen, S. F., and Sumpter, J. P. (2013). The read-across hypothesis and environmental risk assessment of pharmaceuticals. *Environ. Sci. Technol.* 47, 11384–11395. doi: 10.1021/es402065a

Rodrigues, R. M., Heymans, A., De Boe, V., Sachinidis, A., Chaudhari, U., Govaere, O., et al. (2016). Toxicogenomics-based prediction of acetaminophen-induced liver injury using human hepatic cell systems. *Toxicol. Lett.* 240, 50–59. doi: 10.1016/j.toxlet.2015.10.014

Rueda-Zárate, H. A., Imaz-Rosshandler, I., Cárdenas-Ovando, R. A., Castillo-Fernández, J. E., Noguez-Monroy, J., and Rangel-Escareño, C. (2017). A computational toxicogenomics approach identifies a list of highly hepatotoxic compounds from a large microarray database. *PLoS One* 12:e0176284. doi: 10.1371/journal.pone.0176284

Shanks, N., Greek, R., and Greek, J. (2009). Are animal models predictive for humans? *Philos. Ethics Humanit. Med.* 4:2. doi: 10.1186/1747-5341-4-2

Suter, L., Babiss, L. E., and Wheeldon, E. B. (2004). Toxicogenomics in predictive toxicology in drug development. *Chem. Biol.* 11, 161–171. doi: 10.1016/j.chembiol.2004.02.003

Suter, L., Schroeder, S., Meyer, K., Gautier, J. C., Amberg, A., Wendt, M., et al. (2011). Eu framework 6 project: predictive toxicology (PredTox)-overview and outcome. *Toxicol. Appl. Pharmacol.* 252, 73–84. doi: 10.1016/j.taap.2010.10.008

Sutherland, J. J., Jolly, R. A., Goldstein, K. M., and Stevens, J. L. (2016). Assessing concordance of drug-induced transcriptional response in rodent liver and cultured hepatocytes. *PLoS Comput. Biol.* 12:e1004847. doi: 10.1371/journal.pcbi.1004847

Uehara, T., Ono, A., Maruyama, T., Kato, I., Yamada, H., Ohno, Y., et al. (2010). The Japanese toxicogenomics project: application of toxicogenomics. *Mol. Nutr. Food Res.* 54, 218–227. doi: 10.1002/mnfr.200900169

Vink, S. R., Mikkers, J., Bouwman, T., Marquart, H., and Kroese, E. D. (2010). Use of read-across and tiered exposure assessment in risk assessment under REACH – A case study on a phase-in substance. *Regul. Toxicol. Pharmacol.* 58, 64–71. doi: 10.1016/j.yrtph.2010.04.004

Wang, C., Gong, B., Bushel, P. R., Thierry-Mieg, J., Thierry-Mieg, D., Xu, J., et al. (2014). The concordance between RNA-seq and microarray data depends on chemical treatment and transcript abundance. *Nat. Biotechnol.* 32, 926–932. doi: 10.1038/nbt.3001

Wang, K., Zhang, S., Marzolf, B., Troisch, P., Brightman, A., Hu, Z., et al. (2009). Circulating microRNAs, potential biomarkers for drug-induced liver injury. *Proc. Natl. Acad. Sci. U.S.A.* 106, 4402–4407. doi: 10.1073/pnas.0813371106

Ward, J., Kanchagar, C., Veksler-Lublinsky, I., Lee, R. C., McGill, M. R., Jaeschke, H., et al. (2014). Circulating microRNA profiles in human patients

with acetaminophen hepatotoxicity or ischemic hepatitis. *Proc. Natl. Acad. Sci. U.S.A.* 111, 12169–12174. doi: 10.1073/pnas.1412608111

Watanabe, T., Tobe, K., Nakachi, Y., Kondoh, Y., Nakajima, M., Hamada, S., et al. (2007). Differential Gene Expression Induced by Two Genotoxic *N*-nitroso Carcinogens, Phenobarbital and Ethanol in Mouse Liver Examined with Oligonucleotide Microarray and Quantitative Real-time PCR. *Genes Environ.* 29, 115–127. doi: 10.3123/jemsge.29.115

Yang, L., Allen, B. C., and Thomas, R. S. (2007). BMDExpress: a software tool for the benchmark dose analyses of genomic data. *BMC Genomics* 8:387. doi: 10.1186/1471-2164-8-387

Yang, X., Greenhaw, J., Shi, Q., Su, Z., Qian, F., Davis, K., et al. (2012). Identification of urinary microRNA profiles in rats that may diagnose hepatotoxicity. *Toxicol. Sci.* 125, 335–344. doi: 10.1093/toxsci/kfr321

Yu, K., Gong, B., Lee, M., Liu, Z., Xu, J., Perkins, R., et al. (2014). Discovering functional modules by topic modeling RNA-seq based toxicogenomic data. *Chem. Res. Toxicol.* 27, 1528–1536. doi: 10.1021/tx50 0148n

Zhu, H., Bouhifd, M., Kleinstreuer, N., Kroese, E. D., Liu, Z., Luechtefeld, T., et al. (2016). t(4) report(): supporting read-across using biological data. *ALTEX* 33, 167–182. doi: 10.14573/altex.1601252

frontiers
in Genetics

Check for
updates

A Pipeline for High-Throughput Concentration Response Modeling of Gene Expression for Toxicogenomics

John S. House[1,2]*, Fabian A. Grimm[3], Dereje D. Jima[1,2], Yi-Hui Zhou[1,4], Ivan Rusyn[3] and Fred A. Wright[1,4,5]

[1] Bioinformatics Research Center, North Carolina State University, Raleigh, NC, United States, [2] Center for Human Health and the Environment, North Carolina State University, Raleigh, NC, United States, [3] Department of Veterinary Integrative Biosciences, Texas A&M University, College Station, TX, United States, [4] Department of Biological Sciences, North Carolina State University, Raleigh, NC, United States, [5] Department of Statistics, North Carolina State University, Raleigh, NC, United States

Cell-based assays are an attractive option to measure gene expression response to exposure, but the cost of whole-transcriptome RNA sequencing has been a barrier to the use of gene expression profiling for *in vitro* toxicity screening. In addition, standard RNA sequencing adds variability due to variable transcript length and amplification. Targeted probe-sequencing technologies such as TempO-Seq, with transcriptomic representation that can vary from hundreds of genes to the entire transcriptome, may reduce some components of variation. Analyses of high-throughput toxicogenomics data require renewed attention to read-calling algorithms and simplified dose–response modeling for datasets with relatively few samples. Using data from induced pluripotent stem cell-derived cardiomyocytes treated with chemicals at varying concentrations, we describe here and make available a pipeline for handling expression data generated by TempO-Seq to align reads, clean and normalize raw count data, identify differentially expressed genes, and calculate transcriptomic concentration–response points of departure. The methods are extensible to other forms of concentration–response gene-expression data, and we discuss the utility of the methods for assessing variation in susceptibility and the diseased cellular state.

Keywords: expression-based dose–response modeling, dose–response modeling, bioinformatics-pipeline, toxicogenomics, bioinformatics & computational biology, iPSCs, cardiomyocytes, expression profiling

INTRODUCTION

Among the key challenges in contemporary toxicity testing is addressing increasing numbers of commodity chemicals with insufficient toxicity characterization, a trend that is at least partially attributable to the limitations associated with *in vivo* testing strategies. Additional challenges are associated with animal to human extrapolation, as well as concerns over the ethics and expense of animal testing. These challenges were described in the National Toxicology Program's (NTP) 2004 *Vision and Roadmap for the 21st Century*, and the National Research Council's (NRC) report on *Toxicity Testing in the 21st Century* (National Research Council, 2007), which envisioned a strategic shift from exclusive reliance on animal-derived data in chemical regulation to the implementation of novel data streams, including high-throughput *in vitro* testing, omics data, and computational modeling. More recently the NRC report, *A Framework to Guide Selection*

of Chemical Alternatives (National Research Council, 2014) and the National Academies report, *Using 21st Century Science to Improve Risk-Related Evaluations* (National Academies, 2017), articulated the need to transition from expensive and incomplete animal testing to high-throughput exposure assessment of new and existing chemicals. Thus, the use of novel data sources to conduct human and animal health risk assessment, including genomic, epigenomic, cell, and *in silico*-based streams, has become imperative.

Gene-expression data are also important in evaluating effects of chemicals on cells and tissues. The Library of Integrated Network-Based Cellular Signatures (LINCS) is a database of over 1 million gene expression signatures, or perturbations, generated using a targeted hybridized bead-base flow sorter (Peck et al., 2006; Duan et al., 2014) from either drug/chemical exposure or biological knockdown/knockout with 50 different cell types (Campillos et al., 2008) on a ∼1,000 gene-set with full-transcriptome imputation. The related Connectivity Map (Lamb et al., 2006), a database of transcriptional multiplexed microarray technology from multiple cancer cell lines exposed to ∼5,000 drugs and small-molecule compounds, has since been combined into the NIH LINCS database. Further, the DrugMatrix® database contains, in addition to a myriad of phenotypic endpoints, microarray gene-expression data in the rat for over 600 different compounds in multiple tissues. Lastly, the Toxicogenomics Project-Genomics Assisted Toxicity Evaluation Systems (TG-GATEs) database has compiled toxicological endpoints and gene expression data from rats (*in vivo* and *in vitro* primary hepatocytes) and humans (*in vitro* primary hepatocytes) on 170 hepato- and renal-toxicants at multiple doses and time points (Uehara et al., 2010; Igarashi et al., 2015).

The advent of next-generation sequencing (NGS) technology has allowed for dramatic advances in the characterization of genomic, epigenomic, and gene expression endpoints. In contrast to reverse transcriptase PCR and microarrays, NGS results in reduction or elimination of numerous sources of variation (Okoniewski and Miller, 2006; Klebanov et al., 2007; Royce et al., 2007). Anticipating a reduced cost in interrogating only a portion of the transcriptome, phase III of the Tox21 initiative has included the development of the S1500 human gene-set (Merrick et al., 2015) to use for chemical and drug screening. This S1500 gene-set was designed to be representative of the human transcriptome, inclusive of the original L1000 (LINCS) gene-set, and optimized for pathway coverage and co-expression information. The main goal is to represent diversity of expression response to disease and chemical exposure in a cost-effective manner.

RNA-Seq, although considered the gold standard for gene expression (Ellis et al., 2013), does have shortcomings. These include bias introduced during mRNA enrichment and library preparation (Han et al., 2015), as well as substantial monetary, computing hardware, and bioinformatics costs. Targeted sequencing technology, such as TempO-Seq[TM] (Templated Oligo assay with Sequencing readout), which was originally adapted from RASL-seq (Li et al., 2012), specifically targets unpurified RNA in cellular lysates. Two detector oligos are used that can only be ligated when hybridized next to each other

on RNA, and confer specificity and eliminate positional bias introduced by poly-(A)+ selection. Sequencing and sample-specific adapters are then hybridized to the original probes for sequencing (Yeakley et al., 2017), allowing for assessment of differentially expressed transcripts in a high-throughput manner while alleviating some of the shortcomings of untargeted RNA-seq. These potential improvements in cost and reduction in sources of variation are attractive for high-throughput concentration–response transcriptomic profiling.

To date, much of the effort to characterize toxicity transcriptomic endpoints has focused on individual concentration–responses from drugs and a small number of environmental chemicals. Targeted sequencing allows for more economical interrogation of the transcriptome and opens the door for high-quantity, high-throughput assessment of drug/chemical concentration–response. We report here a pipeline for utilizing TempO-Seq (BioSpyder Technologies, Inc., Carlsbad, CA, United States), a targeted RNA sequencing technology, to assess gene-transcript concentration–response relationships to chemical exposure. Much of the pipeline is extensible to any transcriptional profiling of chemical response, where sample sizes are likely to be modest. For proof of principle, we illustrate using 2,982 selected genes that include the "S1500+"[1] gene-set. Induced pluripotent stem cells (iPSC)-cardiomyocytes were treated with three different doses of four chemicals to assess their effects on gene expression and concentration–response point of departure (POD) (GEO accession number GSE105050).

MATERIALS AND METHODS

Chemicals and Biologicals

iCell cardiomyocytes (cat. no.: CMC-100-010-001), and cardiomyocyte plating and maintenance media were purchased from Cellular Dynamics International (Madison, WI, United States). Reference chemicals isoproterenol (CAS: 7683-59-2) and propranolol (CAS: 525-66-6) were purchased as part of EarlyTox Cardiotoxicity Screening kits (cat. no.: R8211) from Molecular Devices LLC (Sunnyvale, CA, United States). Nifedipine (CAS: 21829-25-4) and dofetilide (CAS: 115256-11-6) were purchased from APEXBio (Houston, TX, united States). Dimethyl sulfoxide (DMSO, cat. no.: sc-358801, CAS: 67-68-5) was purchased from Santa Cruz Biotechnology (Dallas, TX, United States). Penicillin/streptomycin solution (cat. no.: 10378016) and 0.4% Trypan Blue solution (cat. no.: 15250061) were obtained from Life Technologies (Grand Island, NY, United States).

Cardiomyocyte Cell Culture

iCell cardiomyocytes were plated and maintained according to the manufacturer's recommendations (Cellular Dynamics International, Madison, WI, United States) and in accordance with previously published protocols with minor adjustments (Sirenko et al., 2013a,b, 2017; Grimm et al., 2015, 2016).

---

[1]https://federalregister.gov/a/2015-08529

Individual units of cardiomyocytes were thawed for 4 min in a 37°C water bath and subsequently resuspended in 10 ml of cardiomyocyte plating medium containing 1:500 penicillin/streptomycin solution. Following microscopic cell counting using the trypan blue exclusion method, the cell density was adjusted to a final plating density of 2×10^5 cells/ml. Twenty-five microliters of cell suspension was then transferred per well to a 384-well microplate, yielding a final cell density of 5,000/well. Tissue-culture treated microplates (cat. no.: 353962, Corning Life Sciences, Corning, NY, United States) were gelatinized for 2 h at 37°C with 25 µl 0.1% gelatin in water before cardiomyocytes were plated. After disposal of the gelatin solution and addition of the cell suspension, microplates were kept at room temperature for 30 min. Cells were then incubated at 37°C and 5% CO_2 for 48 h. The plating medium was then exchanged with 40 µl of maintenance medium containing 1:500 penicillin/streptomycin solution per well. Maintenance medium was replaced every 48–72 h until day 13 post-plating. The maintenance medium was then exchanged with 50 µl fresh medium per well and incubated overnight. Cells were treated the next morning (day 14 post-plating) with 12.5 µl 5× chemical solutions in 0.5% DMSO (v/v) in media (vehicle) in addition to untreated or vehicle-treated negative control wells, and incubated at 37°C and 5% CO_2. Following 24 h of incubation, the cell medium was discarded, and cardiomyocytes were lysed with 10 µl 1× lysis buffer provided in the TempO-Seq assay kit (BioSpyder Technologies, Inc., Carlsbad, CA, United States). Lysate-containing microplates were agitated at 300 rpm using a benchtop microplate shaker and stored at −80°C until further use.

TempO-Seq Library Preparation and Sequencing

Differential gene expression patterns and concentration–response relationships were analyzed using TempO-Seq[TM] (BioSpyder Technologies, Inc., Carlsbad, CA, United States) (Yeakley et al., 2017), a targeted RNA sequencing technology focused on a surrogate transcriptome panel comprising 2,982 transcripts, as described previously (Biopyder Toxpanel Library DO-01-096) (Grimm et al., 2016). The sequencing library was prepared according to the manufacturer's guidelines and as previously described (Grimm et al., 2015). In brief, RNA in 2 µl of each cell lysate was hybridized with the provided detector oligo pool mix (2 µl per sample) using the following thermocycler settings: 10 min at 70°C, followed by gradual decrease to 45°C over 49 min, and ending with 45°C for 1 min. Subsequent steps included nuclease digestion (90 min at 37°C) ligation step (60 min at 37°C, followed by heat denaturation at 80°C for 30 min) following addition of 24 µl nuclease mix and 24 µl ligation mix. Ten microliters of each ligation product was then transferred to a 96-well amplification microplate containing 10 µl of PCR mix per well. The ligation products were then uniquely labeled during product amplification, when well-specific, "barcoded" primer pairs were introduced to templates. Sequence-based barcoding is an essential step allowing for correct identification and recognition of transcript-specific sequencing counts. Five microliters of sample amplicons from each well was subsequently pooled into a single sequencing library. The TempO-Seq library was further processed using a PCR clean-up kit (Clontech, Mountain View, CA, United States) prior to sequencing at Texas A&M University Genomics & Bioinformatics Services. Sequencing was achieved using a 50 single-end read mode in a rapid flow cell (two sequencing lanes for increased sequencing depth; mean reads per gene = 212) on a HiSeq 2500 Ultra-High-Throughput Sequencing System (Illumina, San Diego, CA, united States). The high-expression genes listed in Supplementary Table 1 were attenuated to allow for more sequencing depth. Sequence cluster identification, quality pre-filtering, base calling, and uncertainty assessment were conducted in real time using Illumina's HCS 2.2.68 and RTA 1.18.66.3 software with default parameter settings. Sequencing readouts were demultiplexed to generate FASTQ files, and passed all internal quality controls (GEO accession number GSE105050).

Temposeqcount Application: Availability and Implementation

Temposeqcount installs all dependencies in a Python virtual environment. It is released as an open-source software under the GNU General Public License and available from https://github.com/demis001/temposeqcount. Complete installation instructions are provided. Note that a unix operating system is required for this portion only.

Pathway Analysis

The $\log_2$(fold change) (l2fc) and *p*-values from DESeq2 for dofetilide and nifedipine were analyzed through the use of IPA (Ingenuity® Systems[2]). A core analysis was run using: User Dataset as reference, cutoff of $p < 0.05$, and all other values as default. The cardiac-related pathways (with *p*-value < 0.05) for *Tox Functions under Diseases* and *Functions* were chosen for Supplementary Figure 2.

Differential Gene Expression and Concentration Response

Sample dataset, hash file, figures, and R Scripts for generating all figures and processes are available from https://github.com/jshousephd/HT-CBA.

RESULTS AND DISCUSSION

Process Overview

The transcriptomic analysis methods discussed here focus on assessment of differentially expressed genes (DEGs) and concentration–response assessment using counts from TempO-Seq experiments. However, the methods and processes described herein are extensible to most types of count data. The sample dataset used and provided for illustration consists of 2 sets of 12 vehicle controls and 4 chemical treatments at three concentration levels (0.1, 1.0, and 10 µM). Two inotropic agents,

---

[2]www.ingenuity.com

the β-adrenergic receptor agonist and antagonist isoproterenol and propranolol, are known negative controls for cardiac QT prolongation. Nifedipine is a calcium channel blocker used to treat hypertension and dofetilide is an antiarrhythmic.

As shown in **Figure 1**, the analysis pipeline can be broken into four major steps: (1) generation of the count matrix from sequenced reads (coded in *python*), (2) quality control and count normalization, (3) identification and visualization of DEGs, and (4) concentration–response modeling and POD assessment. POD assessment combines output from *tcpl* (Filer et al., 2017) augmented with additional model fitting as described below.

TempO-Seq Count Matrix Generation

Prior to assessment of DEGs and concentration response modeling, raw sequenced reads are aligned to probe sequences and counted. TempO-Seq is a high-throughput targeted sequencing technology that uses template-dependent oligo ligation on a multi-well plate. Generation of a count matrix from TempO-Seq data requires far less computing resources

than traditional whole-transcriptome RNA-seq. Since the reads are generated from targeted probes, the reference file is several orders of magnitude smaller than a genome reference. For this application, sequencing reads were de-multiplexed by the sequencing facility. After de-multiplexing, a single experimental layout can result in up to 384 or 1,536 fastq files (depending on plate format), each file with reads resulting from an experimental condition. In traditional RNA-seq, individual fastq files are each aligned to a reference sequence individually using a short read aligner such as STAR, BWA, or bowtie (Langmead et al., 2009; Li and Durbin, 2010; Dobin et al., 2013) and then counted using a different command line utility. However, these routines are less useful in a TempO-Seq experiment due to the large number of fastq files generated in a single Tempo-Seq run and the provision of reference sequences.

Accordingly, we developed an application called *temposeqcount* to facilitate this process using a Ruffus framework in Python (Goodstadt, 2010) which is illustrated in **Figure 2**. Briefly, the application accepts the manufacturer provided probe

FIGURE 1 | Pipeline overview. The pipeline consists of four major parts: I. Generation of the count matrix from sequenced reads, II. Quality-control and generation of normalized counts, III. Identification of differentially expressed genes (DEGs), and IV. Assessment of point of departure (POD) from count data.

FIGURE 2 | *temposeqcount* overview. Self-contained and implemented in a Ruffus framework. A directory of fastq files and the probe manifest are the only inputs to generate the count matrix required for the remainder of the pipeline.

manifest CSV and a directory of fastq files as input. Initially, the probe sequence is parsed from the manifest file which generates the probe fasta and pseudo-gtf annotation files. The probe fasta file is indexed using *genomeGenerate* function in the STAR aligner (Dobin et al., 2013) and STAR aligner is used to align a fastq file to indexed probe sequences. Lastly, htseq-count in the HTSeq (Anders et al., 2015) application is used to count probe-aligned reads. The *temposeqcount* application accepts STAR aligned bam and pseudo annotation-gtf files internally to generate a count for each sample. As output, an alignment summary is generated and count files are merged and formatted into a single count matrix (K) consisting of gene$_i$ rows and treatment$_j$ columns for downstream analysis.

QC and Normalization of Counts

The remainder of the pipeline (**Figure 1**; steps II, III, and IV) consists of R scripts that are publicly available[3]. The inputs to the rest of the pipeline consist of the count matrix (K) generated

in step I, and an experimental layout file (hereafter referred to as hash file) from the experimenter. Although we are using this process for counts from the TempO-Seq assay, the pipeline here can be applied to other types of high-throughput sequencing data. Multiple attributes can be included in the hash file for each treatment, but column names of the count matrix must have a corresponding column entry in the hash file. Gene features are filtered for >1 count per row across the experimental count matrix, which resulted in 100 genes removed, leaving 2,882 for subsequent analyses (**Figure 3D**). Prior to normalization, sample count totals are evaluated graphically (**Figure 3A**) and samples (columns) failing to exceed a user defined minimum count threshold (we used 100,000 per sample for the library of 2,982 features in the TempO-seq kits used in these experiments) are removed from subsequent analyses. The design of concentration–response experiments for multiple compounds often uses shared controls. For the data examined herein, there were 24 vehicle controls. In the pipeline, controls are examined by principal component analysis (PCA; **Figure 3B**). In addition, we examine the average correlations of each control sample with the

---

[3]https://github.com/jshousephd/HT-CBA
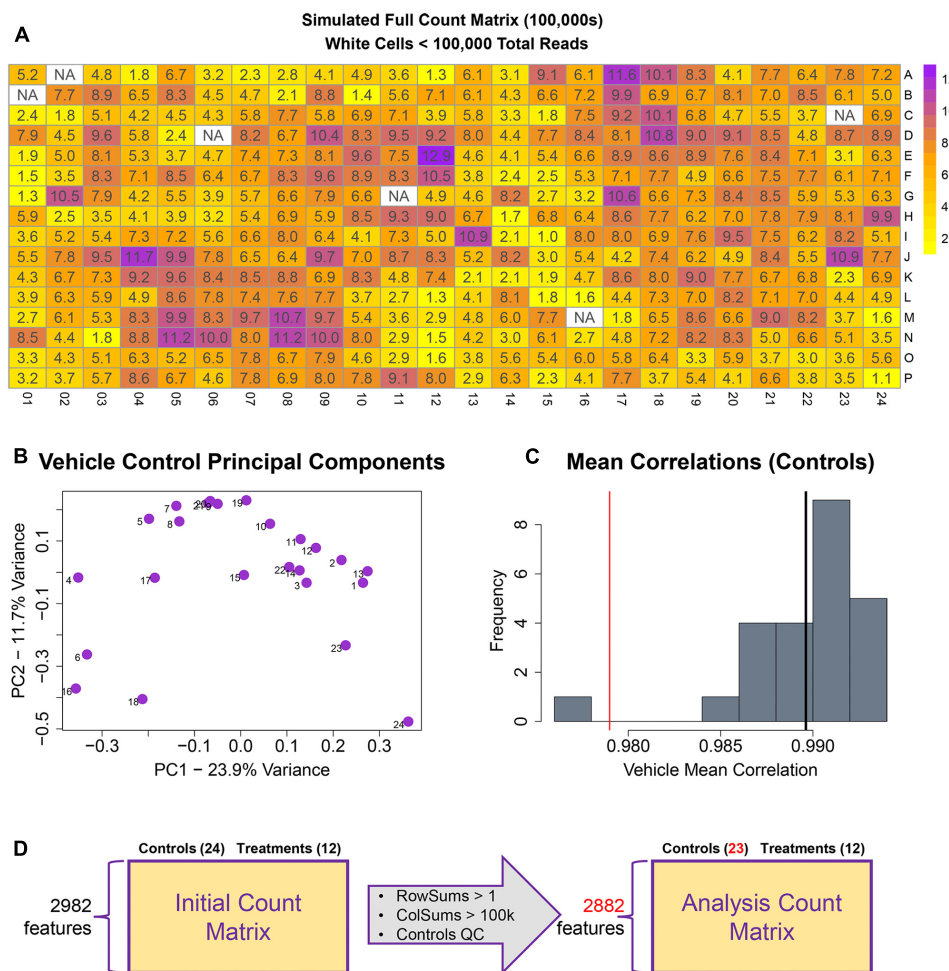
FIGURE 3 | Raw count quality control assessment. **(A)** Simulated raw counts for a full 384-well TempO-Seq experiment. **(B)** First two principal components of vehicle controls. **(C)** Histogram of sample correlations for all vehicle controls. Black bar is the mean of sample correlations across all genes. Red bar represents 3 standard deviations from mean. **(D)** Summary of QC changes to the analysis matrix. 100 features were dropped due to <2 counts across all columns (2982–2882) and one control was dropped (leaving 23) due to low correlation with other control samples. No samples had <100 k counts.

remaining samples (**Figure 3C**), an approach termed the "D statistic" and similar to that used by the GTEx Consortium (Consortium, 2015) to filter low-quality samples. In this analysis, we removed control samples for which the average correlation with remaining control samples was 3 standard deviations lower than a mean computed for all controls (**Figure 3C**, red line). Pairwise correlations and scatterplots were also examined for controls prior to normalization (Supplementary Figure 1). The final analysis count matrix (**Figure 3D**) was then normalized experiment-wide at the treatment level with DESeq2, which models read counts using a negative binomial distribution and normalizes based on a model that uses dispersion estimates for each gene across all treatments (Love et al., 2014).

Analysis of Differential Gene Expression

Differentially expressed genes were determined using DESeq2 prior to concentration response modeling. DEGs were first identified for the maximum concentration of each treatment,

with $\log_2$ fold change (l2fc) values, p-values, and adjusted p-values computed for each gene and combined into a single dataset for each chemical. For the sample data, as seen in the summary plot of the number of DEGs per treatment (**Figure 4A**), nifedipine was the most transcriptionally active treatment with identified differential gene expression for nearly a third of interrogated transcripts (918/2,882). Dofetilide affected 425 transcripts, while isoproterenol and propranolol had little effect on transcription, with 32 and 29 identified DEGs, respectively (**Figure 4A**).

A typical TempO-Seq experiment may have 50–75 unique concentration-chemical combinations making it important to examine how they group together regarding differential gene expression. There are several ways to do this, and we have illustrated some of them in **Figure 4** with the sample dataset. Using all l2fc values for each chemical, the first three principal components were plotted (**Figure 4B**). In the sample data, the two drugs with effects on the heart rhythm (isoproterenol
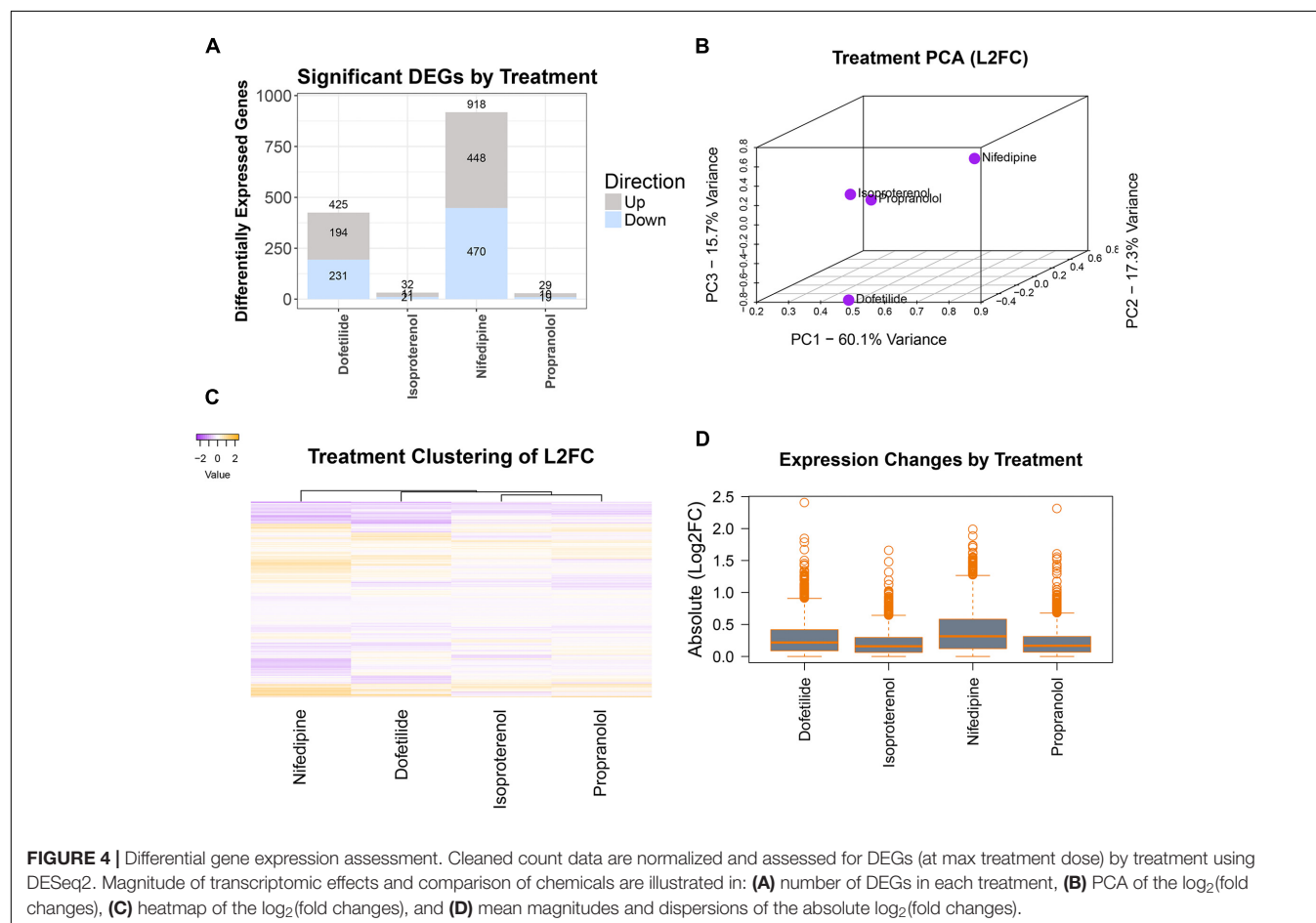
and propranolol) clustered together while the others (nifedipine and dofetilide) were quite distinct. A heatmap of l2fc values further illustrates the similarities between the isoproterenol and propranolol, while highlighting how both nifedipine and dofetilide are each different from these QT-prolongation controls and from each other (**Figure 4C**). The overall magnitude of transcription effects from each chemical is shown in the boxplot of absolute l2fc values (**Figure 4D**).

Concentration–Response Modeling Decision Logic

A critical step in human health assessment for a chemical compound is the determination of the POD from a dose/concentration–response relationship. In the sample data, iPSC cardiomyocytes were exposed to either vehicle or three increasing drug concentrations (0.1, 1.0, and 10 μM). Vehicle controls were assigned a dose value on the log_{10} scale that is one average dose distance below the lowest treatment concentration (**Figure 5B**). Our dataset contained 24 vehicle controls and a single treatment at each of the three concentrations. Thus, since one control was removed in QC, the dosing vector for this experiment consists of 23 values of −2, followed by −1, 0, and 1, while the response vector consists of normalized counts data at each control/treatment. To allow for zero counts, normalized

counts were $log_2(counts + 0.5)$ transformed and mean-centered to vehicle controls. An example plot of the data is shown in **Figure 5B**.

To facilitate decision-making accompanying our concentration–response modeling, we created a process tree that utilizes statistical flags (**Figures 5A,B**). The q-values (FDR values for each p-value threshold), are calculated for: (1) the moment-corrected correlation (MCC) trend test (Zhou and Wright, 2015) for the entire concentration–response range (**Figure 5B**, purple oval), (2) MCC for treatment doses only (**Figure 5B**, green oval), and (3) Wilcoxon's statistic for a difference between vehicle control values vs. treatments (**Figure 5B** green oval vs. orange oval). MCC is a trend procedure that is intended to be powerful while retaining robustness to a wide variety of distributional forms of the data test (Zhou and Wright, 2015). These flags were used to decide which gene/treatment combinations should be fit for a concentration–response (**Figure 5A**). If the overall trend (**Figure 5B**, purple) in the concentration–response is not significant, the experimental maximum dose is assigned as the POD. For those genes where the overall trend is significant, control counts are compared to treatment counts using a Wilcoxon Rank Sum test. If these two groups are different from each other (**Figure 5B**, green vs. orange), the gene/treatment is selected for concentration–response modeling. If controls are not

FIGURE 4 | Differential gene expression assessment. Cleaned count data are normalized and assessed for DEGs (at max treatment dose) by treatment using DESeq2. Magnitude of transcriptomic effects and comparison of chemicals are illustrated in: **(A)** number of DEGs in each treatment, **(B)** PCA of the log_2(fold changes), **(C)** heatmap of the log_2(fold changes), and **(D)** mean magnitudes and dispersions of the absolute log_2(fold changes).

FIGURE 5 | Statistical flag generation for concentration-response modeling. **(A)** Flowchart of decision logic used in determining the gene-treatment combinations to assess for concentration response. **(B)** Example concentration–response curve used for statistical flag generation. Concentration–response relationships where overall ρ (purple circle) is significant, and controls not equal to treatments (green vs. orange), are chosen for subsequent concentration response modeling, as are relationships where ρ for treatments only (green circle) are significant.

different from treatments but did pass the overall trend test for significance, we then assess whether a trend exists in treatments only (**Figure 5B**, green). If the treatment-only trend is significant, it is also selected for POD modeling. For those remaining genes that are not significant for the treatment only trend test, a report is generated for the remaining items for manual experimenter follow-up.

Concentration–Response Modeling and Point of Departure Calculation

As seen in **Figure 5B**, our data had only a single replicate at each of three concentrations of a given chemical. To illustrate clearly dose response modeling we show simulated data for 12

control replicates and three concentrations with three replicates each (**Figures 6A,B**). For those gene/treatments identified for concentration–response modeling, "dose" vectors and response counts are first fit with *tcpl* functions (Filer et al., 2017) with a constant model that represents a null fit, a gain-loss model, and a three-parameter hill model with the "floor" set to zero (**Figure 6A**). We also assess a four-parameter hill function fit using the R-DRM package (Ritz et al., 2015) where the "floor" is not set to zero, and assess the best-fitting model by the smallest Akaike information criterion (AIC) which penalizes model over-fitting. Following selection of the best model, the POD is assessed by determining the concentration that elicits a 1 standard deviation departure from the control mean (**Figures 6A,B**; dotted purple line), although other POD approaches or benchmark dose (Sirenko et al., 2013b; Wignall et al., 2014; Filer et al., 2017) could be used. For the sample with only three treatment concentrations, the three-parameter hill function was the predominant "winner" in terms of minimum AIC, although a four-parameter hill function provides the best fit in 28% of the fitted curves (**Figure 6C**). Although nifedipine caused more than twice the number of differentially expressed genes as dofetilide (**Figure 4A**), dofetilide actually had a smaller mean POD (**Figure 6D**). Propranolol is not shown in **Figure 6D** as it had no genes exhibiting concentration–response relationships.

TempO-seq is a new technology with few publications to date and no standard pipeline for analysis. A recent study by Yeakley et al. (2017) (GEO GSE91395_Dose_Response_Read_Counts.xlsx) also followed a concentration–response study design. We thus used their data as another highlight speed and utility of our pipeline. First, using our pipeline we found 4,178 genes that were differentially expressed in MCF-7 cells treated with 1 μM Trichostatin A, while Yeakley et al. (2017) had reported 4,154 differentially expressed genes. We then used our concentration–response pipeline to calculate POD estimates for several top upregulated genes and two of the novel genes they reported for Trichostatin A response. These are graphically represented in Supplementary Figure 2.

CONCLUSION AND SUMMARY

As sequencing has become more affordable, the number of experiments with sequencing (expression) data has grown exponentially, and multiple computational tools are available for different steps in the analysis. Similarly, many dose–response modeling tools are available (Wignall et al., 2014; Filer et al., 2017; Sirenko et al., 2017), including for gene expression data at the level of genes and pathways (Yang et al., 2007).

Gene expression data have been an important contributor to the mechanistic studies in toxicology and other biomolecular fields (Luo et al., 2017). Toxicogenomics is a valuable tool for predictive modeling (Uehara et al., 2008; De Abrew et al., 2015) and read-across (Low et al., 2011; Grimm et al., 2016). However, the high cost of gene expression studies has largely precluded the use of toxicogenomics as a standard tool for dose–response assessment in studies of adverse effects of drugs and
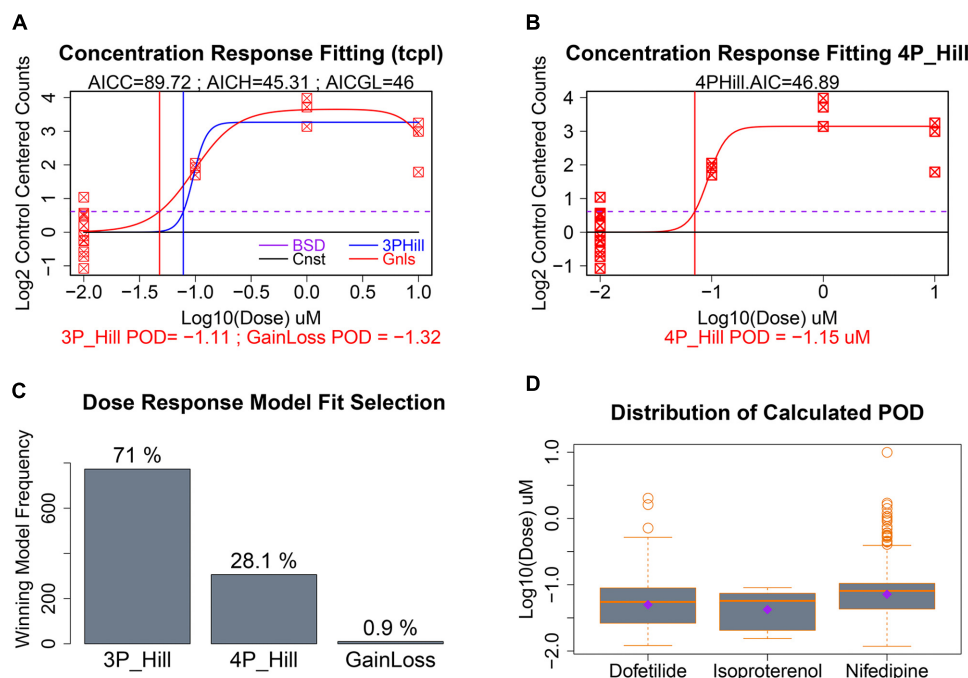
FIGURE 6 | Concentration–response modeling and POD calculation. **(A)** Baseline deviation (BSD) is set to 1 SD from mean of controls (dotted purple line). The tcpl package is used to fit (1) constant, (2) gain loss, and (3) three-parameter hill models. **(B)** The DRC package in R is used to fit a four-parameter hill function and the model with the smallest AIC is chosen as "best" with corresponding POD used where fitted model crosses BSD. **(C)** Distribution of the winning model for these data. **(D)** Evaluation of the dispersion, mean, and median of calculated POD deviations by chemical treatment.

chemicals. While some databases do include dose- and time-dependent profiling of hundreds of drugs and chemicals (Luo et al., 2017), the potential power of dose–response expression profiling has not been fully harnessed. The potential of dose–response toxicogenomics data as a truly predictive tool was first demonstrated by Thomas et al. (2011, 2013), who showed that gene expression data-based PODs derived from short-term studies are well-correlated with the PODs for the apical endpoints from 90-day and 2-year animal studies (Farmahin et al., 2017). A recent study demonstrated the value of dose–response genomics in a comparative analysis of chlorinated solvents in liver and kidney (Zhou et al., 2017). Thus, with the advent of lower cost and higher throughput platforms for gene expression profiling, dose–response modeling will become a major output of these experiments, including in *in vitro* studies.

The methods outlined in this manuscript provide a framework for highly automated assessment of transcriptomic concentration–response POD estimates. Although we have used targeted sequencing data, these methods are extensible to any kind of concentration–response count data, including whole-transcriptome RNA sequencing. Probe-based targeted RNA-seq technology does have many advantages. These procedures are extremely fast and can be run on desktop computers instead of computing clusters. The biases associated with the purification and library creation in RNA-seq are not applicable with this technology. However, it is important to note that this targeted RNA-seq technology does not capture underlying genetic

variation in a region of interest. The probes are highly specific to the 50-mer being interrogated. The methods described here have been made freely available, to provide tools to characterize the transcriptomic dose response and concentration response effects of drugs and chemicals in novel targeted-probe high-throughput formats.

The gene expression response signatures identified by our pipeline can be used for hazard assessment, drug repurposing, and disease characterization (Stegmaier et al., 2004; Hieronymus et al., 2006; Wei et al., 2006; Sirota et al., 2011). We note that agonism is the characterized mode of action for isoproterenol and propranolol, and antagonism of the β-adrenergic receptor for positive and negative inotropic effects. We report few transcriptional effects from treatment with either of these drugs. One possible interpretation is the presence of very little non-receptor-mediated mode of action. In contrast, treatment of iPSC cardiomyocytes with nifedipine, a calcium channel blocker used to treat hypertension, resulted in substantial transcriptional changes. Dofetilide, an antiarrhythmic agent that increases the QT interval by selectively blocking the rapid component of the cardiac ion channel delayed rectifier current (Roukoz and Saliba, 2007), also resulted in substantial transcriptional changes. Further extension of these results into pathway analysis and biological read across will facilitate additional decision-making processes for hazard and risk characterization, drug repurposing, and hypothesis formation. Although our sample data provided contained few replicates at each dose, we still were able to

identify DEGs and calculate POD estimates. As a proof of principle, we examined DEGs for dofetilide and nifedipine using Ingenuity's Pathway Analysis. The top identified toxicology pathways were heart and liver related for dofetilide, and more diverse for nifedipine. The statistically significant overlapping cardiac-related pathways are shown in Supplementary Figure 3.

In summary, with the advent of cheaper sequencing technology and targeted sequencing technology, it has become feasible and necessary to utilize the value of gene expression data in high-throughput experiments, for dose response characterization, for perturbation signature identification, and for biological read-across assessment. The methods herein provide a reproducible, largely automated framework to utilize such sequencing data to identify treatment-induced DEGs and concentration response estimates.

AUTHOR CONTRIBUTIONS

JH: experimental design, methodology development, programming, writing and editing, and submitting. FG: experimental design, all wet-lab work, and writing and editing. DJ: experimental design, methodology development, programming, and writing and editing. Y-HZ: methodology development. IR: study conception, experimental design, support, and writing and editing. FW: study conception, experimental design, support, writing and editing.

REFERENCES

Anders, S., Pyl, P. T., and Huber, W. (2015). HTSeq–a python framework to work with high-throughput sequencing data. *Bioinformatics* 31, 166–169. doi: 10.1093/bioinformatics/btu638

Campillos, M., Kuhn, M., Gavin, A. C., Jensen, L. J., and Bork, P. (2008). Drug target identification using side-effect similarity. *Science* 321, 263–266. doi: 10.1126/science.1158140

Consortium, G. T. (2015). Human genomics. The genotype-tissue expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* 348, 648–660. doi: 10.1126/science.1262110

De Abrew, K. N., Overmann, G. J., Adams, R. L., Tiesman, J. P., Dunavent, J., Shan, Y. K., et al. (2015). A novel transcriptomics based in vitro method to compare and predict hepatotoxicity based on mode of action. *Toxicology* 328, 29–39. doi: 10.1016/j.tox.2014.11.008

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., et al. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21. doi: 10.1093/bioinformatics/bts635

Duan, Q., Flynn, C., Niepel, M., Hafner, M., Muhlich, J. L., Fernandez, N. F., et al. (2014). LINCS canvas browser: interactive web app to query, browse and interrogate LINCS L1000 gene expression signatures. *Nucleic Acids Res.* 42, W449–W460. doi: 10.1093/nar/gku476

Ellis, S. E., Gupta, S., Ashar, F. N., Bader, J. S., West, A. B., and Arking, D. E. (2013). RNA-Seq optimization with eQTL gold standards. *BMC Genomics* 14:892. doi: 10.1186/1471-2164-14-892

Farmahin, R., Williams, A., Kuo, B., Chepelev, N. L., Thomas, R. S., Barton-Maclaren, T. S., et al. (2017). Recommended approaches in the application of toxicogenomics to derive points of departure for chemical risk assessment. *Arch. Toxicol.* 91, 2045–2065. doi: 10.1007/s00204-016-1886-5

Filer, D. L., Kothiya, P., Setzer, R. W., Judson, R. S., and Martin, M. T. (2017). tcpl: the ToxCast pipeline for high-throughput screening data. *Bioinformatics* 33, 618–620. doi: 10.1093/bioinformatics/btw680

FUNDING

ACKNOWLEDGMENT

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2017.00168/full#supplementary-material

Goodstadt, L. (2010). Ruffus: a lightweight Python library for computational pipelines. *Bioinformatics* 26, 2778–2779. doi: 10.1093/bioinformatics/btq524

Grimm, F. A., Iwata, Y., Sirenko, O., Bittner, M., and Rusyn, I. (2015). High-content assay multiplexing for toxicity screening in induced pluripotent stem cell-derived cardiomyocytes and hepatocytes. *Assay Drug Dev. Technol.* 13, 529–546. doi: 10.1089/adt.2015.659

Grimm, F. A., Iwata, Y., Sirenko, O., Chappell, G. A., Wright, F. A., Reif, D. M., et al. (2016). A chemical-biological similarity-based grouping of complex substances as a prototype approach for evaluating chemical alternatives. *Green Chem.* 18, 4407–4419. doi: 10.1039/c6gc01147k

Han, Y., Gao, S., Muegge, K., Zhang, W., and Zhou, B. (2015). Advanced applications of RNA sequencing and challenges. *Bioinform. Biol. Insights* 9(Suppl. 1), 29–46. doi: 10.4137/BBI.S28991

Hieronymus, H., Lamb, J., Ross, K. N., Peng, X. P., Clement, C., Rodina, A., et al. (2006). Gene expression signature-based chemical genomic prediction identifies a novel class of HSP90 pathway modulators. *Cancer Cell* 10, 321–330. doi: 10.1016/j.ccr.2006.09.005

Igarashi, Y., Nakatsu, N., Yamashita, T., Ono, A., Ohno, Y., Urushidani, T., et al. (2015). Open TG-GATEs: a large-scale toxicogenomics database. *Nucleic Acids Res.* 43, D921–D927. doi: 10.1093/nar/gku955

Klebanov, L., Chen, L., and Yakovlev, A. (2007). Revisiting adverse effects of cross-hybridization in Affymetrix gene expression data: do they matter for correlation analysis? *Biol. Direct* 2:28. doi: 10.1186/1745-6150-2-28

Lamb, J., Crawford, E. D., Peck, D., Modell, J. W., Blat, I. C., Wrobel, M. J., et al. (2006). The connectivity map: using gene-expression signatures to connect small molecules, genes, and disease. *Science* 313, 1929–1935. doi: 10.1126/science.1132939

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10:R25. doi: 10.1186/gb-2009-10-3-r25

Li, H., and Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 26, 589–595. doi: 10.1093/bioinformatics/btp698

Li, H., Qiu, J., and Fu, X. D. (2012). RASL-seq for massively parallel and quantitative analysis of gene expression. *Curr. Protoc. Mol. Biol.* Chapter 4:Unit 4.13.1–9. doi: 10.1002/0471142727.mb0413s98

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15:550. doi: 10.1186/s13059-014-0550-8

Low, Y., Uehara, T., Minowa, Y., Yamada, H., Ohno, Y., Urushidani, T., et al. (2011). Predicting drug-induced hepatotoxicity using QSAR and toxicogenomics approaches. *Chem. Res. Toxicol.* 24, 1251–1262. doi: 10.1021/tx200148a

Luo, G., Shen, Y., Yang, L., Lu, A., and Xiang, Z. (2017). A review of drug-induced liver injury databases. *Arch. Toxicol.* doi: 10.1007/s00204-017-2024-8 [Epub ahead of print].

Merrick, B. A., Paules, R. S., and Tice, R. R. (2015). Intersection of toxicogenomics and high throughput screening in the Tox21 program: an NIEHS perspective. *Int. J. Biotechnol.* 14, 7–27. doi: 10.1504/IJBT.2015.074797

National Academies (2017). *Using 21st Century Science to Improve Risk-Related Evaluations*. Washington, DC: The National Academies Press.

National Research Council (2014). *A Framework to Guide Selection of Chemical Alternatives*. Washington, DC: The National Academies Press.

National Research Council (2007). *Toxicity Testing in the 21st Century*. Washington, DC: The National Academies Press.

Okoniewski, M. J., and Miller, C. J. (2006). Hybridization interactions between probesets in short oligo microarrays lead to spurious correlations. *BMC Bioinformatics* 7:276. doi: 10.1186/1471-2105-7-276

Peck, D., Crawford, E. D., Ross, K. N., Stegmaier, K., Golub, T. R., and Lamb, J. (2006). A method for high-throughput gene expression signature analysis. *Genome Biol.* 7:R61. doi: 10.1186/gb-2006-7-7-r61

Ritz, C., Baty, F., Streibig, J. C., and Gerhard, D. (2015). Dose-response analysis using R. *PLOS ONE* 10:e0146021. doi: 10.1371/journal.pone.0146021

Roukoz, H., and Saliba, W. (2007). Dofetilide: a new class III antiarrhythmic agent. *Exp. Rev. Cardiovasc. Ther.* 5, 9–19. doi: 10.1586/14779072.5.1.9

Royce, T. E., Rozowsky, J. S., and Gerstein, M. B. (2007). Toward a universal microarray: prediction of gene expression through nearest-neighbor probe sequence identification. *Nucleic Acids Res.* 35:e99. doi: 10.1093/nar/gkm549

Sirenko, O., Crittenden, C., Callamaras, N., Hesley, J., Chen, Y. W., Funes, C., et al. (2013a). Multiparameter in vitro assessment of compound effects on cardiomyocyte physiology using iPSC cells. *J. Biomol. Screen.* 18, 39–53. doi: 10.1177/1087057112457590

Sirenko, O., Cromwell, E. F., Crittenden, C., Wignall, J. A., Wright, F. A., and Rusyn, I. (2013b). Assessment of beating parameters in human induced pluripotent stem cells enables quantitative in vitro screening for cardiotoxicity. *Toxicol. Appl. Pharmacol.* 273, 500–507. doi: 10.1016/j.taap.2013.09.017

Sirenko, O., Grimm, F. A., Ryan, K. R., Iwata, Y., Chiu, W. A., Parham, F., et al. (2017). In vitro cardiotoxicity assessment of environmental chemicals using an organotypic human induced pluripotent stem cell-derived model. *Toxicol. Appl. Pharmacol.* 322, 60–74. doi: 10.1016/j.taap.2017.02.020

Sirota, M., Dudley, J. T., Kim, J., Chiang, A. P., Morgan, A. A., Sweet-Cordero, A., et al. (2011). Discovery and preclinical validation of drug indications using compendia of public gene expression data. *Sci. Transl. Med.* 3:96ra77. doi: 10.1126/scitranslmed.3001318

Stegmaier, K., Ross, K. N., Colavito, S. A., O'Malley, S., Stockwell, B. R., and Golub, T. R. (2004). Gene expression-based high-throughput screening(GE-HTS) and application to leukemia differentiation. *Nat. Genet.* 36, 257–263. doi: 10.1038/ng1305

Thomas, R. S., Clewell, H. J., Allen, B. C., Wesselkamper, S. C., Wang, N. C., Lambert, J. C., et al. (2011). Application of transcriptional benchmark dose values in quantitative cancer and noncancer risk assessment. *Toxicol. Sci.* 120, 194–205. doi: 10.1093/toxsci/kfq355

Thomas, R. S., Wesselkamper, S. C., Wang, N. C., Zhao, Q. J., Petersen, D. D., Lambert, J. C., et al. (2013). Temporal concordance between apical and transcriptional points of departure for chemical risk assessment. *Toxicol. Sci.* 134, 180–194. doi: 10.1093/toxsci/kft094

Uehara, T., Hirode, M., Ono, A., Kiyosawa, N., Omura, K., Shimizu, T., et al. (2008). A toxicogenomics approach for early assessment of potential non-genotoxic hepatocarcinogenicity of chemicals in rats. *Toxicology* 250, 15–26. doi: 10.1016/j.tox.2008.05.013

Uehara, T., Ono, A., Maruyama, T., Kato, I., Yamada, H., Ohno, Y., et al. (2010). The Japanese toxicogenomics project: application of toxicogenomics. *Mol. Nutr. Food Res.* 54, 218–227. doi: 10.1002/mnfr.200900169

Wei, G., Twomey, D., Lamb, J., Schlis, K., Agarwal, J., Stam, R. W., et al. (2006). Gene expression-based chemical genomics identifies rapamycin as a modulator of MCL1 and glucocorticoid resistance. *Cancer Cell* 10, 331–342. doi: 10.1016/j.ccr.2006.09.006

Wignall, J. A., Shapiro, A. J., Wright, F. A., Woodruff, T. J., Chiu, W. A., Guyton, K. Z., et al. (2014). Standardizing benchmark dose calculations to improve science-based decisions in human health assessments. *Environ. Health Perspect.* 122, 499–505. doi: 10.1289/ehp.1307539

Yang, L., Allen, B. C., and Thomas, R. S. (2007). BMDExpress: a software tool for the benchmark dose analyses of genomic data. *BMC Genomics* 8:387. doi: 10.1186/1471-2164-8-387

Yeakley, J. M., Shepard, P. J., Goyena, D. E., VanSteenhouse, H. C., McComb, J. D., and Seligmann, B. E. (2017). A trichostatin A expression signature identified by TempO-Seq targeted whole transcriptome profiling. *PLOS ONE* 12:e0178302. doi: 10.1371/journal.pone.0178302

Zhou, Y. H., Cichocki, J. A., Soldatow, V. Y., Scholl, E., Gallins, P., Jima, D., et al. (2017). Comparative dose-response analysis of liver and kidney transcriptomic effects of trichloroethylene and tetrachloroethylene in B6C3F1 mouse. *Toxicol. Sci.* doi: 10.1093/toxsci/kfx165 [Epub ahead of print].

Zhou, Y. H., and Wright, F. A. (2015). Hypothesis testing at the extremes: fast and robust association for high-throughput data. *Biostatistics* 16, 611–625. doi: 10.1093/biostatistics/kxv007

A Leveraged Signal-to-Noise Ratio (LSTNR) Method to Extract Differentially Expressed Genes and Multivariate Patterns of Expression From Noisy and Low-Replication RNAseq Data

Oswaldo A. Lozoya*, Janine H. Santos and Richard P. Woychik*

Genome Integrity and Structural Biology Laboratory, National Institute of Environmental Health Sciences, National Institutes of Health, Durham, NC, United States

To life scientists, one important feature offered by RNAseq, a next-generation sequencing tool used to estimate changes in gene expression levels, lies in its unprecedented resolution. It can score countable differences in transcript numbers among thousands of genes and between experimental groups, all at once. However, its high cost limits experimental designs to very small sample sizes, usually $N = 3$, which often results in statistically underpowered analysis and poor reproducibility. All these issues are compounded by the presence of experimental noise, which is harder to distinguish from instrumental error when sample sizes are limiting (e.g., small-budget pilot tests), experimental populations exhibit biologically heterogeneous or diffuse expression phenotypes (e.g., patient samples), or when discriminating among transcriptional signatures of closely related experimental conditions (e.g., toxicological modes of action, or MOAs). Here, we present a leveraged signal-to-noise ratio (LSTNR) thresholding method, founded on generalized linear modeling (GLM) of aligned read detection limits to extract differentially expressed genes (DEGs) from noisy low-replication RNAseq data. The LSTNR method uses an agnostic independent filtering strategy to define the dynamic range of detected aggregate read counts per gene, and assigns statistical weights that prioritize genes with better sequencing resolution in differential expression analyses. To assess its performance, we implemented the LSTNR method to analyze three separate datasets: first, using a systematically noisy *in silico* dataset, we demonstrated that LSTNR can extract pre-designed patterns of expression and discriminate between "noise" and "true" differentially expressed pseudogenes at a 100% success rate; then, we illustrated how the LSTNR method can assign patient-derived breast cancer specimens correctly to one out of their four reported molecular subtypes (luminal A, luminal B, Her2-enriched and basal-like); and last, we showed the ability to retrieve five different modes of action (MOA) elicited in livers of rats exposed to three toxicants under three nutritional routes by using the LSTNR method. By combining differential measurements with resolving power

to detect DEGs, the LSTNR method offers an alternative approach to interrogate noisy and low-replication RNAseq datasets, which handles multiple biological conditions at once, and defines benchmarks to validate RNAseq experiments with standard benchtop assays.

INTRODUCTION

At their core, RNAseq and qPCR are two flavors of the same principle: by quantifying how many copies of different molecular templates are accumulated after a discrete number of duplication rounds, it is possible to estimate their abundance in the original sample as long as the amplification has occurred in exponential fashion (Livak and Schmittgen, 2001; Pfaffl, 2001). Viewed under this light, the realization that RNAseq data analysis offers the same challenges qPCR faces is somewhat expectable; nevertheless, discriminating good RNAseq data from bad is all but impossible if following the traditional in-depth inspection of qPCR data quality.

For starters, RNAseq can collect information from thousands of genes in a massively parallel fashion within a single experiment; given the scale of generated data, this means inspection of differential expression levels from each individual gene one at a time is impractical. Also, the discrete form of raw RNAseq output, i.e., countable reads, is of a different nature and statistical behavior than the output of other techniques, such as qPCR and hybridization microarrays, in the form of a digitized continuous-valued signal mass (Roy et al., 2011). Finally, in RNAseq experiments the expected variation in total number of representative reads for each detected transcript within a sample depends, among other factors, on each transcript's size, abundance, and GC-composition. Efforts to control the effects of such sequencing biases during library assembly have been posited early on during the development of RNAseq technologies (Auer and Doerge, 2010; Bullard et al., 2010); at the same time, development of various programmatic strategies to adjust against those biases during statistical analyses has continued ever since (Hansen et al., 2010; Aird et al., 2011; Risso et al., 2011; Benjamini and Speed, 2012; Wu et al., 2013; Law et al., 2014; Finotello and Di Camillo, 2015).

In many pipelines for RNAseq analysis, read outputs are transformed to a normalized measurement of relative expression between two samples, such as fold-change differences. However, performing read normalization can be problematic because it "divides out" the net output of detected reads. Without that information, it is impossible to determine whether observed experimental variation is consistent with the detection capacity of sequencing hardware or not. As a result, the ability to discern between a real signal and instrumental noise is lost. These issues are magnified in experiments with low replicate numbers—a common limitation that RNAseq users face due to costs of these technologies—or when specimens under inspection show highly heterogeneous transcriptional profiles within statistical groups (Hansen et al., 2011; Oberg et al.,

2012; Robles et al., 2012). To this day, financial constraints to acquire and maintain RNAseq instrumentation, combined with complex structure of output data, remain the primary obstacles preventing RNAseq technologies to join clinical diagnostic settings when characterizing transcriptional profiles of patient-derived specimens (Nazarov et al., 2017).

Still, RNAseq remains the most powerful technique to assay gene expression genome-wide. The accuracy in gene expression measurements afforded by RNAseq relies on the volume and diversity of aggregated cDNA fragments, each distinguished from all others by their nucleotide sequences (Cloonan et al., 2008; Mortazavi et al., 2008; Oshlack et al., 2010); conversely, errors in RNAseq measurements are usually ascribed to nucleotide-integration errors during PCR amplification (Finotello and Di Camillo, 2015) because it is assumed that the exquisite sensitivity of detection hardware in next-generation sequencers guarantees accuracy. However, this assumption is not entirely correct: digital detection systems will generate output data from faithfully replicated sequences, trace contaminant templates and misconstrued sequences all the same depending on the quality of base-calling. This means that distinguishing detection noise that stems from PCR bias, instrumental error, contamination or any of them combined is not only difficult, but perhaps irrelevant—all of them happen, and all of them distort gene expression measurements the worst when calculated from low read counts. In other words, while base-calling in RNAseq depends on how precisely nucleotides are detected, measuring expression differences is a resolution problem.

Challenges to resolving expression differences by RNAseq are complicated further when the number of experimental samples is small and rates of detection for differentially expressed genes (DEGs) is poor. In those scenarios, the traditional route is to increase the sequencing depth for each sample with additional rounds of sequencing. Adding sequencing rounds surely increases the number of aligned reads per gene (or coverage); however, no matter how low the rate of misconstrued and contaminant reads that contribute to the total read output may be, they will amount to higher numbers of "phantom reads" undistinguishable from sequencing noise as more sequencing rounds pile up (Tarazona et al., 2011; Li and Tibshirani, 2013). This means that, without a threshold to discriminate the rate of contaminant templates that randomly align to a reference genome, the assumption that aligned reads measure the different cDNA copies in a sample's RNA pool is as "true" as assuming that they represent sequencing artifacts. Determining such a threshold of expectable detected artifacts from RNAseq data calls for a statistical treatment of

raw sequencing output, one that deems collected reads as a combination of faithful and artefactual sequences that align to a reference genome.

Here, we present a leveraged signal-to-noise ratio (LSTNR) thresholding method, founded on generalized linear modeling (GLM) of aligned read detection limits to extract DEGs from count-based and noisy low-replication RNAseq experiments. The LSTNR method uses an agnostic independent filtering strategy to define the dynamic range of detected aggregate read counts per gene that can be explained by the variation in sequencing coverage across genes and experimental replicates. This approach not only determines a minimum read count density that true expressed genes must accrue for reliable detection, but also qualifies genes as more or less reliable for differential expression measurements based on how distant they are from the reliable detection minimum. By taking into account that expression measurements based on read counts have different levels of resolution for different genes, the LSTNR method relies on one fundamental difference between the nature of scale-dependent dispersion in sequencing output (uniquely aligned reads) and the behavior of differential expression estimates *secondary to the original output* (log-transformed relative fold changes between genes or treatment groups): that RNAseq-based estimates of gene expression differences between groups are as robust as the sequencing depth that underlie them.

MATERIALS AND METHODS

Implementation of LSTNR Method

A diagram of the LSTNR method pipeline is depicted in **Figure 1**. Briefly, expression levels of individual genes (Ensembl annotation) were calculated as the normalized rate of deduplicated and uniquely aligned reads per million of total sequenced reads (RPM) overlapping their annotated genomic coordinates (reference genome: hg19). Statistical tests of differential gene expression were performed using a weighed two-way ANOVA model (gene × group blocks) of log_2-transformed fold changes (Log2FC) in RPM relative to gene-wise mean RPM either from all samples (for phenotype profiling experiments) or from a reference group (for treatment vs. control experiments) with $N \geq 3$ replicates per group. Resolution weights of genes corresponded to the cumulative hazard of gene-wise significance scores from two-way ANOVA testing of the linear predictor of RPM from GLM of the natural parameter $\boldsymbol{B(\vartheta)}$ to an exponential continuous-valued distribution with a canonical inverse link function (Nelder and Wedderburn, 1972). Gene-wise significance of Log2FC variation based on weighed ANOVA inference testing were adjusted by the Benjamini–Hochberg method for multiple comparisons (Benjamini and Hochberg, 1995). Further details on the statistical treatment of count-level data under the LSTNR method are available in the Supplementary Methods section. All metrics and statistical analyses were carried out using JMP$^{®}$ 13.0.0 64-bit statistical software (SAS, Cary, NC).

Stratification of Significantly Expressed Genes

Significant genes (SGs) were identified as those with significantly different weighed ANOVA scores (FDR adj. $p < 0.05$); DEGs equaled the subset of SGs with a minimum practical effect size $\delta_{Log2FC} > \delta_{Effect}$ and *post-hoc* pairwise-significant Log2FC differences between at least two groups (Student's t-test $p < 0.05$). As a reproducibility benchmark, we refer to LSTNRs as the subset of SGs in which a minimum practical effect size $\delta_{Log2FC} > \delta_{Effect}$ was detected and at least one group exhibits average Log2FC signal vs. baseline greater than transcriptome-wide measurement noise (or SNR > 1) where noise is defined as the 95% Tolerance Interval (Odeh and Owen, 1980; Hahn and Meeker, 1991; Tamhane and Dunlop, 2000) of gene × group residuals among SGs. Finally, we refer to the subset of genes classified as both DEGs and LSTNRs as DEGs with reproducible expectation estimates (DEGREEs)—i.e., genes with relevant effect size, significant *post-hoc* pairwise differences, and prospective SNR > 1.

Extraction of Candidate Genes for Transcriptional Profiling and Biomarker Analysis

To extract a list of DEGREEs with the highest transcriptional profiling potential and minimal Type II error rates, we calculated within-gene observed effect sizes (Δ_{Log2FC}), as well as estimated retrospective statistical power means ($\ll\pi\gg$) and 95% lower confidence limits (π_{low}). DEGREEs with $\pi > 90\%$ were listed in descending order by Δ_{Log2FC} first, and then by π_{low}. The difference between successive values of Δ_{Log2FC} going down the list of ranked DEGREEs (also known as lag differences) were calculated, with the expectation that genes with higher retrospective statistical power also show larger Δ_{Log2FC} values; when true, this expectation results in a list of negative-valued lagging differences. The minimal subset of characteristic genes, or Profiler DEGREEs, corresponds to the subset of DEGREEs ranked by π_{low} that all show negative Δ_{Log2FC} lag differences before the first positive instance is found.

To obtain a reductive set of biomarker gene candidates, we performed sequential partition tree analysis (without repetition) based on the list of Profiler DEGREEs. The predictive machine-learning power of partitioned data is reported graphically via receiver operating characteristic (ROC) curves specific to each known phenotype. The minimum number of biomarkers needed for partitioning equals the number of phenotypes being partitioned minus one. Partitioning performance was then evaluated by all-at-once discriminant analysis based exclusively on the data from the biomarkers selected by sequential partitioning. Predictive machine-learning power of the canonical multivariate factors of discrimination, estimated using only biomarker data, is reported graphically via ROC curves for each expected phenotype. Phenotype segregation is depicted in multivariate space using 2-component canonical factor plots.
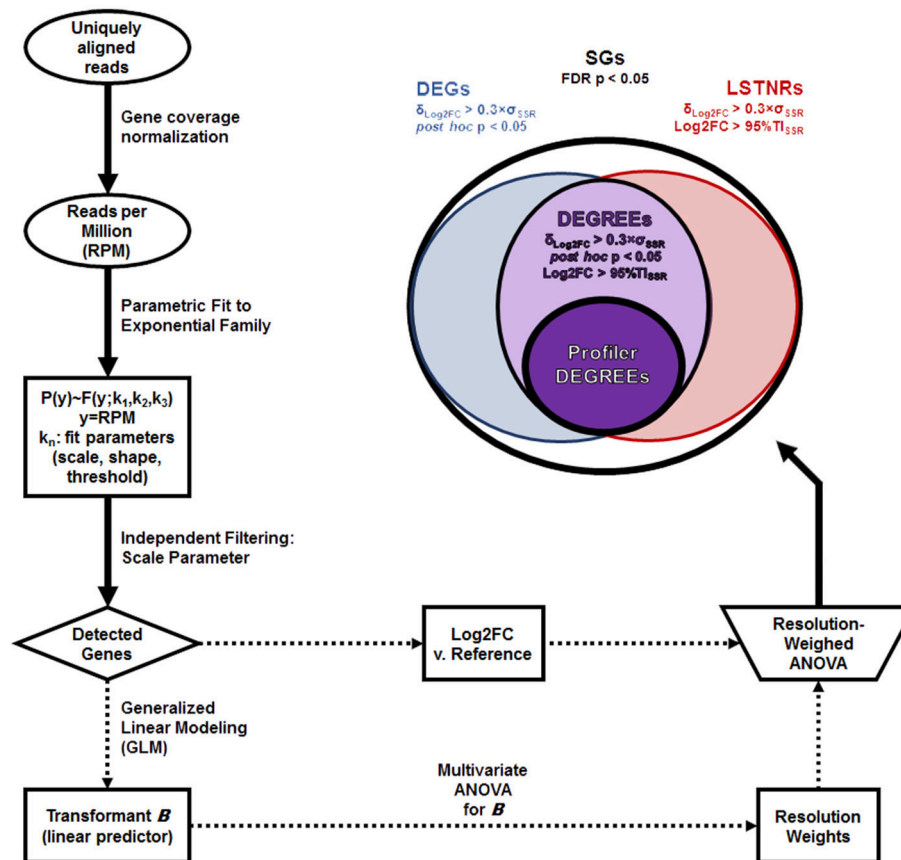
FIGURE 1 | LSTNR method workflow. The schematic depicts the main steps involved in detection of statistically significant genes (SGs) starting from uniquely aligned reads for each gene in individual samples: coverage normalization as reads per million total sequenced reads (RPM); parametric distribution fitting; independent filtering based on fit parameters; generalized linear modeling (GLM), leading to gene resolution weights; log-fold expression measurements (Log2FC) vs. a fixed reference; and two-way resolution-weighed ANOVA of Log2FC. Top right: stratification criteria for SGs and their respective nomenclature: differentially expressed genes (DEGs), leveraged signal-to-noise-ratio genes (LSTNRs), and DEGs with reproducible expectation estimates (DEGREEs). The subset of Profiler DEGREEs corresponds to DEGREEs with a mean retrospective statistical power $\ll\pi\gg > 90\%$ which, when ranked by within-gene observed effect sizes (Δ_{Log2FC}) first and by their 95% lower confidence limit of retrospective statistical power (π_{low}) next, show monotonically decreasing Δ_{Log2FC} values. Profiler DEGREEs can then be used for benchtop validation or for additional statistical analysis to obtain a reductive set of prospective biomarkers through a variety of machine-learning analytics (e.g., partition trees, canonical factor analysis).

Analyzed Datasets

For this work, an *in silico* dataset of simulated RNAseq counts—originally assembled in the development of EPIG-seq (Li and Bushel, 2016)—was used to validate the performance of the LSTNR method. In addition, the ability of the LSTNR method to extract transcriptional signatures and expression patterns from experimental data was tested on two publicly available RNAseq data sets: one for breast cancer primary tumors under 4 breast cancer molecular subtypes deposited in The Cancer Genome Atlas (TCGA) (Cancer Genome Atlas, 2012), and another one from the MAQC phase III SEQC crowd source toxicogenomics (TGxSEQC) effort using livers of male Sprague-Dawley rats after exposure to hepatotoxic agents sharing modes of action (MOA) (Gong et al., 2014; Wang et al., 2014). Further details for each of the three datasets in this study, as well as gene annotation conventions used for each, are outlined in the Supplementary Methods section.

RESULTS

EPIG-Seq Simulated Data

We used a data set of simulated RNAseq counts to validate the performance of the LSTNR method. The simulated data set was originally assembled in the development of EPIG-seq, a similarity scoring methodology for count-based data that catalogs co-expressed genes under patterns of differential expression among multiple conditions (Li and Bushel, 2016). To perform independent filtering, we fit empirically observed RPM averages from each of the 20,000 simulated pseudogenes across the entire 140-pseudoreplicate set to parametric distribution functions. We found the best fit model corresponded to a 3-parameter lognormal distribution (**Figure 2A**), which is equivalent to a normal distribution of RPM in logarithmic scale, with a threshold value γ representing the minimum and positive value of gene-wise average RPM supported by the
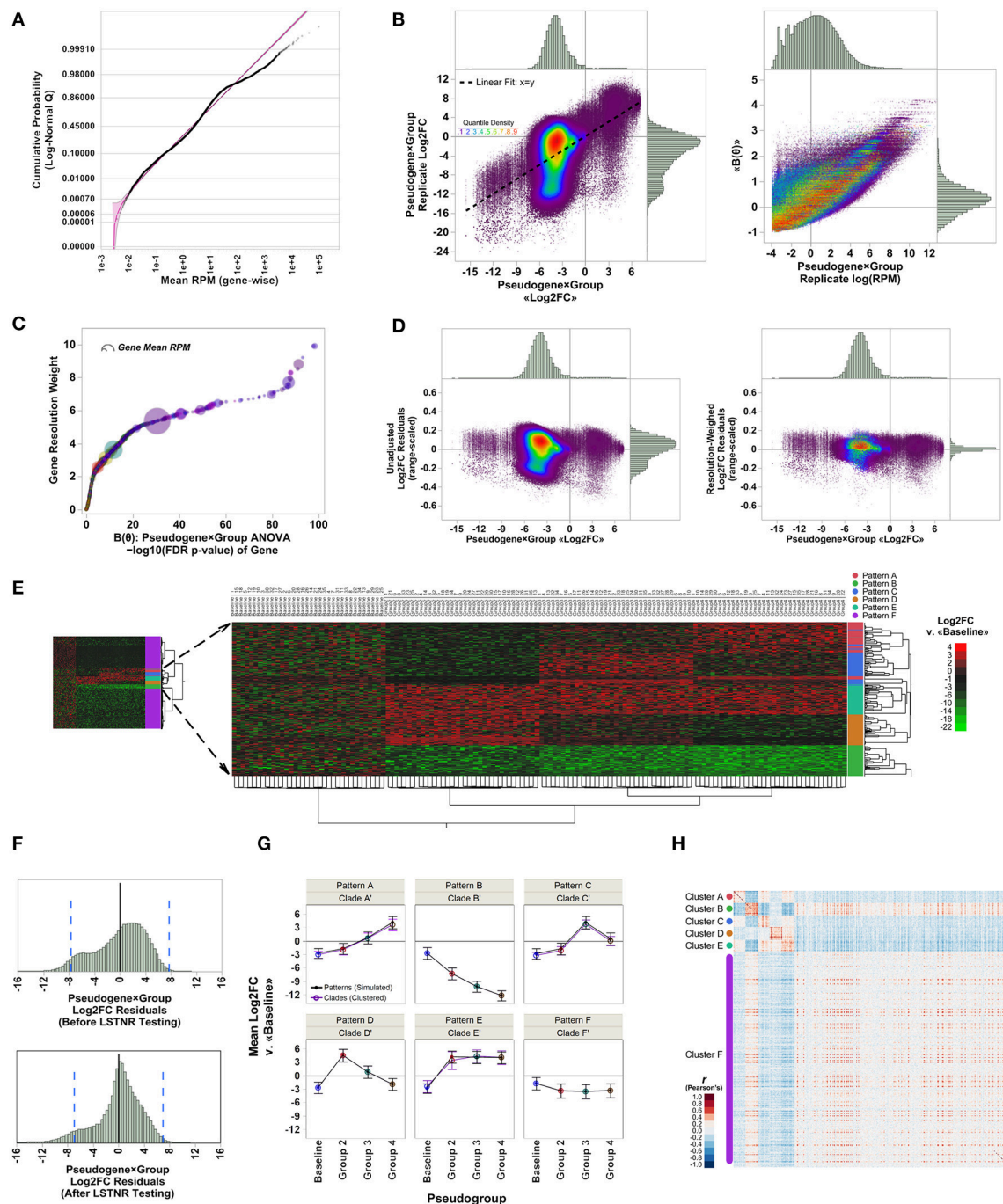
FIGURE 2 | Analysis of EPIG-seq *in silico* test data by the LSTNR method. **(A)** Quantile plot of pseudogene-averaged RPM across pseudoreplicates, overlaid onto their best-fit threshold lognormal distribution parametric model (purple); shading around the parametric fit represents the simultaneous 95% confidence interval of predicted means. **(B)** Linear fitting and distribution of relative expression metrics for pseudogene × group blocks with respect to individual pseudoreplicates. Left: average of Log2FC measurements vs. Log2FC from individual pseudoreplicates; right: values of the canonical link function log (RPM) of pseudoreplicates vs. pseudogene × group block averages of the linear predictor function, $\ll\!\boldsymbol{B}(\theta)\!\gg$, estimated by GLM. Colors of individual data points corresponds to the estimated quantile density of Log2FC values in mean vs. replicate space (left panel) as indicated by the adjacent color scale. **(C)** Gene resolution weights as a function of FDR-adjusted significance levels of pseudogenes as determined by pseudogene × group two-way ANOVA based on the linear predictor $\boldsymbol{B}(\theta)$. Point coloring corresponds to that in B; size of each data point is representative of pseudogene-averaged RPM across pseudoreplicates. **(D)** Distribution of range-scaled residuals around the average of pseudogene × group blocks before (left) and after (right) multiplying Log2FC measurements by gene resolution weights depicted in **(C)**; point coloring corresponds to that in **(B)**. **(E)** Heatmap plot for two-way unsupervised clustering of pseudoreplicates (horizontal) and 4,541 significant pseudogenes (FDR $p < 0.05$) detected by LSTNR analysis based on Log2FC vs. the mean RPM in the baseline group. Left: bird view of the entire heatmap; center: magnification to subset

(Continued)

FIGURE 2 | of 1,000 pseudogenes from five simulated co-expression patterns and their correspondence with inferred hierarchical clades; right: heatmap is colored on a green-black-red gradient scale of Log2FC values relative to baseline (green, downregulated; black, same; red, upregulated). **(F)** Distribution of net residuals around pseudogene×group Log2FC averages, based on 20,000 analyzed pseudogenes (before LSTNR testing) and 4,541 pseudogenes with statistically significant resolution-weighed differences (after LSTNR testing). Dotted blue lines to the left and right of the x-axis origin enclose the predicted 95% tolerance interval of residuals in each respective plot. **(G)** Average Log2FC expression $\pm$ s.d. of pseudogenes in simulated patterns vs. inferred clades of co-expression across groups. **(H)** Heatmap depicts Pearson's correlation coefficient r values between 4,541 statistically significant pseudogenes. Pseudogenes are displayed by groups of inferred co-expression clades; heatmap is colored on a blue-white-red gradient scale of Pearson's correlation coefficient r values (blue, negative correlation; white, not correlated; red, positive correlation). $N = 35$ pseudoreplicates in each of 4 simulated condition groups: baseline, and groups 1 to 3; number of simulated co-expression patterns: five, with 200 pseudogenes each. Simulation comprises 140 total pseudoreplicates with reads distributed among 20,000 total pseudogenes.

fitted distribution. We reasoned the fitted threshold $\gamma = 2.4 \times 10^{-3}$ (95% prediction CI: 0.9×10^{-3}-3.5×10^{-3}) was the best candidate value to use for independent filtering across pseudogenes with simulated reads. In the case of this simulated data set, independent filtering against the threshold parameter γ did not exclude any pseudogenes from subsequent analysis—meaning all listed pseudogenes were "detectable" assuming the estimated error model of simulated read counts was properly fit by a lognormal distribution (**Table 1**).

The next step to implement the LSTNR method was to find a metric that places RPM-values from each of the detected pseudogenes (i.e., that passed independent filtering) in the context of the entire dataset to describe the relative *resolution* in detected RPM-values for each pseudogene. This can be achieved through generalized linear modeling, or GLM (Nelder and Wedderburn, 1972). In the case of the EPIG-seq *in silico* data, which followed a lognormal distribution, rewriting in exponential family form showed the linear predictor $B = \log(\text{RPM})$ (**Figure 2B**). Further algebraic inspection revealed the best model to match averages of lognormally distributed B with their underlying variation is a normal distribution. Thus, transformant values of B did not require a normalizing manipulation (i.e., $\eta = B$) going further down the LSTNR pipeline.

Transformant log(RPM) B-values were tested by two-way ANOVA to estimate transformant significance scores for each pseudogene; after multiple testing adjustment by the false-discovery rate approach (Benjamini and Hochberg, 1995), we found transformant FDR $p < 0.05$ in 9,492 of the 20,000 total pseudogenes, meaning 47.46% of all pseudogenes showed variability in read counts in one or more groups that was statistically discernible (*resolvable*) from that of all pseudoreplicates combined (**Table 1**). We then ranked those transformant significance scores from least to most significant and calculated their position within the ranking; this is the complement of the cumulative density function, also known as the survival function. Finally, to create a metric that gives higher weight to pseudogenes detected with better resolution, we took the negative logarithm of the survival for transformant significance scores function (known as the cumulative hazard rate). This metric is smallest for pseudogenes with low resolution in RPM-values that change little across replicates, and largest for pseudogenes with RPM-values that are either very large, very variable or both (**Figure 2C**). We estimated these metrics and assigned them to their corresponding pseudogenes; later on, we used them as "resolution weights" to account for different resolution levels for pseudogenes based on their net read counts.

TABLE 1 | Step-by-step output as numbers of qualifying genes along the LSTNR analytical pipeline for a validation *in silico* dataset (courtesy of Li and Bushel, 2016; doi: 10.1186/s12864-016-2584-7).

Criteria	*In Silico*: EPIG-seq ($N = 140$)
Simulated pseudogenes	20,000
Distribution of pseudogene-wise RPM means	$P(y) \sim Normal_{3P}(\log[y]; \sigma', \mu', \gamma);$ $y = \text{RPM}$ $\sigma' = 25.4$ $\mu' = 25.4$ $\gamma = 2.5 \times 10^{-3}$ RPM
Independent filtering: pseudogenes with average $y > \gamma$	20,000
Linearized normalizing transformant: GLM Linear Predictor	Log[y]
Transformant two-way ANOVA: resolved pseudogenes across groups with respect to pseudogene-wise mean	9,492
Resolution-weighed ANOVA: significant pseudogenes with FDR adj. $p < 0.05$ based on differences in resolution-weighed RPM log-fold changes (Log2FC) relative to baseline group	4,541

To detect significantly expressed pseudogenes, we first calculated log$_2$-fold changes (Log2FC) relative to the average RPM in the baseline group for each gene; then, we combined Log2FC of genes with their resolution weights in a two-way multivariate ANOVA model. We reasoned that scaling Log2FC relative expression measurements by their resolution metric would also homogenize the scale of variation around the means of individual genes in each experimental statistical group. Indeed, we found that the relative dispersion and homogeneity of Log2FC residuals improved after scaling by gene resolution weights (**Figure 2D**). In all, we detected 4,541 statistically significant pseudogenes by resolution-weighed two-way ANOVA (FDR adj. $p < 0.05$). This pool of statistically significant pseudogenes contained all 1,000 differentially expressed pseudogenes with simulated co-expression patterns, as well as an extra pool of 3,541 pseudogenes from the unpatterned group exhibiting only random noise (**Figure 2E**).

One important aspect to consider in benchtop validation of RNAseq experiments is how well the dispersion in RNAseq output can be projected. This projection is critical to confirm

RNAseq estimates by qPCR, since it helps determine both: (a) how large should RNAseq expression differences be in validation assays to be reliable; and (b) which genes are the most reliable validation candidates based on their RNAseq expression differences. With that principle in mind, we calculated 95% tolerance intervals around the mean $\log_2(\text{RPM})$ measurements of all 20,000 pseudogenes in the EPIG-seq simulated data set that passed independent filtering, as well as among the 4,541 statistically significant pseudogenes identified after resolution-weighed ANOVA. We found $\log_2(\text{RPM})^{95\%\text{TI}} = \pm 7.7$ among all 20,000 pseudogenes, and $\log_2(\text{RPM})^{95\%\text{TI}} = \pm 7.0$ among 4,541 statistically significant ones. In other words, assuming these data were derived from a "true" sample of biological specimens, one could project with 95% confidence that 95% of differences in expression between groups, and for any particular pseudogene, may be ∼200-fold off from their "true" expected value, or ∼130-fold when adjusted for noise among significant pseudogenes, due only to experimental variability between replicate experiments (**Figure 2F**).

The unpatterned pseudogenes with statistically significant expression levels detected through LSTNR could aggregate to excessive levels of "background noise." This "noise" could be detrimental to statistical clustering or discriminant analyses, and may undermine the capacity to extract "true" expression patterns. To address this point, we performed naïve hierarchical clustering (Ward's method) and found that, even though unpatterned pseudogenes accounted for most of the detected differential pseudogenes overall, all differentially expressed pseudogenes were agglomerated in the hierarchical tree under simulated patterns. We also found that the 3,541 unpatterned but statistically significant pseudogenes were segregated apart from the 1,000 patterned ones (**Figure 2E**). Furthermore, 927 out of the 1,000 patterned differential pseudogenes were assigned into five well-separated clades of expression trends that matched the co-expression patterns prescribed *in silico* (contingency analysis Pearson's *p* < 0.0001). Altogether, the LSTNR method detected statistically significant pseudogenes belonging to both patterned and unpatterned expression trends, but did not compromise the ability to discriminate between both kinds of pseudogenes, nor their correct expression patterns, by standard clustering analyses.

Still, any benefit of implementing the LSTNR workflow is only substantial if it matches or surpasses the performance of already existing methodologies to tease out coordinated patterns of differential gene expression, such as EPIG (a pipeline tailored for microarray data) and EPIG-seq (a modified version of EPIG to handle count-based data) (Chou et al., 2007; Li and Bushel, 2016). To address this point, we assembled the confusion matrix of pseudo-gene cluster assignments per LSTNR analyses of the *in silico* dataset, and compared to those obtained by EPIG and EPIG-seq analyses (courtesy of Li and Bushel, 2016) in terms of the sensitivity (the true positive rate) and specificity (the true negative rate) of inferred cluster membership among differential pseudo-genes detected per platform (**Table 2**). We found the LSTNR method showed >94.5% specificity rates across simulated clusters, much like those from EPIG and EPIG-seq, thus indicating that pseudo-genes identified as differentially expressed are rarely misclassified

under their originally prescribed expression patterns regardless of the chosen methodology. However, LSTNR detected more differential pseudo-genes than either EPIG or EPIG-seq, as shown by improved sensitivity rates per cluster when using LSTNR (77.5–100%) vs. either EPIG (17.5–68%) or EPIG-seq (55.5–84.5%). Put together, these results indicate that the LSTNR method successfully extracts more DEGs, all while grouping them by their true underlying expression patterns at the same or improved rates, than other similar pipelines.

Among the 927 differential pseudogenes correctly assigned to simulated patterns of expression detected by the LSTNR method, 400 were classified into clades that matched their *in silico* counterparts in full (patterns B and D); in contrast, the EPIG-seq method matched pseudogenes completely to their simulated pattern only for B. The five hierarchical clades identified by LSTNR also matched their respective *in silico* patterns in terms of average expression levels within statistical groups (**Figure 2G**). We found similar pattern-detection performance among the 4,541 statistically significant pseudogenes detected with the LSTNR method when clustering them by the Pearson product-moment correlation scores of their Log2FC vs. the baseline average (**Figure 2H**). In sum, the LSTNR method not only identified the five simulated expression patterns just like the EPIG-seq pipeline, but did so by assigning differential pseudogenes to their correct patterns with greater accuracy, by routing read count data to traditionally robust statistical tests (e.g., multivariate ANOVA), and without relying on user-defined parameters.

Breast Cancer Classification by Hierarchical Clustering of LSTNR-Detected DEGs

Next, we used RNAseq data from breast ductal carcinoma biopsies deposited in The Cancer Genome Atlas (TCGA) to test the LSTNR method. The data set collected for this assessment consisted of 160 primary tumors classified under 4 breast cancer molecular subtypes groups in equal sample sizes (luminal A, luminal B, Her2-enriched and basal-like) and 40 matching normal breast tissue biopsies as controls (Cancer Genome Atlas, 2012). Besides consisting of patient-derived data that is clinically relevant, this collection of TCGA specimens presents some practical and common challenges associated with clinical data, including high variability between patients and between cohorts from an epidemiological perspective, and batch effects from a technical one. All these challenges, which severely undermine significance testing, have clinical consequences: if the only solution to understand the performance of RNAseq as a diagnostic tool is to design large patient cohorts, it risks rendering RNAseq too expensive or slow to be useful in a clinical setting. Here, our goal was to implement the LSTNR method and define how successful it was in sorting out the correct breast cancer molecular subtypes (and their transcriptional signatures) based on a much smaller cohort of patient-derived specimens than the TCGA database itself.

We analyzed TCGA dataset following two approaches: (a) we split the specimens from each of the four molecular subtypes

TABLE 2 | Confusion matrices for differential pseudo-gene pattern assignment by EPIG, EPIG-seq, and LSTNR implementation based on a validation *in silico* dataset (EPIG and EPIG-seq: courtesy of Li and Bushel, 2016; doi: 10.1186/s12864-016-2584-7.

| Method	Clustered Clades	Simulated Expression Pattern					Total	Sensitivity (%)	Specificity (%)
		Pattern A	Pattern B	Pattern C	Pattern D	Pattern E			
EPIG	Group A	0	0	46	0	0	46	23	100
	Group B	0	0	0	0	136	136	68	100
	Group C	0	39	0	0	0	39	19.5	100
	Group D	0	0	0	60	0	60	30	100
	Group E	35	0	0	0	0	35	17.5	100
EPIG-Seq	Group A	7	132	11	0	0	150	66	98.2
	Group B	0	0	169	0	0	169	84.5	100
	Group C	135	26	5	0	0	166	67.5	96.9
	Group D	0	0	0	15	166	181	83	98.5
	Group E	9	23	0	111	43	186	55.5	92.5
LSTNR	Clade A'	177	0	12	0	4	193	88.5	98.4
	Clade B'	0	200	0	0	0	200	100	100
	Clade C'	1	0	155	0	1	157	77.5	99.8
	Clade D'	0	0	0	200	0	200	100	100
	Clade E'	22	0	33	0	195	250	97.5	94.5

into four independent and mutually exclusive sample subsets (or *realizations*) of equal size, performed parallel statistical analyses for each, and identified the subset of DEGs detected in common by all four separate per-realization analyses; and (b) we analyzed the entire dataset all at once in a single run. Altogether, this strategy comprised a total of 5 separate implementations of the LSTNR method on patient-derived breast cancer data: four per-realization analyses and one single-shot analysis. By following this analytical design, we sought to establish the degree of observed concordance between DEGs detected by a single-shot analysis vs. the consensus from multiple tests on subsampled statistical groups. We also aimed at determining whether capturing more DEGs with an all-encompassing single-shot analysis is superior in quality and performance to performing split tests in tandem and extracting a consensus DEG list.

A total of 20,532 annotated genes (hg19) with uniquely aligned reads were represented in any one of the specimens used to test the LSTNR method. Empirically observed RPM averages within realizations were fit to parametric distribution functions independently. In all cases, the best-fit model for within-gene RPM averages corresponded to a 3-parameter Weibull distribution function $P(y) \sim \text{Weibull}_{3P}(y;\alpha,\text{ß},\gamma)$ where $y = \text{RPM}$ (**Figure 3A**); the estimates for the threshold parameter γ in each separate realization ranged between 1.5×10^{-3}-11×10^{-3} RPM (**Table 3**). The threshold parameter γ, representing the minimum average RPM-value explained by the parametric Weibull fit, was added across the dataset to circumvent arithmetic issues with zero-valued data when estimating relative expression ratios.

It is well-known that, for GLM, Weibull distributions with fixed shape parameter ß can be restated by simple algebraic substitution in the form of an exponential distribution, for which the linear predictor $B \sim (y-\gamma)^{\text{ß}-1}$ and the normalizing link function $\eta \sim 1/B$ (Nelder and Wedderburn, 1972; Aitkin and Clayton, 1980). Therefore, to determine the resolution level of detected reads per gene, we calculated transformant **B**-values from within-gene RPM averages, normalized their distribution by using their reciprocal values, and tested them by multivariate ANOVA (see Supplementary Methods for details). Analysis of the entire dataset at once also identified, with parameters almost equal to the ones estimated from realizations, a 3-Parameter Weibull distribution as the best parametric fit (**Table 3**).

The number of retained genes following independent filtering with respect to the Weibull scale parameter α (**Table 3**) was about the same in all five analyses (8,005–8,562 across realizations; 8,538 single-shot), yet the estimated number of significantly resolved genes, defined by genes with transformant FDR $p < 0.05$, was sensitive to the realization they came from. The number of resolved genes varied by up to one order of magnitude (e.g., 381 genes in realization 2 and 4,295 genes in realization 1). Even though the number of resolved genes estimated by single-shot analyses (2,851 total with transformant FDR $p < 0.05$) fell within the order of magnitude resolved genes by the other four per-realization analyses, it was still over 50% larger than the average of 1,899 resolved genes per realization (**Table 3**). These findings illustrate that the independent filtering criteria used to exclude poorly represented genes were equally valid whether they were calculated from a subset or from the entire set of individual specimens in an experiment. In contrast, estimates of the dynamic range of sequencing representation are off-target when using different data subsets. Therefore, expecting independent RNAseq experiments to detect the same pool of significantly resolved genes is sensible only when the replicate sets in one experiment are statistically indistinguishable from the replicate
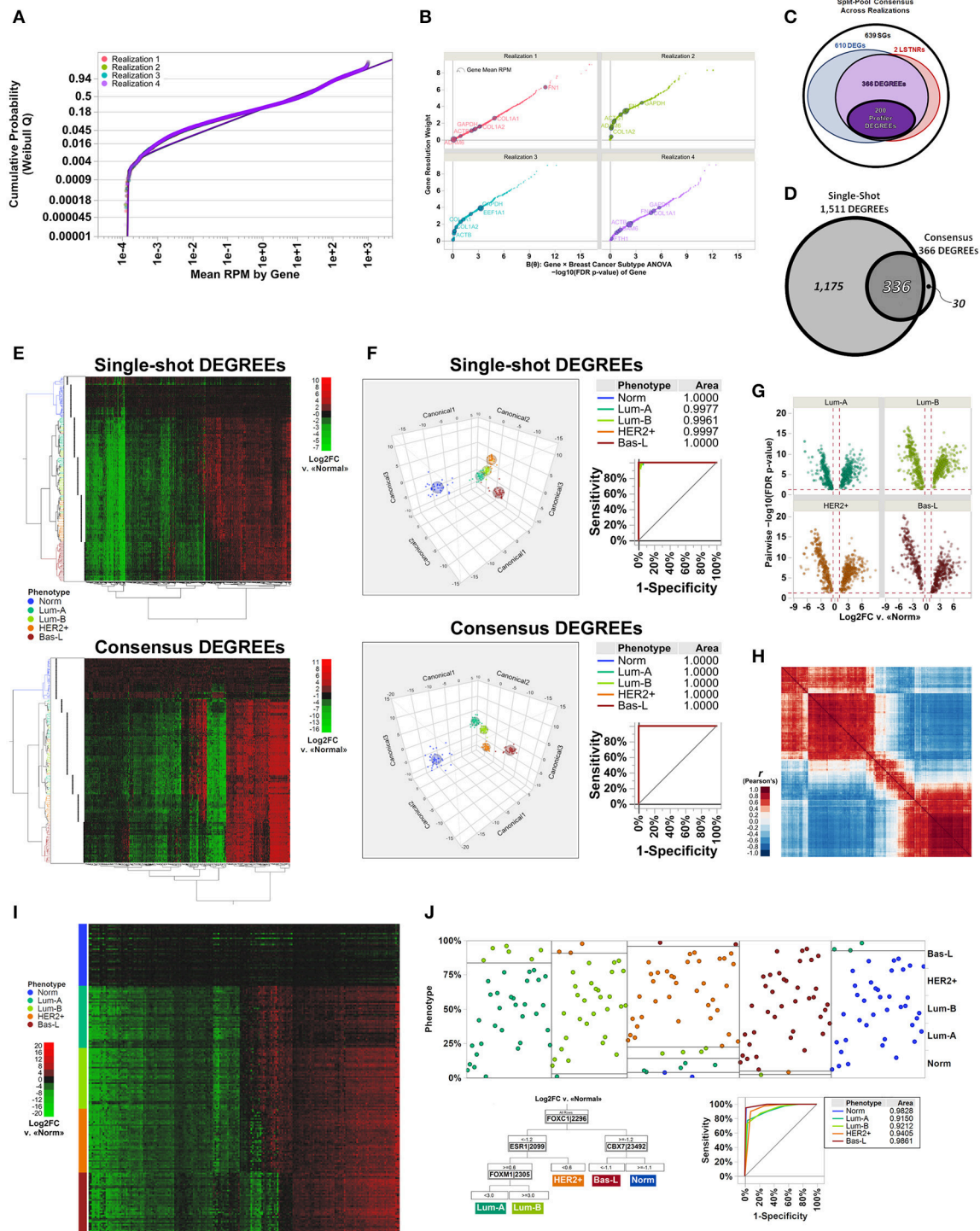
FIGURE 3 | Molecular subtype discrimination of transcriptional signatures from patient-derived breast cancer specimens using the LSTNR method by split-pool and single-shot approaches. **(A)** Quantile plot of gene-averaged RPM across replicates, overlaid onto best-fit threshold Weibull distribution parametric models per individual realization; fit lines and points are colored by the realization they were tested under. **(B)** Gene resolution weights calculated within each realization, as a function of FDR-adjusted significance levels of genes, based on linear predictor $\boldsymbol{B}(\boldsymbol{\theta})$ gene × group two-way ANOVA. Point coloring indicates realization tested; size of each data point is representative of gene-averaged RPM across replicates within each realization; also, the top 10 genes found in all realizations with highest RPM scores, and their identities, are shown with dark outlines. **(C)** Venn diagram depicting the number of shared gene symbols among consensus SGs, DEGs, LSTNRs, and DEGREEs identified in all realizations. **(D)** Venn diagram depicting the concordance between DEGREEs detected by all-at-once single-shot analysis of the TCGA

(Continued)

FIGURE 3 | dataset vs. the set of consensus DEGREEs identified across separate realizations. **(E)** Heatmap plots for two-way unsupervised clustering of all specimens (horizontal) and DEGREEs detected by single-shot analysis (top) or consensus across realizations (bottom) based on differential expression levels. Coloring of row dendrograms and labels represent the phenotype as annotated in TCGA database for each specimen. Dot columns on the left of heatmap plots depict inferred specimen clades via unsupervised clustering of depicted genes in each heatmap separately. Right: heatmaps are colored on green-black-red gradient scales of Log2FC values relative to baseline (green, downregulated; black, same; red, upregulated). **(F)** Three-factor discriminant analysis and plots of tested specimens based on Log2FC measurements of single-shot DEGREEs (top) vs. consensus DEGREEs (bottom). Coloring of lines and points in depicted plots represent the phenotype as annotated in TCGA database for each specimen. ROC curves and their AUC values are also shown. **(G)** Volcano plots of 200 Profiler DEGs across all breast cancer subtype specimens relative to normal group; y-axis: *post-hoc* pairwise significance scores of Log2FC measurements; x-axis: mean Log2FC vs. mean RPM of normal specimens. **(H)** Heatmap depicts Pearson's correlation coefficient *r*-values between 200 Profiler DEGREEs. Order of Profiler DEGREEs corresponds to their arrangement by unsupervised clustering, as shown in **(E)**. Heatmap is colored on a blue-white-red gradient scale of Pearson's correlation coefficient *r*-values (blue, negative correlation; white, not correlated; red: positive correlation). **(I)** Heatmap plots of all breast cancer specimens sorted by phenotype and replicate number (rows) vs. Profiler DEGREEs detected by consensus across realizations (columns) based on differential expression levels. Order of Profiler DEGREEs corresponds to their arrangement by unsupervised clustering, as shown in **(E,H)**. Coloring of row labels represent the phenotype as annotated in TCGA database for each specimen. Right: heatmap is colored on a green-black-red gradient scale of Log2FC values relative to normal specimens (green, downregulated; black, same; red, upregulated). **(J)** Inferred diagnostic biomarkers of breast cancer subtypes by sequential partitioning tree analysis of Profiler DEGREEs. Coloring of lines and points in depicted plots represent the phenotype as annotated in TCGA database for each specimen. Partition analysis ROC curves and their AUC values are also shown. $N = 40$ independent specimens per breast cancer molecular subtype: normal, luminal A, luminal B, HER2+, and basal-like. Specimen classification as annotated in TCGA database. TCGA dataset comprises 200 total individual specimens with reads aligned onto 20,532 genes overall, analyzed as a whole (single-shot) or evenly split into four separate and mutually exclusive realizations (split-pool). Reference genome: hg19.

sets in another—an expectation that is grossly misguided among patient-derived specimens.

Significance scores of estimated RPM transformants are only as good as the samples they reflect, and will change depending on whether variation within statistical groups is more or less heterogeneous for different subsets. We premised that each of the subsets is representative of the entire set, and that ranking of the significance scores of genes should be similar across subsets no matter their actual *p*-values; that was the case when we calculated resolution weights of genes per realization (**Figure 3B**). Interestingly, we found that many genes at the high-end of the resolution weight curve tended to show high read count numbers, yet many genes with high RPM averages showed poor significance scores for their transformant ***B***-values. These results suggested that average RPM of genes, though likely to capture finer differences, were not reliable predictors of gene expression resolution.

We then performed resolution-weighed multivariate ANOVA tests of Log2FC differences relative to the average of normal tissues, either per realization or single-shot. We found that, by including resolution scaling, the number of detected significant genes (SGs) became more consistent: between 4,465 and 5,086 SGs from the four independently analyzed realizations, and 6,193 SGs by the single-shot analysis (**Table 3**; see Table S1 for gene symbols). These results were striking because the resolution scaling strategy of the LSTNR method effectively stabilized the number of detected SGs across all five separate tests, even though the number of genes with significant transformant variation (FDR $p < 0.05$) were quite different in each of these separate analyses (as mentioned earlier).

In all, we found a common pool of 1,617 SGs across all four realizations (**Table 3**; see Table S1 for gene symbols), corresponding to 32–36% concordance rates with respect to the total SGs in each one separately (range of SGs per realization: between 4,465 and 5,086). Those observed concordance rates carried down to increasingly stringent levels of stratified practical significance across separate realization tests (see section Materials and Methods and Supplemental

Materials for further details): from the pool of 1,617 common SGs from per-realization analyses, we detected 976 DEGs (vs. number of DEGs between 3,377 and 3,736 in separate realizations; 26–29% concordance rates), and 368 LSTNRs (vs. number of LSTNRs between 1,102 and 1,370 in separate realizations; 27–33% concordance rates). Overall, between DEGs and LSTNRs, we found an intersecting set of 366 genes (37% of DEGs), which we termed DEGs with reproducible expectation estimates (DEGREEs) because they: (a) were statistically significant when adjusted for sequencing resolution; (b) showed Log2FC variation between groups greater than a reference 5% practical effect size; (c) exhibited *post-hoc* pairwise significant differences in Log2FC between groups; and (d) exhibited Log2FC differences with SNR > 1 relative to transcriptome-wide Log2FC measurement error (**Figure 3C**). Likewise, refining the pool of 6,193 single-shot SGs with additional stringency resulted in 6,093 DEGs and, among them, 1,511 LSTNRs all contained within the DEG pool. Hence, all 1,511 LSTNRs (or 25% of DEGs) were also DEGREEs (**Table 3**; see Table S1 for gene symbols).

Next, we interrogated whether detected SGs were similar between per-realization and single-shot analyses at different levels of statistical stringency. Of the 1,617 common SGs across realizations, 1,509 (over 93%) were also among the 6,193 SGs detected in the single-shot analysis using the entire data set. This means that about three out of four (75.6%) of all SGs identified by single-shot testing of the entire dataset are not replicated in parallel tests of mutually exclusive subsets of samples from the same experiment. We then evaluated the concordance rates between intersecting genes across per-realization vs. single-shot analyses, and detected final overlaps of 908 DEGs (93% concordance w.r.t. common per-realization pools; 15% concordance w.r.t. single-shot pool), 337 LSTNRs (92% concordance w.r.t. common per-realization pools; 22% concordance w.r.t. single-shot pool), and 336 DEGREEs overall (92% concordance w.r.t. common per-realization pools; 22% concordance w.r.t. single-shot pool) (**Figure 3D** and **Table 3**; see Table S1 for gene symbols).

TABLE 3 | Step-by-step output as numbers of qualifying genes along the LSTNR analytical pipeline for RNAseq data deposited in TCGA from four realizations of patient-derived breast cancer transcriptomes across four molecular subtypes (courtesy of Li and Bushel, 2016; doi: 10.1186/s12864-016-2584-7).

Criteria	Breast Cancer Molecular Subtypes (TCGA) (*N* = 200)				
	Realization 1 (*N* = 50)	Realization 2 (*N* = 50)	Realization 3 (*N* = 50)	Realization 4 (*N* = 50)	All Specimens (*N* = 200)
Genes with uniquely aligned reads	20,532				
Distribution of gene-wise RPM means	$P(y) \sim Weibull_{3P}(y;\alpha,\beta,\gamma)$; y = RPM				
	$\alpha = 25.4$ RPM $\beta = 0.53$ $\gamma = 9.9 \times 10^{-3}$ RPM	$\alpha = 24.1$ RPM $\beta = 0.53$ $\gamma = 6.6 \times 10^{-3}$ RPM	$\alpha = 25.0$ RPM $\beta = 0.54$ $\gamma = 1.1 \times 10^{-2}$ RPM	$\alpha = 21.9$ RPM $\beta = 0.49$ $\gamma = 1.6 \times 10^{-3}$ RPM	$\alpha = 22.2$ RPM $\beta = 0.49$ $\gamma = 1.5 \times 10^{-3}$ RPM
Independent filtering: Genes with average $y > \alpha$	*8,005*	*8,110*	*8,083*	*8,562*	*8,538*
Linearized normalizing transformant: GLM Linear Predictor	$(y-\gamma)^{\beta-1}$				
Transformant two-way ANOVA: resolved genes across groups with respect to gene-wise mean	*4,295*	*381*	*638*	*2,281*	*2,851*
Resolution-Weighed ANOVA: Significant Genes (SGs) with FDR adj. $p < 0.05$ based on differences in resolution-weighed RPM log-fold changes (Log2FC) relative to baseline condition	*4,465*	*5,086*	*4,537*	*4,618*	*6193*
			Altogether: 7,749		**Final Overlap: 1,509**
			Intersection: 1,617		
Differential expression: DEGs = subset of SGs that exhibit both: • resolution-weighed effect size above 5% of gene-wise variation ($\delta_{Log2FC} > 0.3 \times \sigma_{SSR}$); and • *post-hoc* pairwise-significant Log2FC differences between at least two groups (Student's t-test $p < 0.05$)	*3,736*	*3,377*	*3,497*	*3,617*	*6,093*
			Altogether: 6,407		**Final Overlap: 908**
			Intersection: 976		
Reproducibility: LSTNRs = subset of SGs that exhibit both: • resolution-weighed effect size above 5% of gene-wise variation ($\delta_{Log2FC} > 0.3 \times \sigma_{SSR}$); and • at least one group with Log2FC differences vs. baseline greater than 95% Tolerance Interval of gene×group residuals among SGs (*post-hoc* pairwise-significance not required)	*1,370*	*1,102*	*1,130*	*1,210*	*1,511*
			Altogether: 2,193		**Final Overlap: 337**
			Intersection: 368		
Expectable DEGs: DEGREEs = Ensembl-annotated DEGs with a reproducible expectation estimate (i.e., DEGs that are also LSTNRs) and official Entrez symbol			*Intersection: 366*		*1,511*
					Final Overlap: 336

(Continued)

TABLE 3 | Continued

Criteria	Breast Cancer Molecular Subtypes (TCGA) (N = 200)				
	Realization 1 (N = 50)	Realization 2 (N = 50)	Realization 3 (N = 50)	Realization 4 (N = 50)	All Specimens (N = 200)
Transcriptional profiling: Profiler DEGREEs = top DEGREEs ranked by retrospective statistical power with monotonically decreasing within-gene effect sizes Δ_{Log2FC}	200 Profiler DEGREEs (consensus)				
Diagnostic targets: Biomarkers = minimal subset of Profiler DEGREEs with predictive discriminant power based on sequential partition tree analysis (ROC scores>0.9 per phenotype)	CBX7, ESR1, FOXC1, and FOXM1				

We also asked to what extent using DEGREEs as a gene set of choice for transcriptome-based phenotype segregation was adequate, and whether using a consensus minimal set was superior than using a larger set of DEGREEs extracted from a single-shot differential expression analysis. To do this, we performed two-way unsupervised hierarchical clustering (Ward method) of the entire set of 200 specimens based on Log2FC expression differences vs. normal breast biopsies with: (a) 366 consensus DEGREEs across all per-realization analyses; and (b) the 1,511 DEGREEs identified from the single-shot analyses. We found that unguided sorting of specimens in hierarchical clusters was in agreement with the known molecular classification of the specimens, irrespective of the set of DEGREEs used (**Figure 3E**). An alternative test, discriminant analysis, also showed that separation of specimens in multifactorial space was highly predictive of the correct phenotype. Surprisingly, the same tests also revealed that discrimination and predictive power both were superior when using the consensus set of 366 DEGREEs (0% misclassified specimens) instead of the larger set of 1,511 DEGREEs from single-shot analysis (3.5% misclassified specimens) (**Figure 3F**). This outcome suggests that the LSTNR method not only extracted the same phenotype groups by splitting specimens into parallel data subsets and extracting a consensus DEG set rather than using the entire cohort at once, but profiled them successfully using far fewer genes and specimens, and with greater accuracy, than ever reported (Perou et al., 2000; Sørlie et al., 2001; Cancer Genome Atlas, 2012).

The objective of collecting clinical data from large cohorts, as those in TCGA, is to determine what are the most reliable diagnostic signatures that distinguish closely related diseases. Thus, we tested the ability to discriminate breast cancer subtypes when using the smallest possible set of genes with reproducible expression differences that we detected by LSTNR. To that end, we selected a minimal set of 200 consensus DEGREEs that exhibited, simultaneously, the largest within-gene retrospective statistical power ($\ll\pi\gg$ >90%) and within-gene effect size Δ_{Log2FC}; we termed these Profiler DEGREEs (**Table 3**; see Table S1 for gene symbols). We observed >1.66-fold statistically

significant *post-hoc* pairwise absolute differences ($p < 0.05$) in at least one of the four breast cancer subtypes vs. normal tissue biopsies in each of the 200 Profiler DEGREEs (**Figure 3G**), which also showed highly correlative associations (**Figure 3H**). The increased statistical strength of Profiler DEGREEs over all other sequenced genes was sufficient to distinguish different expression patterns among the 200 breast tissue specimens by ordering them by subtype (**Figure 3I**).

Finally, to extract a minimal set of diagnostic biomarkers, we performed sequential partition tree analysis and defined what subset of genes among the 200 Profiler DEGREEs had the highest discriminative power. We found that breast cancer subtypes could be assigned with >90% diagnostic accuracy, as indicated by the area under the ROC curve of the partition tree, by differential expression analysis with respect to a normal breast tissue reference using only 4 genes: CBX7, ESR1, FOXC1, and FOXM1 (**Figure 3J**). Of the four biomarkers we detected, ESR1 is the only one in common with the reported signature for luminal breast cancers (ESR1, GATA3, FOXA1, XBP1, and cMYB) (Cancer Genome Atlas, 2012). Notably, only ESR1, GATA3, and XBP1 are also present in our list of Profiler DEGREEs. The fact that we detected closely related transcription factors other than FOXA1 and cMYB specifically, i.e., four members from the FOX family (FOXC1, FOXM1, FOXN3, FOXO1) and one from the MYB family (MYBL2), lends confidence to our analysis, which captured the same underlying biological mechanisms that distinguish each breast cancer subtype. However, our analysis achieved the same results with a considerably smaller cohort (N = 200) than the one reported previously by the TCGA consortium (N = 825) (Cancer Genome Atlas, 2012).

In all, these results confirm that single-shot analyses of differential gene expression yields higher numbers of detected significant differences as sample size grows. However, they also suggest that statistical testing of excessively large sample sizes at once comes at the expense of experimental reproducibility, regardless of statistical stringency. Instead, data splitting into modestly sized subsets of samples for parallel statistical analysis can extract a minimal set of representative DEGREEs shared

among parallel per-realization tests. We showed that the resulting consensus DEGREE set shows a higher chance of reproducibility. Furthermore, split data processing allows for multi-threaded computation, which in principle accelerates data analysis speed and performs at a much more accessible computational footprint than a single-shot analysis.

Most importantly, a minimal consensus set of DEGREEs deduced by parallel subsampled analyses is not only more statistically powerful in theory, but also more useful in practice. Because any inferred biomarkers need to be experimentally validated, diagnosing between disease phenotypes by gene expression assays is more amenable and affordable at the bench when the number of targets is kept at their fewest; it is also less prone to practitioner's mistakes, and easier to translate to clinical practice. In the case of breast cancer subtypes, the LSTNR method excelled in extracting a theoretical minimum number of diagnostic biomarkers (a total of 4) necessary to discriminate among 4 predetermined subtypes. In practice, diagnostic testing can be carried out to validate findings, to confirm observations by other researchers, or for clinical purposes where sample numbers are limiting. Our data indicate using the 4 biomarkers detected by LSTNR to diagnose breast cancer subtypes would be at least as reproducible as using 50-gene subtype predictor microarrays reported previously (Parker et al., 2009)—only much cheaper.

Discrimination of Hepatotoxic MOAs Following Chemical Exposure in Rats

The LSTNR method, like many others, can discriminate among disease conditions with known molecular signatures (such as breast cancer) in spite of the heterogeneity in the gene expression profiles among patients because the transcriptional signatures are often overt. In the case of specimens deposited in TCGA, the evidence supporting the classification of each specimens is a combination of both histopathological criteria and transcriptional profiling, an approach pioneered almost two decades ago and refined since by Perou and others based on microarray data (Perou et al., 2000; Sørlie et al., 2001; Parker et al., 2009). This means that, although successful in phenotyping breast cancer subtypes, LSTNR implementation in the context of breast cancer subtypes is somewhat recursive.

The same cannot be said when characterizing transcriptional signatures from data generated by toxicogenomics experiments, in which healthy specimens are exposed transiently to chemical insults and assayed soon after to understand which genes respond to the exposure, and to what extent their response is coordinate or follows particular signaling pathways (also known as the mode of action, MOA). Usually, these studies use large sample sizes, but that is often the case because specimens tested for the same MOA are often exposed to more than one eliciting chemical agents, each with minimal sample sizes. In the end, this means most toxicogenomics datasets cannot be split into realization subsets, and so they must be inspected with single-shot analyses. For all these reasons, identifying MOA from toxicogenomics studies can be more challenging than transcriptional profiling of disease phenotypes.

To assess the performance of LSTNR for toxicogenomics analyses, we used a MOA training RNAseq dataset generated through the MAQC phase III SEQC crowdsource toxicogenomics effort (TGxSEQC). The TGxSEQC training dataset consists of liver transcriptomes from male Sprague-Dawley rats following exposure to hepatotoxicants (Gong et al., 2014; Wang et al., 2014). The experimental designed included five known modes of action, each induced by exposure to three different chemical agents unique to each MOA, with a replicate size of $N = 3$ per combined MOA × Agent group. In all, the experimental design comprised 45 individual specimens equally split among 15 different agents and stratified under 5 MOAs with $N = 9$. Because the purpose of this particular TGxSEQC crowdsourced experiment was to create a benchmark training set to attest performance of newly developed statistical pipelines, additional levels of complexity were introduced in the experimental design by controlling their exposure scheme: some of the chemical agents were supplied by intraperitoneal injections, and others by oral gavage using either nutritive (corn oil) or non-nutritive (water) delivery vehicles. In addition to the 45 agent-exposed specimens, an additional set of 9 livers was also collected as a control group from male rats that underwent mock treatments by different combinations of exposure schemes (i.e., intraperitoneal vs. oral gavage, and corn oil vs. water as oral vehicles).

The TGxSEQC dataset consisted of 30,852 RefSeq-annotated transcripts (rn6 reference genome) uniquely aligned reads, which we refer to as "genes" hereafter for simplicity. The best-fit model to the average RPM-values within annotated transcripts across all 54 individual liver specimens was a 3-parameter Weibull distribution function $P(y) \sim Weibull_{3P}(y;\alpha,\beta,\gamma)$ where $y = RPM$. We then performed independent filtering with respect to the Weibull scale parameter $\alpha = 6.7$ RPM and retained 9,593 transcripts for differential expression analysis (**Table 4**). Next, to determine the resolution level of detected reads per gene, we calculated transformant $\boldsymbol{B}$-values from within-gene RPM averages by GLM with linear predictor $\boldsymbol{B} \sim (y-\gamma)^{\beta-1}$ and normalizing link function $\eta \sim 1/\boldsymbol{B}$ as appropriate for fixed-ß Weibull distributions (Nelder and Wedderburn, 1972; Aitkin and Clayton, 1980). Then, transformant $\boldsymbol{B}$-values were tested by gene×MOA multivariate ANOVA, i.e., irrespective of the chemical agent, route, or nutritional status of the vehicle of each specimen. We identified 3,975 significantly resolved genes, based on a transformant $\boldsymbol{B}$ FDR $p < 0.05$ criterion (**Table 4**). Gene resolution weights were estimated from the cumulative hazard rate of the gene × MOA multivariate ANOVA significance scores of $\boldsymbol{B}$. Also, we added the threshold parameter $\gamma = 2.5 \times 10^{-3}$ RPM to each individual replicate across all transcripts, and estimated Log2FC differences relative to the average of mock-treated controls. Then, we performed a resolution-weighed gene × MOA multivariate ANOVA of Log2FC differences. We detected a total 5,983 SGs; among them, we found 5,864 DEGs and, of those, 1,953 were also LSTNRs. The list of LSTNRs included 386 non-protein encoding transcripts based on their RefSeq annotation (i.e., XR_, XM_, or NR_ accession prefix), which we discarded from subsequent analysis. The remaining 1,567 LSTNRs corresponding to mRNA transcripts

TABLE 4 | Step-by-step output as numbers of qualifying genes along the LSTNR analytical pipeline for liver transcriptomes from male Sprague-Dawley rats after toxicant exposure based on the mode-of-action training RNAseq dataset by the MAQC phase III SEQC crowdsource toxicogenomics (TGxSEQC) effort (GEO accession number: GSE55347).

| Criteria	Hepatotoxicity: Mode-of-Action Rat Models (N = 54)
Genes with uniquely aligned reads	30,852
Distribution of gene-wise RPM means	$P(y) \sim Weibull_{3P}(y;\alpha,\beta,\gamma)$; y = RPM α = 6.7 RPM β = 0.38 $\gamma = 2.5 \times 10^{-3}$ RPM
Independent filtering: Genes with average RPM $y > \alpha$	9,593
Linearized normalizing transformant: GLM Linear Predictor	$(y - \gamma)^{\beta - 1}$
Transformant two-way ANOVA: resolved genes across groups with respect to gene-wise mean	3,975
Resolution-Weighed ANOVA: Significant Genes (SGs) with FDR adj. p < 0.05 based on differences in resolution-weighed RPM log-fold changes (Log2FC) relative to baseline condition	5,983
Differential expression: DEGs = subset SGs that exhibit both: • resolution-weighed effect size above 5% of gene-wise variation ($\delta_{Log2FC} > 0.3 \times \sigma_{SSR}$); and • *post-hoc* pairwise-significant Log2FC differences between at least two MOAs (Student's t-test $p < 0.05$)	5,864
Reproducibility: LSTNR genes = subset of SGs that exhibit both: • resolution-weighed effect size above 5% of gene-wise variation ($\delta_{Log2FC} > 0.3 \times \sigma_{SSR}$); and • at least one group with Log2FC differences vs. baseline greater than 95% Tolerance Interval of gene×group residuals among SGs (*post-hoc* pairwise-significance not required)	1,953
Expectable DEGs: DEGREEs = Ensembl-annotated DEGs with a reproducible expectation estimate (i.e., DEGs that are also LSTNRs) and official Entrez symbol	1,510
Transcriptional profiling: Profiler DEGREEs = top DEGREEs ranked by retrospective statistical power with monotonically decreasing within-gene effect sizes Δ_{Log2FC}	**65 Profiler DEGREEs**
Diagnostic targets: Biomarkers = minimal subset of Profiler DEGREEs with predictive discriminant power based on sequential partition tree analysis (ROC scores >0.9 per phenotype)	Ucp3, Tmem86b, Sugct, Acaa1b, Hadhb, Tfam, Acaa1a, and Gsdmd

(NM_ accession prefix) were consolidated into a final list 1,510 DEGREEs with official non-duplicate Entrez gene symbols. Then, after comparing the ranks of retrospective statistical power and within-DEGREE effect size for Log2FC measurements across MOA groups, we identified a minimal set of 65 Profiler DEGREEs for transcriptional profiling (**Table 4**; see Table S2 for gene symbols).

We asked to what extent the DEGREEs detected by our method captured the underlying organization of specimens into MOA groups. Using two-way unsupervised hierarchical clustering (Ward method), we found the set of 1,510 DEGREEs sorted not only 43 of the 45 treated specimens into their original MOA and chemical agent groupings, but also recovered their stratification by route of exposure, nutritional status of vehicle, and one of two means of pathway activation: receptor mediated (RM) and non-receptor mediated (NRM). We also detected five co-expression patterns among DEGREEs, which were supported by separate clustering of correlation scores for the differential expression of DEGREEs across MOA groupings (**Figure 4A**). Still, we found that clustering based on 1,510 DEGREEs resulted in interleaved MOA groups, except for PPARA (**Figure 4A**). This kind of transcriptional cross-talk among 1,510 DEGREEs has been previously reported for this same dataset, suggesting that downstream pathways elicited by different chemical agents under the same MOA classification can converge to the same regulatory hubs through different molecular cascades and lead to distinctive transcriptional effects; conversely, chemical agents classified under different MOA can exhibit similar transcriptional signatures just as well (Funderburk et al., 2017).

Next, we investigated the capacity of Profiler DEGREEs to segregate MOA groups, and their underlying strata, compared to the pool of DEGREEs. Unsupervised clustering based on the 65 Profiler DEGREEs revealed seven MOA clades among treated specimens (I-VII) and, like its larger counterpart with 1,510 DEGREEs, five minimal gene co-expression patterns (a–e) (**Figure 4B**). Using only Profiler DEGREEs also recovered the experimental stratification of the treated specimens with the exception of 6 specimens interspersed between the Cytotoxic, DNA damage, and CAR/PXR MOA groups (**Figure 4B**). Yet, hierarchical clustering by Profiler DEGREEs also segregated each of the MOA from each other, unlike clustering with the entire pool of 1,510 DEGREEs (**Figure 4B**). This highly discriminative capacity indicates that Profiler DEGREEs represent the most discriminative subset of DEGs to extract transcriptional signatures.

The ability to segregate entire MOA groups from each other based on Profiler DEGREEs allowed comparing transcriptional profiles of each MOA across the five expression patterns. Among RM MOA groups, PPARA showed the strongest and most dissimilar transcriptional response; it was also the MOA with the most pronounced effects across the entire dataset and across all co-expression patterns (**Figure 4C**). Both Ahr and CAR/PXR exhibited modest Log2FC levels relative to the effects elicited in the PPARA groups; interestingly, expression regulation trends in Ahr groups were opposite to those in PPARA across all five expression patterns (a–e), and in three
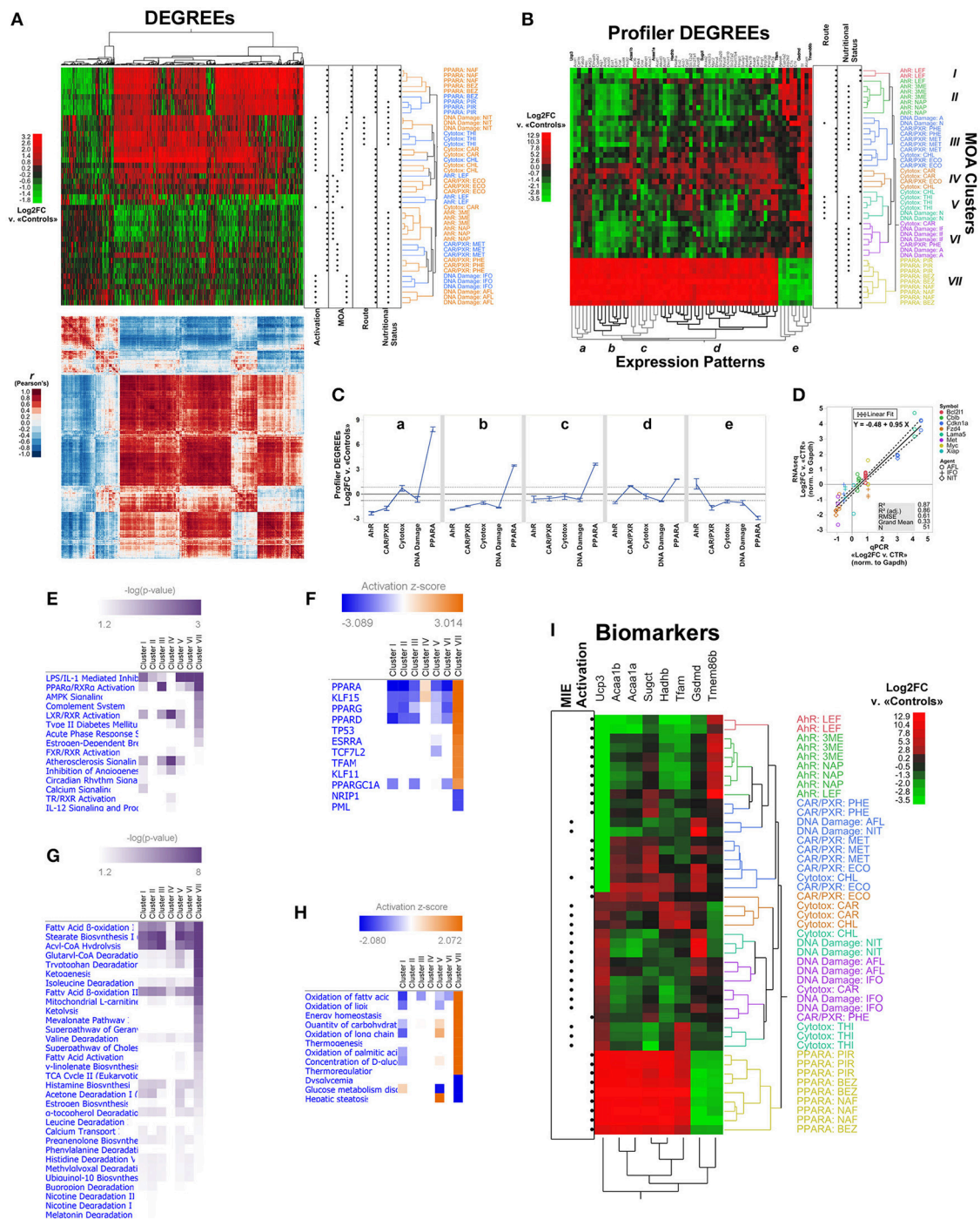
FIGURE 4 | LSTNR method analysis of the MOA training RNAseq dataset by the TGxSEQC crowdsource effort. **(A)** Heatmap plots of expression differences (top) and Pearson's r correlation coefficients (bottom) for 45 Sprague-Dawley rat liver specimens exposed to toxic agents classified under 5 different MOAs (rows; N = 9 per MOA) and 1,510 DEGREEs (columns) based on Log2FC measurements vs. the average of 9 mock-treated controls. The expression heatmap of DEGREEs (top) is colored on a green-black-red gradient scale (left) of Log2FC values (green downregulated; black, same; red, upregulated); the correlation heatmap of DEGREEs (bottom) is colored on a blue-white-red gradient scale (left) of Pearson's correlation coefficient r-values (blue, negative correlation; white, not correlated; red, positive correlation). Order of DEGREEs in both heatmaps (columns) is illustrated by the column dendrogram in the expression heatmap, and was based on two-way unsupervised hierarchical clustering of DEGREEs and specimens based on Log2FC expression differences (expression heatmap, top). Dot plots on the right of the expression heatmap (top) depict experimental classifications of treated specimens based on: means of transcriptional activation of MIEs (left to right: non-receptor mediated, or NRM vs. receptor-mediated or RM); MOA (left to right: Ahr, CAR/PXR, Cytotoxic, DNA damage, PPARA); route of exposure (left to right: intraperitoneal injection, or oral gavage); and nutritional status of vehicle (left to right: not nutritive solution, or nutritive oil-based vehicle). Inferred specimen clades, represented in the

(Continued)

FIGURE 4 | expression heatmap (top) by row dendrograms and their respective MOA × Agent labels, are shown with alternating orange and blue coloring for clarity. **(B)** Expression heatmap for two-way unsupervised clustering of Log2FC differences, as in **(A)**, using only 65 Profiler DEGREEs (columns). Profiler DEGREE columns are labeled by their respective Entrez gene symbol (top); the names of 8 biomarkers selected from the set of Profiler DEGREEs by sequential partition tree analysis are shown raised and in bold. Coloring of the row dendrogram and its labels (left) represent inferred grouping of treated specimens into 7 MOA clusters (I-VII). Inferred co-expression patterns represented by the column dendrogram (bottom; a–e) are shown with alternating gray and black coloring for clarity. Top left: heatmap is colored on a green-black-red gradient scale (top left) of Log2FC values relative to mock-treated controls (green, downregulated; black, same; red, upregulated). Dot plots on the right of the expression heatmap depict experimental classifications of treated specimens based on route of exposure (left to right: intraperitoneal injection, or oral gavage) and nutritional status of vehicle (left to right: not nutritive solution, or nutritive oil-based vehicle). **(C)** Average Log2FC expression ± s.e.m. based on Profiler DEGREEs, split into co-expression patterns from **(B)**, and separated by experimental MOA classification. **(D)** Concordance of relative expression estimates normalized to house-keeping glyceraldehyde-3-phosphate dehydrogenase (Gapdh) gene between RNA-seq and qPCR experiments for eight annotated gene symbols across hepatotoxic agents with DNA damage MOA; all assays were performed from cDNA templates derived using the same total RNA extracts for both techniques ($N = 3$ per hepatotoxic agent); qPCR assays were performed in technical duplicates per reaction plate with matched untreated samples as controls (CTR). Overall linear regression (regression mean: solid black line; ±95 CI of regression: dashed black lines) corresponds to qPCR-based average normalized Log2FC expression levels of each gene (x-axis) vs. normalized Log2FC expression measurements per sample among mRNA transcripts with matching gene symbols detected by RNA-seq (y-axis). **(E)** Enriched signaling pathways and **(F)** inferred upstream regulators for Profiler DEGREEs with |Log2FC|>0.82 under each MOA cluster identified in **(B)**, based on Ingenuity Knowledge Base ontologies; equal analyses are depicted in regards to **(G)** metabolic pathways and **(H)** inferred disease and biological functions, respectively. Pathways depicted in **(E,G)** heatmaps showed enriched representation $p < 0.05$ in at least one MOA cluster; upstream regulators and functions in **(F,H)** showed both enriched representation $p < 0.05$ and predictive $|z| > 2.0$ in at least one MOA cluster. Intensity of purple coloring in pathway heatmaps **(E,G)** represent increasing levels of significance; coloring of activation heatmaps **(F,H)** on a blue-white-orange gradient scale depicts prediction z-score values (blue, inhibited; white, inactive; orange, activated). **(I)** Expression heatmap for two-way unsupervised clustering of Log2FC differences based on the 8 biomarkers highlighted in **(B)** (columns). Coloring of the row dendrogram and its labels (right) matches the 7 MOA clusters (I-VII) depicted in **(B)**. Top right: heatmap is colored on a green-black-red gradient scale (top left) of Log2FC values relative to mock-treated controls (green, downregulated; black, same; red: upregulated). Dot plots (left) depict experimental classifications of treated specimens based means of transcriptional activation of MIEs (left to right: non-receptor mediated, or NRM vs. receptor-mediated or RM). $N = 9$ independent specimens per MOA classification with three different hepatotoxic agents each ($N = 3$ per MOA×Agent combination): Ahr (3ME = 3-methylcholantrene, 300 mg/kg/day, 5 days; NAP = β-naphthoflavone, 1,500 mg/kg/day, 5 days; LEF = leflunomide, 60 mg/kg/day, 5 days); CAR/PXR (ECO = econazole, 334 mg/kg/day, 5 days; MET = methimazole, 100 mg/kg/day, 3 days; PHE = phenobarbital, 54 mg/kg/day, 5 days); Cytotoxic (CAR = carbon tetrachloride, 1,175 mg/kg/day, 7 days; CHL = chloroform, 600 mg/kg/day, 5 days; THI = thioacetamide, 200 mg/kg/day, 5 days); DNA damage (AFL = aflatoxin B1, 0.3 mg/kg/day, 5 days; IFO = ifosfamide, 143 mg/kg/day, 3 days; NIT = N-nitrosodimethylamine, 10 mg/kg/day, 5 days); and PPARA (BEZ = bezafibrate, 617 mg/kg/day, 7 days; NAF = nafenopin, 338 mg/kg/day, 5 days; PIR = pirinixic acid, 364 mg/kg/day, 5 days). Log2FC measurements were calculated relative to 9 vehicle-only mock-treated controls. The TGxSEQC training dataset comprises 54 total individual specimens with reads aligned onto 30,852 RefSeq-annotated transcripts overall. To designate the final list of Profiler DEGREEs, Log2FC of differentially expressed transcripts (resolution-weighed ANOVA FDR $p < 0.05$ and $\delta_{Log2FC} > 0.3 \times \sigma_{SSR}$, and both post-hoc $p < 0.05$ and Log2FC > 95%TI$_{SSR}$ in at least one MOA vs. average of controls) annotated as mRNA-encoding (RefSeq NM_ suffix) were consolidated by the average of Log2FC values under matching and nonduplicate gene symbols from the Entrez database. Reference genome: rn6.

for CAR/PXR groups (a-c). Of the three RM groups, CAR/PXR exhibited the weakest differential expression levels. In general, MOA groups with NRM activation show muted expression effects compared to RM MOAs, in both cases reminiscent of CAR/PXR expression trends, and with DNA damage agents showing slightly greater effects than cytotoxic ones (**Figure 4C**). Even then, LSTNR analysis produced accurate estimates of weak differential expression levels, such as those elicited by DNA damage MOA agents, that showed significant concordance with matching qPCR validation data (Gong et al., 2014) as shown in **Figure 4D**.

To infer the biological mechanisms underlying the transcriptional signatures detected by unsupervised clustering, we performed pathway enrichment analysis using Profiler DEGREEs for each of the seven MOA clades separately (I-VII; **Figure 4B**) via the Ingenuity® platform. Profiler DEGREEs in each MOA clade were filtered for inclusion against a threshold |Log2FC|>0.82 (i.e., >1.76 expression fold-changes vs. average of controls). This threshold is the LSTNR filter, and equals the 95% tolerance interval for SNR = 1 among SGs based on residuals of gene Log2FC means within MOA × Agent groups. Expectably, PPARα activation was among the top enriched pathways, indicative not only of the strength of differential expression with PPARA MOA agents, but also of the signature in the AHR and CXR/PXR groups with the same genes, but with opposite expression regulation. Viewed together, the top enriched signaling pathways were consistent with acute-phase inflammation-related mechanisms (e.g., IL1, IL12, complement system), transcriptional regulation of xenobiotic and steroidal pathways (activation of RXR, LXR, FXR, TR, and estrogen-dependent pathways), and mitochondrion-linked metabolic remodeling (AMPK and Ca^{2+} signaling) (**Figure 4E**).

The results for signaling pathway enrichment analysis were further supported by sets of predicted upstream regulators via activation z-scores (**Figure 4F**). In particular, from the perspective of the PPARA MOA group, exclusively represented by Clade VII, the strongest predicted regulator was PPARA itself, and was also one among three other inferred activated factors characteristic of PPAR signaling (PPARG, PPARD, and PPARGC1A). Other factors included KLF15 and PML, showing opposite activation scores from each other, and both of which modulate TP53 activation (also predicted) by regulating the activity of lysine acetyltransferases such as EP300 and KAT6A in opposite manners (Haldar et al., 2010; Rokudai et al., 2013). Interestingly, our analysis also inferred the activation of TCF7L2, a transcription factor known to associate with genomic enhancers coincident with epigenetic H3K27ac post-translational modifications (Frietze et al., 2012). Activation of TFAM, the master transcription factor of mtDNA (Ekstrand et al., 2004) and one of the upregulated Profiler DEGREEs in PPARA MOA specimens based on our analysis (**Figure 4B**), was also predicted by the transcriptional signatures of the rest of the Profiler DEGREEs (**Figure 4F**).

Inferred regulation scores for upstream regulators were opposite to those in PPARA specimens for all other clades, except for clade IV. The distinct behavior of clade IV, which included two of the three liver specimens treated with carbon tetrachloride, did present distinctive enrichment of LXR/RXR activation and atherosclerosis signaling pathways, along with inferred activation of PPARA and KLF15. Those results for clade IV were consistent with known effects of carbon tetrachloride, a traditional model of chronic liver injury that elicits fibrogenic activity in hepatic stellate cells and loss of fenestration along liver sinusoids due to thickening of basal membranes with fibrillar collagens; both these responses are relayed via LXR signaling (Beaven et al., 2011; Xing et al., 2016). In this context, activation of PPAR signaling would offer a counteracting mechanism to deactivate fibrogenic stellate cells; in fact, PPARA is a known transcriptional regulator with reported protective roles against steatosis in hepatocytes (Tsuchida and Friedman, 2017).

When inspected from a metabolic standpoint, pathway enrichment analysis pointed to three pillars of mitochondrial function: fatty acid β-oxidation, degradation of xenobiotic agents, and sourcing of TCA cycle intermediates via amino acid catabolism (**Figure 4G**). The same predictions were inferred based on the activation z-scores of biological functions in the Ingenuity Knowledge Base (**Figure 4H**). Clade IV was the least enriched for pathways involving lipogenic activity, perhaps indicative of stalled fatty acid synthesis in steatotic hepatocytes (Ogrodnik et al., 2017). Once again, clade VII (i.e., PPARA MOA group) presented the most overt levels of significance and predictive inference, and predicted the opposite biological response with respect to all other clades, except for clade V (**Figure 4H**).

Of note, clade V consists of two agents from different MOA groups: thioacetamide (THI) from the cytotoxic MOA group, and N-nitrosodimethylamine (NIT) from the DNA damage group (**Figure 4B**). Still, the ability of THI to elicit secondary oxidative DNA damage has long been reported in its function as a free-radical generator (Clawson et al., 1997). Enrichment analysis and predicted biological activities suggested that both agents under clade V elicited increased biosynthesis of ketogenic precursors, oxidation of long-chain fatty acids, defective glucose metabolism, and hepatic steatosis—all of which are processes that synergize with mitochondrial respiration and rely on the integrity of the mitochondrial genome. We interpreted the combination of predicted glucose disorders and enhanced metabolism of ketogenic intermediates triggered by exposure to clade V agents to have a compensatory function in response to mtDNA damage. In a sense, these endogenous responses to NIT and THI exposure would be analogous to the effects of ketogenic diets in the Deletor mouse model, in which high-fat feeds ameliorate the chronic bioenergetic crisis that stems from defective maintenance of mtDNA integrity and copy number (Ahola-Erkkilä et al., 2010).

Last, we performed sequential tree partitioning analysis of Profiler DEGREEs to extract candidate biomarkers and test their ability to discriminate MOA groups compared to using all 65 Profiler DEGREEs. The process of biomarker selection was a two-tier process: first, tree partitioning analysis was performed with all MOA groups; then, to preempt disparately

larger transcriptional responses in PPARA MOA groups from undercutting our analysis, we repeated the partitioning analysis without including PPARA specimens. We detected eight biomarkers from both analyses combined: Ucp3, Acaa1b, Acaa1a, Sugct, Hadhb, Tfam, Gsdmd, and Tmem86b (**Figure 4I**). Altogether, these eight biomarkers represented four out of the five co-expression patterns we identified by clustering all 65 Profiler DEGREEs (**Figure 4B**). The ability to sort treated specimens by biomarker-based clustering into MOA clades was commensurate to that of Profiler DEGREEs with one notable exception, clade V, split into two well defined biomarker-based groups: the first one, for NIT-treated specimens, showed upregulated Gsdmd and downregulated Tmem86b vs. controls; the second one, for THI-treated specimens, showed upregulation of Tfam and downregulation of Sugct instead (**Figure 4I**). Among the eight biomarkers, Ucp3 showed the most striking differences among MOAs, and clearly distinguished three types of transcriptional responses. In the case of the RM Ahr and CAR/PXR MOAs, the response was consistent with silencing of Ucp3 expression; instead, RM activation of PPARA MOAs exhibited the largest levels of Ucp3 overexpression throughout; finally, NRM activation of cytotoxic and DNA damage pathways were met by modest upregulation of Ucp3 (**Figure 4I**).

In principle, the purpose of defining a minimal set of biomarkers by subsequent refinements of candidate gene signatures—e.g., DEGREEs down to Profiler DEGREEs, and Profiler DEGREEs down to biomarkers—merely seeks to determine a realistic and manageable set of testable target genes for an experimentalist to carry out validation and reproducibility assays with at the bench. The underlying assumption is that the reductive process involved in curating among candidate targets will not only reflect similar or improved sample partitioning for the selected gene subset vs. the original list of candidates (as shown in **Figure 4I**), but that the predictive strength of each selected gene improves as the list of genes becomes smaller with higher statistical stringency. To test this, we performed bootstrap forest models using all 1,510 DEGREEs, only the 65 Profiler DEGREEs, or only the 8 biomarkers as candidate lists vs. the inferred MOA clusters (I-VII) depicted in **Figure 4B**; then, we tracked the predictive strength of the 8 selected biomarker genes common to all lists in terms of G^2—the likelihood-ratio test—which: (a) represents the relative contribution of a gene among all others to assigning samples into expected clusters across bootstrapped partition trees; and (b) approximates a χ^2 distribution to estimate the statistical significance of each gene's predictive capacity. As we surmised, the contributions of biomarker genes to MOA cluster assignment showed increasing significance as the tested candidate gene lists were curated in favor of genes with more robust expression differences (**Table 5**). Notably, Ucp3 ranked as the top contributor among any candidate genes retained in all tiers of statistical stringency (**Table 5**). This last observation agreed with our earlier interpretation (based on unsupervised clustering of samples using 1,510 DEGREEs, 65 Profiler DEGREEs, or 8 biomarkers) that Ucp3 overwhelmingly outpaced any other individual gene in the training dataset in its ability to

TABLE 5 | Partitioning contribution statistics by likelihood ratio tests of eight selected biomarkers via bootstrap forest modeling of sample classification under seven inferred MOA clusters with reference gene lists at different tiers of statistical stringency (mode-of-action training RNAseq dataset, TGxSEQC; GEO accession number: GSE55347).

Gene Symbol	1,510 DEGREEs	65 Profiler DEGREEs	8 Biomarkers	Statistic
Ucp3	8.4178	14.2830	14.5505	G^2
	(9.89%)	(17.12%)	(22.92%)	(G^2_{gene}/G^2_{total}, %)
	1	1	1	Rank by G^2 [highest: 1]
	0.2091	0.0266	0.0241	p-value [df: MOA Clusters-1]
Tmem86b	0.0000	2.4276	8.5938	
	(0.00%)	(2.91%)	(13.54%)	
	1363	6	2	
	1.0000	0.8765	0.1977	
Hadhb	0.3465	1.4197	8.3181	
	(0.41%)	(1.70%)	(13.11%)	
	83	20	3	
	0.9992	0.9647	0.2157	
Gsdmd	0.0000	0.1950	7.5273	
	(0.00%)	(0.23%)	(11.86%)	
	603	50	4	
	1.0000	0.9999	0.2748	
Acaa1b	0.0000	2.3052	7.3439	
	(0.00%)	(2.76%)	(11.57%)	
	142	9	5	
	1.0000	0.8896	0.2902	
Acaa1a	0.4146	2.1847	6.5803	
	(0.49%)	(2.62%)	(10.37%)	
	61	12	6	
	0.9987	0.9020	0.3614	
Tfam	0.0000	2.2218	6.3974	
	(0.00%)	(2.66%)	(10.08%)	
	1331	10	7	
	1.0000	0.8982	0.3802	
Sugct	0.0000	0.0000	4.1599	
	(0.00%)	(0.00%)	(6.55%)	
	1308	56	8	
	1.0000	1.0000	0.6551	

assign samples to their inferred MOA cluster memberships (**Figure 4I**).

Differentially regulated Ucp3 expression across MOAs is particularly relevant in the context of mitochondrial metabolism. Besides functioning as an uncoupler of oxidative phosphorylation, Ucp3 also facilitates fatty acid metabolism and ROS regulation in mitochondria (MacLellan et al., 2005), perhaps through a role for fatty acids as competent ROS scavengers (Lemke et al., 2014). We interpreted the stark contrasts between MOA groups on the basis of Ucp3 expression alone to be indicative of remodeled lipid store management in cells, perhaps to alleviate insults on mtDNA integrity or ROS imbalances. This interpretation is supported by the functions of all other biomarkers detected by the LSTNR method (**Figure 4I**):

(a) Sugct, an enzyme involved in lysine degradation that metabolizes glutarate (Marlaire et al., 2014); (b) Acaa1a and Acaa1b, both enzymes that participate in lipid metabolism in peroxisomal compartments (Schram et al., 1987; Ferdinandusse et al., 2001); (c) Hadhb, a critical subunit of the mitochondrial trifunctional protein that governs fatty acid beta-oxidation inside mitochondria (Spiekerkoetter et al., 2004); (d) the gene encoding for lysoplasmalogenase, Tmem86b (Braverman and Moser, 2012); (e) Gsdmd, or gasdermin D, a lipid-porating protein that effects pyroptotic cell death during inflammation (Rathkey et al., 2017); and (f) Tfam, the master transcription factor for mtDNA (Woo et al., 2012; Stiles et al., 2016).

From a biological perspective, implementation of the LSTNR method unveiled different levels of transcriptional cross-talk in response to chemical exposure. Perhaps unsurprisingly, the ability to discriminate different forms of liver toxicity by transcriptional profiling was founded on maintenance of mitochondrial function, in particular by remodeling lipid metabolism to counterbalance toxic effects on aerobic respiration machinery. Still, the LSTNR method did harbor one particular strength: it provided a systematic strategy to distinguish between interconnected transcriptional signatures and characteristic ones based on tiers of statistical stringency. In that regard, the different tiers of differential gene expression that LSTNR dissects may outline the difference between targeted (or causal) triggers and their secondary (or consequential) effects in transcriptional responses to toxicants. We believe these types of results from LSTNR implementation, as shown by the analysis of the TGxSEQC dataset, reflect the underlying basis of bottom-up transcriptional networks that become more intertwined as more gene nodes are added—and yet, the LSTNR method can unravel them systematically from the top down.

DISCUSSION

Independent Filtering in LSTNR: Data Worth Making Is Not Always Worth Keeping

A key element of RNAseq analysis is choosing a sensible threshold that reflects the dynamic range of gene detection based on total aligned reads from a sequencing run. In our analyses of the TCGA and TGxSEQC datasets, we used the scale parameter α of a 3-parameter Weibull fit to gene RPM means as the independent filtering threshold. Mathematically, this ensured no more than 63% of genes with aligned reads will be retained for analysis, since by definition the α parameter is the 37th percentile of a Weibull-distributed variable (Aitkin and Clayton, 1980). In contrast, we found that our approach to independent filtering, based on a lognormal distribution, estimated a dynamic range that was valid for every pseudogene in the EPIG-seq *in silico* dataset. This outcome, which rarely occurs in the analysis of *bona fide* experimental data sets, is consistent with using simulated data since: (a) pseudogene averages in statistical groups are prescribed under known patterns for all pseudoreplicates; and (b) read count variation is modeled around those fixed averages across the entire dataset. Given these properties of simulated

data, it follows that an explanatory parametric fit of pseudogene averages should include every pseudogene if those gene averages are also parametrically tailored; if true, the estimated dynamic range of average RPM-values should be valid for all pseudogenes and, consequently, all pseudogenes should pass independent filtering—just as we saw in the EPIG-seq *in silico* dataset.

It is worth mentioning that the point of independent filtering has little to do with estimates of relative expression. In fact, fold-change differences across groups are scored within genes in RNAseq, meaning differential expression measurements for individual genes are separate from each other whether they fall within the dynamic range of detection or not. However, subjecting all genes to multivariate differential expression analyses, including underresolved ones indistinguishable from instrumental noise, undermines the ability to adjust significance tests for multiple comparisons (Benjamini and Hochberg, 1995; Tamhane and Dunlop, 2000). Therefore, if DEGs are selected based on significance scores alone, their numbers will be inflated (Type I error) if underresolved genes are not discarded ahead of inferential testing.

Empirical Fitting of Gene-Wise Coverage

It is in this context that effect size filtering at the gene level becomes particularly relevant to the analysis of deeply sequenced RNAseq data sets. Depending on the type of experimental design, additional levels of DEG discrimination may be needed, for example, when dealing with transient expression differences, such as mtDNA depletion time courses in DN-POLG cells (Martínez-Reyes et al., 2016). In time course experiments, differences in gene expression are expectably smaller between successive time points than comparisons between start and end points. With shorter sampling intervals in an experiment come smaller expression differences—yet instrumental noise, which is relatively fixed, may be more prominent than the dispersion of differential expression measurements. This is one major reason why relying exclusively on *pairwise* significance tests to detect DEGs across 3+ groups are prone to Type II errors. If a practical effect size criterion is not imposed to discriminate genes that show statistical significance (or not) in RNAseq tests with small sample sizes, it is difficult to separate between genes whose significant expression differences are more likely to be real, rather than anecdotal, in an underpowered experimental design (Ioannidis, 2005).

Average sequencing depth across detected genes in RNAseq experiments account predominantly for random dispersion and instrumental error combined—except in DEGs. In that sense, one can interpret the behavior of accrued reads in each detected gene as a fingerprint of how read counts vary with sequencing depth. From such perspective, the within-gene sample size is the total number of experimental samples, and the between genes sample size is the number of genes retained after independent filtering. One can then "studentize" read count variation across gene × group statistical blocks by producing an "image" of transformed RPM rates. This can be achieved by fitting a generalized linear model (GLM), which are statistical instruments to account for magnitude and resolution differences all at once in non-linear and non-Gaussian systems (Nelder and Wedderburn, 1972).

To perform GLM, it is necessary to restate an observed probability density function in the general form of the exponential family. The purpose is to devise a transformation that turns RPM data into a set of values showing a linear relationship between samples and their averages, known as a *linear predictor* or transformant $B(\vartheta)$ [or B in short]; furthermore, if the chosen linear predictor B yields transformant values whose variation around within-gene averages behave like a classical distribution from the exponential family (e.g., normal, binomial, Poisson, or exponential), then the transformant values of B can be manipulated algebraically into normally distributed scores using what is known as a *link function* $\eta(B)$ [or η in short]. As a result, η is a representation of RPM that, unlike RPM-values themselves, are linear and normally distributed—which meets requirements for ordinary multivariate ANOVA tests of significance; this is the basis of GLM (see Supplementary Methods) (Nelder and Wedderburn, 1972).

The metric of interest when using GLM to model gene-wise RPM averages is not the actual values of the transformant B-values, but what those values represent: an instrument to explain which genes are *better* resolved based on statistical evidence about their location in the dynamic range of sequenced read detection. In that context, significance levels of genes based on ANOVA testing of the linear predictor B are a proxy for how robust are the differences between groups within one gene vs. all other genes based on their RPM data and its non-linear variation. Therefore, those significance scores rank the experiment's sensitivity to changes in each gene. In the LSTNR method, we used these scores to produce gene "resolution weights" when testing log-fold expression measurements.

The weight function of choice is the cumulative hazard of gene significance scores. The cumulative hazard function, which is the complement of the cumulative density function in negative logarithmic scale, becomes a score for how strong is the expectation that read counts from individual genes estimate "true" relative expression differences between groups. The cumulative hazard function has many advantages in regards to how it represents RPM resolution: it can be determined from bounded population data (i.e., the number of detected genes is a finite number), it is continuous-valued (i.e., not ordinal), and monotonically increasing (larger weights for better resolved genes). Also, it is derived from read count data without being a direct transformation of read count values; the benefit is that the same weight can be applied to each gene across samples or experimental realizations whether net read counts are the same or not between individual replicates. In the context of RNAseq output, the behavior of cumulative hazard functions resembles that of the "tightness" of relative expression measurements around the mean from sequencing data: genes with low read counts are poorly detected and yield noisy estimates of log-fold differences; genes with more aligned reads produce more robust estimates with increasing coverage; and genes near saturated sequencing levels plateau to the absolute resolution limit of the sequencing instrument.

Based on the TCGA dataset, we found the behavior of resolution weights to be consistent with how read count differences in RNAseq data behave: genes with better resolution

will often present either low read counts and large differences between groups, or high read counts with modest differences and tight within-group variability. Since RNAseq is based on PCR amplification, observing tight distributions for abundant transcripts may occur rarely, since amplification errors are propagated exponentially. Using separate implementations of the LSTNR method in each individual realization ameliorated the impact of such features, all too common in noisy RNAseq data. In effect, resolution scaling in the LSTNR method adjusted for differential expression tests for exponential propagation of PCR errors.

In theory, there is an additional benefit worth noting that LSTNR offers. GLM with canonical link functions relies on the statistical sufficiency of exponential family distributions, and the resolution weights of genes are derived from linear predictor ANOVA based on ranks, not scores, of statistical significance. If individual experimental replicates are a true representative sample of a population, and their distribution belongs to the exponential family (as implemented in the LSTNR method), then a notable statistical corollary follows: the resolution power of gene detection from one experiment is valid for separate realizations from the same population. In practical terms, this means the resolution weights for genes detected in one study are valid for any additional sequencing of its replicates, entire repeated experiments, and replicate studies—*regardless of sample size or sequencing depth*. If true, the GLM basis of the LSTNR method makes it a particularly attractive pipeline to determine consensus resolution weight matrices from small training sets heading into consolidated meta-analyses of much larger cohorts. As our findings suggests, the use of LSTNR-derived gene weight matrices would protect DEG analysis against detrimental batch effects and subsampling errors, both of which are characteristic of epidemiological and clinical studies.

Detecting DEGS by Resolution-Weighed Multivariate Analysis in the LSTNR Method

The customary methods of inferential testing among multiple groups are ANOVA models because of their computational efficiency and ease of implementation. The main caveat of ANOVA models pertains statistical power, which strongly depends on sample sizes, normally distributed responses, and homoscedasticity across statistical blocks. These requirements are a major impediment to the management of RNAseq data sets; for one, sample sizes are limiting in RNAseq applications due to cost constraints, since sequencing reagents, equipment and necessary computational support to convert raw sequencing output into genome-aligned reads are all expensive. In regard to normality of measurements, it is common practice to represent expression levels as log-transformed read counts, such that expression fold-changes between groups become arithmetic differences between means with distributions that approximate normality; this is not only questionable for most RNAseq experiments, mainly due to the impedingly small sample sizes mentioned earlier, but also insufficient as it does not address lack of homoscedasticity.

Log2FC values, which are read count ratios, are a measurement of relative expression levels for a gene between conditions; however, Log2FC values by themselves do not offer any information on their own measurement resolution because the read counts used to estimate them are "divided out." Calculating significance scores based on Log2FC values is only statistically fair if all genes show equal levels of read coverage or, as in the LSTNR method, if relative expression measurements are adjusted to reflect different resolution levels among detected genes. Introducing resolution weights to multivariate ANOVA testing of relative expression differences posited attractive statistical advantages. In principle, resolution scaling of log-fold differences should: (a) improve homoscedasticity across genes, as shown in **Figure 2D**; (b) discriminate between highly variable (largest fold-differences) and highly granular (largest read counts) genes, as shown in **Figures 2C**, **3B**; and (c) prioritize genes whose prospective differential expression are the most reproducible should RNAseq experiments be repeated. Above all else, the issue of reproducibility described in (c) was the reason behind our split-pool analyses approach to the TCGA dataset, and the identification of a consensus set of Profiler DEGREEs (see **Table 3**).

When we analyzed the breast cancer dataset from TCGA, we found large discrepancies in the number of significantly resolved genes among separate analyses, yet the number of SGs was very consistent across the board. In our view, this illustrates the main strength behind the LSTNR method: whether particular sample groupings lead to statistically significant dispersion in RPM-values in one subset vs. another is less relevant than the actual ranking of their relative resolution within replicate experiments. Our interpretation is based on the premise that estimates of variability in observed RPM-values within genes are representative of how accurately they were detected *by instrumentation* with respect to each other, not whether accuracy in sequencing was exactly the same for all detected genes between two different replicate experiments. In that sense, our findings showed how coupling of individual sequencing resolution of genes with relative expression differences in the LSTNR method alleviates discrepancies in the number of detectable SGs across replicate experiments, and performs well when the statistical groups are well-defined in advance—e.g., treatments, phenotypes, controls, etc.

The LSTNR Method Can Estimate Signal, Noise, and Reproducibility Benchmarks

Benchmarking the expected dispersion range in RNAseq output is the foundation to a power analysis that outlines what are the smallest differences to expect ahead of deeper sequencing; often, though, access to RNAseq and costs are too prohibitive to pursue any benchmarking efforts. In such cases, the same projection for expected dispersion ranges can be used as a practical signal-to-noise (SNR) threshold of reproducibility that a select subset of genes can be validated against, even if the expression differences measured with RNAseq data for those genes were statistically significant or not. A benchmark SNR threshold of within-gene dispersion in sequenced RPM-values

is a key metric for experimentalists, because it helps justify or rule out additional rounds of sequencing (or patient recruitment in clinical settings) when the statistical significance of any transcriptional differences is marginal—or, if all else fails, to shift focus to better experimental alternatives or re-design a project altogether.

To perform relative expression analysis, we determined a reference expression value in each gene equal to the average $\log_2(\text{RPM})$ in a reference control group. The advantage of establishing a fixed "null hypothesis" reference, instead of using the distribution of samples in the control group, is that sample means and variation from individual replicates can be estimated for all groups, including the baseline condition, and before inferential testing of significance. The benefits are many: in practice, RNAseq experiments often produce expression measurements that, in retrospective, are underpowered, biased or unable to detect relevant biological effects because intrinsic biological variability among specimens is simply too large; in other cases, cost constraints limit studies using RNAseq technology to low replication models (Conesa et al., 2016). Faced with these challenges, one could instead project from existing data (even if it is only available for control samples) how large the dispersion in RNAseq output will be; for example, we calculated the predicted 95% tolerance intervals around the mean Log2FC measurements as a surrogate estimate of expectable measurement variability. Those tolerance levels, which can be calculated directly from the residuals of DEGs, represent a transcriptome-wide signal-to-noise (SNR) thresholds of practical reproducibility. Furthermore, since these are based on the variation of individual Log2FC measurements around means of gene × group blocks, they project the scales of expression differences that should be reproduced by other techniques (e.g., targeted qPCR) or in repeated experiments with reasonable sample sizes (e.g., 95% prediction intervals). In the case of the simulated EPIG-seq data, our results suggested that dispersion of Log2FC measurements was roughly the same whether it was calculated based on all genes passing filter or only on statistically significant ones (**Figure 2G**). This entails that SNR-based criteria for reproducible differential expression can be established ahead of statistical testing based solely on the distribution of Log2FC residuals. In practical terms, would allow confirmatory qPCR assays to be designed from low replication RNAseq studies, simply because noise benchmarking does not require detecting DEGs.

We must point out, going back to the EPIG-seq *in silico* dataset, how the 73 patterned pseudogenes that did not match their simulated trends originated from prescribed *in silico* patterns A, C, or E (**Figure 2E**). All these three patterns show higher average expression levels in all treatment groups vs. the baseline average; in contrast, the matched patterns B and D both exhibit a dominant downregulation trend: expression in pattern B pseudogenes decreases further when looking across treatment groups, and pattern D exhibits a return to baseline after an "expression spike" in the first treatment group (**Figure 2H**). These results, which suggest that downregulation trends are easier to discriminate than upregulation ones, may simply reflect the fact that resolution is finite—meaning the number

of possible values for RPM differences is countable, since they are ultimately based on integer-based read counts. This implies that, when the number of read counts drops relative to a control, it reduces the number of combinations available to measure gene expression ratios. To clustering algorithms, larger downregulation differences may become easier to discriminate as they behave more like piecewise jumps. In contrast, differences of equal magnitude, but upregulated relative to controls, tend toward a continuum because the scale of possible values is refined with increasing numbers of reads per gene—and, by similar logic, upregulation differences are harder to discriminate by clustering analysis. Such behavior of countable differences in read counts has somewhat intuitive implications: for genes with low coverage to be detected as differentially expressed, both the net and proportional differences in read counts between groups must be large and far from the instrumental background so as to be more accurate and less piecewise; for genes with high read counts, significance is possible for smaller and more precise proportional differences, but is only justifiable if the variation in net read counts is also tight. Above everything else, these are the governing premises behind our design strategy for the LSTNR method.

Co-expression Patterns Detected Through LSTNR Method Are Reproducible at Different Statistical Stringency Levels

To define candidate biomarkers for different transcriptional signatures, the LSTNR method defines a subset of Profiler DEGREEs, equal to the largest subset of ranked DEGREEs with monotonically decreasing retrospective statistical power ($\ll\pi\gg$ >90%) and effect size Δ_{Log2FC} *at the same time*. The rationale behind this approach is to maximize discriminatory potential among transcriptional signatures of each breast cancer specimen by accounting for the largest possible differences between subtypes (i.e., effect size) while minimizing the commensurate false-positive rate projected from the available experimental data (statistical power).

In both experimental datasets, Profiler DEGREEs matched the discriminant capacity of other pipelines, but with considerably less genes overall. The strongest evidence for this came from our analysis of the TGxSEQC training dataset. If stratified by agent, the TGxSEQC dataset is a 15-group experimental design with minimal replication level ($N = 3$). Hence, it cannot be split into realizations for parallel testing, and has to be tested through a single-shot analysis. Another challenge to testing the TGxSEQC training set deals with prospective statistical power: given 15 groups, the number of possible comparisons between them is very large. Therefore, if the analysis is performed based on pairwise gene expression differences, there is a large risk of over-correcting for multiple testing. As a result, this would underestimate the number of DEGs, and limit the ability to discriminate specimens from different MOA groups. Consequently, grouping the pool of 45 specimens under too many categories is detrimental to the statistical power of the experimental design as a whole. To circumvent this challenge, and based on the fact that each of the individual chemical agents belongs to one out of five overarching MOAs, we opted to group

specimens under one out of five MOA groups with $N = 9$. Even then, analysis of the TGxSEQC training set using the LSTNR method captured additional sample classifications at different tiers of statistical stringency, based on MOA information alone, and simply by unsupervised hierarchical clustering. In addition, each of the three tiers of discriminant power (i.e., based on 1,510 DEGREEs, 65 Profiler DEGREEs, and 8 biomarkers) were consistent with each other in regards to clustering and detection of co-expression patterns. Most importantly, each increasingly stringent filter improved on the ability of its predecessor to segregate individual MOA groups: clustering by 1,510 separated MOA × Agent groups, but displayed interlaced strata from RM and NRM activation modes; based on 65 Profiler DEGREEs, clustering effectively separated each MOA from all others; and the 8 biomarkers effectively refined the discriminative power between NRM cytotoxic and DNA damage MOA strata.

CONCLUSION

We put forth the LSTNR method as an alternative pipeline to tackle current obstacles in RNAseq analysis: first, it defines a detection limit for genes with respect to random errors in instrumental sequencing and read alignment by fitting the observed distribution of aligned read counts; in return, this empirical fit offers a data-driven threshold for independent filtering. Then, the pipeline accounts for non-linear and non-homogeneously distributed variation in read counts per gene; this is used to generate a gene-wise resolution score based on read counts that, when implemented as a weight function, improves the normality and homogeneity of relative expression measurements. With the LSTNR method, the improvements in homoscedasticity by resolution scaling of Log2FC differences allow experimental designs with multiple groups to be tested by standard ordinary ANOVA techniques. Furthermore, DEGs detected by the LSTNR method capture the same transcriptional signatures at different tiers of statistical stringency with high accuracy; we found the performance of the LSTNR method was so robust that it required less genes than previously reported to discriminate breast cancer phenotypes and hepatotoxic MOAs using the same experimental datasets (Perou et al., 2000; Cancer Genome Atlas, 2012; Gong et al., 2014; Wang et al., 2014; Li and Bushel, 2016; Funderburk et al., 2017). As an added benefit, the LSTNR method can produce noise benchmarking estimates to validate RNAseq experiments against by benchtop techniques, regardless of statistical significance or sample replication levels. Altogether, these features set the LSTNR method apart as an agnostic pipeline that: (a) can be programmed for automated processing of RNAseq data with minimal user intervention; and (b) can estimate thresholds of experimental reproducibility for confirmatory assays using RNAseq studies with limiting sample sizes.

AUTHOR CONTRIBUTIONS

OL conceptualized and implemented the analytical methodology described in the study, performed data analysis, and wrote the manuscript. OL and JS wrote the initial manuscript drafts leading to submission. JS and RW supervised the selection of simulated and experimental RNAseq data sets to validate and test the statistical approach described herein. All authors contributed in the design of the study and interpretation of the results, as well as revised, read, and approved the submitted version.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2018.00176/full#supplementary-material

REFERENCES

Ahola-Erkkilä, S., Carroll, C. J., Peltola-Mjösund, K., Tulkki, V., Mattila, I., Seppänen-Laakso, T., et al. (2010). Ketogenic diet slows down mitochondrial myopathy progression in mice. *Hum. Mol. Genet.* 19, 1974–1984. doi: 10.1093/hmg/ddq076

Aird, D., Ross, M. G., Chen, W. S., Danielsson, M., Fennell, T., Russ, C., et al. (2011). Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. *Genome Biol.* 12:R18. doi: 10.1186/gb-2011-12-2-r18

Aitkin, M., and Clayton, D. (1980). The fitting of exponential, weibull and extreme value distributions to complex censored survival data using GLIM. *J. R. Stat. Soc. Ser. C* 29, 156–163. doi: 10.2307/2986301

Auer, P. L., and Doerge, R. W. (2010). Statistical design and analysis of RNA sequencing data. *Genetics* 185, 405–416. doi: 10.1534/genetics.110.114983

Beaven, S. W., Wroblewski, K., Wang, J., Hong, C., Bensinger, S., Tsukamoto, H., et al. (2011). Liver X receptor signaling is a determinant of stellate cell activation and susceptibility to fibrotic liver disease. *Gastroenterology* 140, 1052–1062. doi: 10.1053/j.gastro.2010.11.053

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B* 57, 289–300.

Benjamini, Y., and Speed, T. P. (2012). Summarizing and correcting the GC content bias in high-throughput sequencing. *Nucleic Acids Res.* 40:e72. doi: 10.1093/nar/gks001

Braverman, N. E., and Moser, A. B. (2012). Functions of plasmalogen lipids in health and disease. *Biochim. Biophys. Acta* 1822, 1442–1452. doi: 10.1016/j.bbadis.2012.05.008

Bullard, J. H., Purdom, E., Hansen, K. D., and Dudoit, S. (2010). Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. *BMC Bioinformatics* 11:94. doi: 10.1186/1471-2105-11-94

Cancer Genome Atlas, N. (2012). Comprehensive molecular portraits of human breast tumours. *Nature* 490, 61–70. doi: 10.1038/nature11412

Chou, J. W., Zhou, T., Kaufmann, W. K., Paules, R. S., and Bushel, P. R. (2007). Extracting gene expression patterns and identifying co-expressed genes from microarray data reveals biologically responsive processes. *BMC Bioinformatics* 8:427. doi: 10.1186/1471-2105-8-427

Clawson, G. A., Benedict, C. M., Kelley, M. R., and Weisz, J. (1997). Focal nuclear hepatocyte response to oxidative damage following low dose thioacetamide intoxication. *Carcinogenesis* 18, 1663–1668. doi: 10.1093/carcin/18.8.1663

Cloonan, N., Forrest, A. R., Kolle, G., Gardiner, B. B., Faulkner, G. J., Brown, M. K., et al. (2008). Stem cell transcriptome profiling via massive-scale mRNA sequencing. *Nat. Methods* 5, 613–619. doi: 10.1038/nmeth.1223

Conesa, A., Madrigal, P.,Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., et al. (2016). A survey of best practices for RNA-seq data analysis. *Genome Biol.* 17:13. doi: 10.1186/s13059-016-0881-8

Ekstrand, M. I., Falkenberg, M., Rantanen, A., Park, C. B., Gaspari, M., Hultenby, K., et al. (2004). Mitochondrial transcription factor A regulates mtDNA copy number in mammals. *Hum. Mol. Genet.* 13, 935–944. doi: 10.1093/hmg/ddh109

Ferdinandusse, S., Denis, S., Mooijer, P. A., Zhang, Z., Reddy, J. K., Spector, A. A., et al. (2001). Identification of the peroxisomal beta-oxidation enzymes involved in the biosynthesis of docosahexaenoic acid. *J. Lipid Res.* 42, 1987–1995.

Finotello, F., and Di Camillo, B. (2015). Measuring differential gene expression with RNA-seq: challenges and strategies for data analysis. *Brief. Funct. Genomics* 14, 130–142. doi: 10.1093/bfgp/elu035

Frietze, S., Wang, R., Yao, L., Tak, Y. G., Ye, Z., Gaddis, M., et al. (2012). Cell type-specific binding patterns reveal that TCF7L2 can be tethered to the genome by association with GATA3. *Genome Biol.* 13:R52. doi: 10.1186/gb-2012-13-9-r52

Funderburk, K. M., Auerbach, S. S., and Bushel, P. R. (2017). Crosstalk between receptor and non-receptor mediated chemical modes of action in rat livers converges through a dysregulated gene expression network at tumor suppressor Tp53. *Front. Genet.* 8:157. doi: 10.3389/fgene.2017.00157

Gong, B., Wang, C., Su, Z., Hong, H., Thierry-Mieg, J., Thierry-Mieg, D., et al. (2014). Transcriptomic profiling of rat liver samples in a comprehensive study design by RNA-Seq. *Sci. Data* 1:140021. doi: 10.1038/sdata.2014.21

Hahn, G. J., and Meeker, W. Q. (1991). *Statistical Intervals: A Guide for Practitioners.* New York, NY: Wiley.

Haldar, S. M., Lu, Y., Jeyaraj, D., Kawanami, D., Cui, Y., Eapen, S. J., et al. (2010). Klf15 deficiency is a molecular link between heart failure and aortic aneurysm formation. *Sci. Transl. Med.* 2:26ra26. doi: 10.1126/scitranslmed.3000502

Hansen, K. D., Brenner, S. E., and Dudoit, S. (2010). Biases in Illumina transcriptome sequencing caused by random hexamer priming. *Nucleic Acids Res.* 38:e131. doi: 10.1093/nar/gkq224

Hansen, K. D., Wu, Z., Irizarry, R. A., and Leek, J. T. (2011). Sequencing technology does not eliminate biological variability. *Nat. Biotechnol.* 29, 572–573. doi: 10.1038/nbt.1910

Ioannidis, J. P. (2005). Why most published research findings are false. *PLoS Med.* 2:e124. doi: 10.1371/journal.pmed.0020124

Law, C. W., Chen, Y., Shi, W., and Smyth, G. K. (2014). voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol.* 15:R29. doi: 10.1186/gb-2014-15-2-r29

Lemke, R. A., Peterson, A. C., Ziegelhoffer, E. C., Westphall, M. S., Tjellström, H., Coon, J. J., et al. (2014). Synthesis and scavenging role of furan fatty acids. *Proc. Natl. Acad. Sci. U.S.A.* 111, E3450–E3457. doi: 10.1073/pnas.1405520111

Li, J., and Bushel, P. R. (2016). EPIG-Seq: extracting patterns and identifying co-expressed genes from RNA-Seq data. *BMC Genomics* 17:255. doi: 10.1186/s12864-016-2584-7

Li, J., and Tibshirani, R. (2013). Finding consistent patterns: a nonparametric approach for identifying differential expression in RNA-Seq data. *Stat. Methods Med. Res.* 22, 519–536. doi: 10.1177/0962280211428386

Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta C}$T method. *Methods* 25, 402–408. doi: 10.1006/meth.2001.1262

MacLellan, J. D., Gerrits, M. F., Gowing, A., Smith, P. J., Wheeler, M. B., and Harper, M. E. (2005). Physiological increases in uncoupling protein 3 augment fatty acid oxidation and decrease reactive oxygen species production without uncoupling respiration in muscle cells. *Diabetes* 54, 2343–2350. doi: 10.2337/diabetes.54.8.2343

Marlaire, S., Van Schaftingen, E., and Veiga-da-Cunha, M. (2014). C7orf10 encodes succinate-hydroxymethylglutarate CoA-transferase, the enzyme that converts glutarate to glutaryl-CoA. *J. Inherit. Metab. Dis.* 37, 13–19. doi: 10.1007/s10545-013-9632-0

Martínez-Reyes, I., Diebold, L. P., Kong, H., Schieber, M., Huang, H., Hensley, C. T., et al. (2016). TCA cycle and mitochondrial membrane potential are necessary for diverse biological functions. *Mol. Cell.* 61, 199–209. doi: 10.1016/j.molcel.2015.12.002

Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* 5, 621–628. doi: 10.1038/nmeth.1226

Nazarov, P. V., Muller, A., Kaoma, T., Nicot, N., Maximo, C., Birembaut, P., et al. (2017). RNA sequencing and transcriptome arrays analyses show opposing results for alternative splicing in patient derived samples. *BMC Genomics* 18:443. doi: 10.1186/s12864-017-3819-y

Nelder, J. A., and Wedderburn, R. W. (1972). Generalized linear models. *J. R. Stat. Soc. Ser. A-G* 135, 370–384.

Oberg, A. L., Bot, B. M., Grill, D. E., Poland, G. A., and Therneau, T. M. (2012). Technical and biological variance structure in mRNA-Seq data: life in the real world. *BMC Genomics* 13:304. doi: 10.1186/1471-2164-13-304

Odeh, R. E., and Owen, D. B. (1980). *Tables for Normal Tolerance Limits, Sampling Plans, and Screening.* New York, NY: Dekker.

Ogrodnik, M., Miwa, S., Tchkonia, T., Tiniakos, D., Wilson, C. L., Lahat, A., et al. (2017). Cellular senescence drives age-dependent hepatic steatosis. *Nat. Commun.* 8:15691. doi: 10.1038/ncomms15691

Oshlack, A., Robinson, M. D., and Young, M. D. (2010). From RNA-seq reads to differential expression results. *Genome Biol.* 11:220. doi: 10.1186/gb-2010-11-12-220

Parker, J. S., Mullins, M., Cheang, M. C., Leung, S., Voduc, D., Vickery, T., et al. (2009). Supervised risk predictor of breast cancer based on intrinsic subtypes. *J. Clin. Oncol.* 27, 1160–1167. doi: 10.1200/JCO.2008.18.1370

Perou, C. M., Sørlie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., Rees, C. A., et al. (2000). Molecular portraits of human breast tumours. *Nature* 406, 747–752. doi: 10.1038/35021093

Pfaffl, M. W. (2001). A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res.* 29:e45. doi: 10.1093/nar/29.9.e45

Rathkey, J. K., Benson, B. L., Chirieleison, S. M., Yang, J., Xiao, T. S., Dubyak, G. R., et al. (2017). Live-cell visualization of gasdermin D-driven pyroptotic cell death. *J. Biol. Chem.* 292, 14649–14658. doi: 10.1074/jbc.M117.797217

Risso, D., Schwartz, K., Sherlock, G., and Dudoit, S. (2011). GC-content normalization for RNA-Seq data. *BMC Bioinformatics* 12:480. doi: 10.1186/1471-2105-12-480

Robles, J. A., Qureshi, S. E., Stephen, S. J., Wilson, S. R., Burden, C. J., and Taylor, J. M. (2012). Efficient experimental design and analysis strategies for the detection of differential expression using RNA-Sequencing. *BMC Genomics* 13:484. doi: 10.1186/1471-2164-13-484

Rokudai, S., Laptenko, O., Arnal, S. M., Taya, Y., Kitabayashi, I., and Prives, C. (2013). MOZ increases p53 acetylation and premature senescence through its complex formation with PML. *Proc. Natl. Acad. Sci. U.S.A.* 110, 3895–3900. doi: 10.1073/pnas.1300490110

Roy, N. C., Altermann, E., Park, Z. A., and McNabb, W. C. (2011). A comparison of analog and Next-Generation transcriptomic tools for mammalian studies. *Brief. Funct. Genomics* 10, 135–150. doi: 10.1093/bfgp/elr005

Schram, A. W., Goldfischer, S., van Roermund, C. W., Brouwer-Kelder, E. M., Collins, J., Hashimoto, T., et al. (1987). Human peroxisomal 3-oxoacyl-coenzyme A thiolase deficiency. *Proc. Natl. Acad. Sci. U.S.A.* 84, 2494–2496. doi: 10.1073/pnas.84.8.2494

Sørlie, T., Perou, C. M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., et al. (2001). Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc. Natl. Acad. Sci. U.S.A.* 98, 10869–10874. doi: 10.1073/pnas.191367098

Spiekerkoetter, U., Khuchua, Z., Yue, Z., Bennett, M. J., and Strauss, A. W. (2004). General mitochondrial trifunctional protein (TFP) deficiency as a result of either alpha- or beta-subunit mutations exhibits similar phenotypes because mutations in either subunit alter TFP complex expression and subunit turnover. *Pediatr. Res.* 55, 190–196. doi: 10.1203/01.PDR.0000103931.80055.06

Stiles, A. R., Simon, M. T., Stover, A., Eftekharian, S., Khanlou, N., Wang, H. L., et al. (2016). Mutations in TFAM, encoding mitochondrial transcription factor A, cause neonatal liver failure associated with mtDNA depletion. *Mol. Genet. Metab.* 119, 91–99. doi: 10.1016/j.ymgme.2016.07.001

Tamhane, A. C., and Dunlop, D. D. (2000). *Statistics and Data Analysis: From Elementary to Intermediate* (Upper Saddle River, NJ: Prentice Hall).

Tarazona, S., Garcia-Alcalde, F., Dopazo, J., Ferrer, A., and Conesa, A. (2011). Differential expression in RNA-seq: a matter of depth. *Genome Res.* 21, 2213–2223. doi: 10.1101/gr.124321.111

Tsuchida, T., and Friedman, S. L. (2017). Mechanisms of hepatic stellate cell activation. *Nat. Rev. Gastroenterol. Hepatol.* 14, 397–411. doi: 10.1038/nrgastro.2017.38

Wang, C., Gong, B., Bushel, P. R., Thierry-Mieg, J., Thierry-Mieg, D., Xu, J., et al. (2014). The concordance between RNA-seq and microarray data depends on chemical treatment and transcript abundance. *Nat. Biotechnol.* 32, 926–932. doi: 10.1038/nbt.3001

Woo, D. K., Green, P. D., Santos, J. H., D'Souza, A. D., Walther, Z., Martin, W. D., et al. (2012). Mitochondrial genome instability and ROS enhance intestinal tumorigenesis in APC(Min/+) mice. *Am. J. Pathol.* 180, 24–31. doi: 10.1016/j.ajpath.2011.10.003

Wu, H., Wang, C., and Wu, Z. (2013). A new shrinkage estimator for dispersion improves differential expression detection in RNA-seq data. *Biostatistics* 14, 232–243. doi: 10.1093/biostatistics/kxs033

Xing, Y., Zhao, T., Gao, X., and Wu, Y. (2016). Liver X receptor alpha is essential for the capillarization of liver sinusoidal endothelial cells in liver injury. *Sci. Rep.* 6:21309. doi: 10.1038/srep 21309

Check for updates

Evaluation of Bioinformatics Approaches for Next-Generation Sequencing Analysis of microRNAs with a Toxicogenomics Study Design

Halil Bisgin[1], Binsheng Gong[2], Yuping Wang[2] and Weida Tong[2]*

[1] Department of Computer Science, Engineering, and Physics, University of Michigan–Flint, Flint, MI, United States,
[2] Division of Bioinformatics and Biostatistics, National Center for Toxicological Research (FDA), Jefferson, AR, United States

MicroRNAs (miRNAs) are key post-transcriptional regulators that affect protein translation by targeting mRNAs. Their role in disease etiology and toxicity are well recognized. Given the rapid advancement of next-generation sequencing techniques, miRNA profiling has been increasingly conducted with RNA-seq, namely miRNA-seq. Analysis of miRNA-seq data requires several steps: (1) mapping the reads to miRBase, (2) considering mismatches during the hairpin alignment (windowing), and (3) counting the reads (quantification). The choice made in each step with respect to the parameter settings could affect miRNA quantification, differentially expressed miRNAs (DEMs) detection and novel miRNA identification. Furthermore, these parameters do not act in isolation and their joint effects impact miRNA-seq results and interpretation. In toxicogenomics, the variation associated with parameter setting should not overpower the treatment effect (such as the dose/time-dependent effect). In this study, four commonly used miRNA-seq analysis tools (i.e., miRDeep2, miRExpress, miRNAkey, sRNAbench) were comparatively evaluated with a standard toxicogenomics study design. We tested 30 different parameter settings on miRNA-seq data generated from thioacetamide-treated rat liver samples for three dose levels across four time points, followed by four normalization options. Because both miRExpress and miRNAkey yielded larger variation than that of the treatment effects across multiple parameter settings, our analyses mainly focused on the side-by-side comparison between miRDeep2 and sRNAbench. While the number of miRNAs detected by miRDeep2 was almost the subset of those detected by sRNAbench, the number of DEMs identified by both tools was comparable under the same parameter settings and normalization method. Change in the number of nucleotides out of the mature sequence in the hairpin alignment (window option) yielded the largest variation for miRNA quantification and DEMs detection. However, such a variation is relatively small compared to the treatment effect when the study focused on DEMs that are more critical to interpret the toxicological effect. While the normalization methods introduced a large variation in DEMs, toxic behavior of thioacetamide showed consistency in the trend of time-dose responses. Overall, miRDeep2 was found to be preferable over other choices when the window option allowed up to three nucleotides from both ends.

Keywords: miRNA, next generation sequencing (NGS), miRNA-seq, miRNA profiling, toxicogenomics, thioacetamide

INTRODUCTION

MicroRNAs (miRNAs), which are ~22 nucleotides (nts) short non-coding RNAs, comprise a class of gene regulatory elements that have a major effect on the stability and translation of messenger RNAs (mRNAs). The regulatory effect of miRNAs has been observed in the development stages, tissue-specific expressions, disease states and toxicological effects. Recent studies reveal that the role of miRNAs is more profound than it was suspected (Zhang and Su, 2009) as they target ~30% of all genes in animals (Lewis et al., 2005) and contribute to a spectrum of human diseases with potential therapeutic options (Chen and Verfaillie, 2014; Finch et al., 2014). For example, in cancer research, miRNAs (e.g., miR-374a) can serve as prognostic markers (Vosa et al., 2011). Despite their low abundance in body fluids, miRNAs have been studied as non-invasive biomarkers for diseases (Cortez et al., 2011).

Due to their crucial role in regulatory processes, miRNAs are also associated with toxicity (An and Hwang, 2014). As toxicogenomics studies aim to elucidate the relation between toxicant exposure and genome-wide gene expression patterns, miRNAs play an important role as key mediators in toxicological research (Lema and Cunningham, 2010). miRNA profiling associated with environmental toxicants, industrial chemicals, and drugs have been studied on different organisms such as humans, mice, and rats (Marsit et al., 2006; Fukushima et al., 2007; Pogribny et al., 2007; Sathyan et al., 2007; Shah et al., 2007). While most of these studies apply qPCR or microarray techniques, RNA sequencing (RNA-seq) has gained a larger role in miRNA profiling due to its use of next-generation sequencing, namely miRNA-seq.

RNA-seq has several advantages over conventional profiling techniques such as qPCR and microarrays, particularly as its cost has lowered in recent years. For example, RNA-seq is not bound with the pre-defined genes to be assayed, which increases the number of genes to be detected and reveals new transcripts (Mestdagh et al., 2014). In addition, RNA-seq has been demonstrated with low variation between platforms compared to the conventional techniques (Consortium, 2014). Furthermore, RNA-seq has a better dynamic detection range (Zhao et al., 2014). However, the choice of analysis methods has been a crucial step for RNA-seq data analysis and subsequent biological interpretation. In-depth analyses of RNA-seq pipelines for mRNA expression analysis were reported but a comprehensive analysis of various bioinformatics pipelines for miRNA-seq applied in toxicogenomics study has yet been conducted. A toxicogenomics study design usually involves both dose- and time-dependent features. It is important that the choice of miRNA-seq analysis methods should not shadow the dose and time-dependent treatment effect. Therefore, this study aims to assess various miRNA pipelines for miRNA profiling exposed to a toxicant with multiple dose and time points. Of note, this study is to assess the joint effect of various parameters involving mapping, quantification, and selection of the profiling tool, rather than novel discoveries and isoform identification.

There are several tools for miRNA-seq data analysis and some have been evaluated for different purposes in previous studies. Li et al. (2012) compared eight tools: miRDeep (Friedlander et al., 2008), miRanalyzer (Hackenberg et al., 2009), miRExpress (Wang et al., 2009), miRTRAP (Hendrix et al., 2010), DSAP (Huang et al., 2010), mirTools (Zhu et al., 2010), MIReNA (Mathelier and Carbone, 2010), miRNAkey (Ronen et al., 2010), and mireap (Sourceforge, 2015). They selected three data sets (*Caenorhabditis elegans*, *Gallus gallus*, and human embryonic stem cells) to evaluate the sensitivity, accuracy and potential for novel miRNA discovery of these tools. They provided a preference list of tools for these three organisms in predicting novel miRNAs. To observe miRNA response to acute nerve crush, Metpally et al. (2013) conducted an analysis on miRDeep2 (Friedländer et al., 2012), miRExpress, and miRNAkey which was followed by the computation of DEMs by DESeq (Anders and Huber, 2010) and EdgeR (Robinson et al., 2010). miRDeep2 was consistently reported to identify the highest number of DEMs. Whereas the former study was focused on novel miRNA discovery in comparison, the latter measured the DEM concordance of tools and provided novel miRNA predictions for a neurological disease. However, to the best of our knowledge, such an extensive evaluation on bioinformatics choices for toxicogenomics data has not been conducted. It is also important to note that miRDeep2 is the evolved version of miRDeep while sRNAbench has recently become the latest version of miRanalyzer. Their parameters in the earlier versions were not extensively studied to confirm whether such changes cause any major downstream perturbations. This not only necessitates revisiting the current versions of these tools and comparing with other tools, but also provides the opportunity for a new study design with a toxicogenomics focus.

We propose a toxicogenomics study design that is based upon time-dose resolution of a toxic treatment to addresses questions. The study conducted focused on the parameter choices of selected tools for miRNA profiling as it may have an impact on DEMs that plays a central role in toxicogenomics for mechanistic interpretation of toxicity. Whereas an ideal design would require a comparison based on a ground truth set of DEMs, we introduced a method that used the number of DEMs as a means in the absence of such ground truth. In specific, we examined the variability in the number of quantified miRNAs and detected DEMs under certain conditions for selected profiling tools: miRDeep2, sRNAbench, miRExpress, and miRNAkey. For a comprehensive analysis including time-dose response, we employed miRNA-seq data acquired through rat liver samples that were treated by thioacetamide, a well-known carcinogen (Fitzhugh and Nelson, 1948), in four time points and at three dose–levels with their concurrent control. The dataset allowed us to measure the effect of treatment as we changed the parameters of the tools as well as the normalization methods. Especially important for downstream analysis was the use of change in the number of DEMs as a measure to assess the stability of a tool as well as the competition between treatment and parameter combinations.

MATERIALS AND METHODS

Data Acquisition and Processing

The data used in this analysis is reported from our lab (Dweep et al., 2017) and available in the GEO repository (GSE87446)[1]. Briefly, thioacetamide was administered to rats for 3, 7, 14, and 28 days with 4.5, 15, and 45 g/kg, daily. On a concurrent basis, another group of healthy rats were kept as a control group. Being the main target organ in this project, liver samples obtained from both the treated and the control groups were further processed for RNA isolation and the isolates exceeding the RNA integrity number (9.0) were sent out for sequencing.

Following the separation of miRNAs from other RNAs, sequencing was conducted on the Illumina HiSeq 2500 platform and sequencing data were de-multiplexed. Resulting fastq files were inspected for quality control purposes. Lanes having low sequence depth were discarded, which in turn caused the loss of one control sample for 3 days. Other discarded lanes only led to the excise of some technical replicates, but did not cause any further exclusion of time-dose points and control samples. Average Phred scores for the sustained lanes varied from 38 to 40. Before employing the profiling tools on the reads, we trimmed the Illumina 3′ adapter by using *fastx_clipper* in the FASTX-Toolkit (Metpally et al., 2013).

miRNA Profiling Tools

Several miRNA-seq tools have been reported with different underlying algorithms and running environments, i.e., web server vs. a stand-alone application. These tools usually provide both expression profiling and miRNA prediction. In this study, we selected four popular tools that can run on a local system since uploading big data sets to the web-based tools is time-consuming and brings further limitations. We used the 21st release of miRBase (Kozomara and Griffiths-Jones, 2014) as the reference database for all tools in the alignment step. In **Table 1**, we summarize the features of the tools by also giving the applicability information of an option and the number of choices we examined for parameters. Most variations are for windowing (more than four options for each tool) and quantification (two options for each tool).

During the biogenesis of miRNAs, there might be errors in the Dicer process and therefore some mismatches can be tolerated out of the mature sequence during mapping. While sRNAbench introduces the terms *WinUp* and *WinDown*, miRDeep2 allows such tolerance by –e and –f parameters in the *quantifier.pl* module. These parameters refer to a window size in both directions of mature sequence and we used windowing as a profiling parameter that took values in {0, 3, 5} and {0, 3} for 3′- and 5′-UTR directions, respectively. Hence, windowing introduces six combinations for the profiling step. However, this was not applicable to miRNAkey and miRExpress. While miRNAkey did not support any extensions, miRExpress worked with the total window size regardless of the direction. For instance, if the window size (or tolerance parameter –g) is set

to 8, miRExpress might be allowing 2 nts from 5′ UTR and 6 nts from 3′ UTR. Alternatively, it might be tolerating 8 nts from 5′ UTR, but none toward 3′ UTR. Finally, those six combinations were only tested for miRDeep2 and sRNAbench as a part of their evaluations.

Read fragments that can be mapped to multiple miRNAs are handled differently by all four tools. In this step, tools distribute mapped reads to known miRNAs in their own way and output the expression values accordingly. miRNAkey employs SEQ-EM for additional preferences for the allocation of multiply aligned reads. In this step, we had both the multiple mapping and single mapping values for sRNAbench and miRDeep2 which were computed on the basis of their outputs.

miRDeep2

miRDeep2 (Friedländer et al., 2012) is a software package which consists of Perl scripts and is the updated version of miRDeep. The use of its components allows both expression profiling and identification of novel miRNAs. Preprocessing of the reads is accomplished through its *mapper.pl* script while quantification is achieved by *quantifier.pl*. For the novel miRNA identification, *mirDeep2.pl* is utilized. miRDeep2 offers various options for each script. Since the underlying alignment tool is Bowtie, alignment parameters can either be set on the command line or modified in the source code. Default parameters consider mismatches outside the mature sequence for both 3′ and 5′ directions through "windowing," that is, tolerating mismatches from both ends. miRDeep2 also introduces further flexibility when counting the mapped reads in quantification, i.e., single mapping, multiple mapping, and unique counts.

sRNAbench

Similar to miRDeep2, sRNAbench is also an improved version of an earlier tool, i.e., miRanalyzer, in which novel features, such as genome and library mapping, have been added. In addition, its prediction capability has been improved for novel miRNAs and isomiR support. It consists of different modules such as preprocessing, mapping, and profiling/detection. For the mapping module, it also employs Bowtie. With a variety of output options, it can also provide a differential expression analysis by using EdgeR. With the parameters *WinUp* and *WinDown* it provides the windowing option that sets a tolerance value for the mismatches toward 3′- and 5′-UTR. Moreover, the user has the option to incorporate genome mapping in the pipeline.

miRExpress

Unlike miRDeep2 and sRNAbench, miRExpress (Wang et al., 2009) employs the Smith-Waterman (SW) algorithm (Smith and Waterman, 1981) for the alignment. It comes with different constituents including preprocessing steps such as adapter removal. miRExpress also differs in that it does not need genome mapping, directly mapping the reads to the known miRNAs documented in miRbase instead. While the tools mentioned above allow for a tolerance value in both directions of the mature sequence, miRExpress only allows a total number of mismatches from either direction. That is, the user sets a window size through

---

the windowing option and miRExpress tolerates any number of mismatches that do not exceed this size.

miRNAKey
Another profiling tool which directly maps the reads to known miRNAs is miRNAKey (Ronen et al., 2010). It comes with a stand-alone Java package that can be used either on the command line or through a GUI. It further allows for adapter removal and performs alignment by using BWA (Smith and Waterman, 1981). As for the counting option in **Table 1**, it utilizes the SEQ-EM algorithm (Pasaniuc et al., 2011) to optimize the distribution of multiply aligned reads.

Applicability of Parameters across Tools
Although the tools mentioned above have unique features, they also have similar parameters that can be compared to some extent. While tool-specific parameters were used for internal evaluation of each tool by measuring the variability and treatment effect in the number of DEMs, compatible ones were utilized for a cross-tool comparison. Windowing and quantification with two counting options (single and multiple) served as two factors for a side-by-side comparison of miRDeep2 and sRNAbench.

Side-by-Side Comparison of sRNAbench and mirDeep2
Since both tools detected around 500 miRNAs with expression values greater than zero, we ranked miRNAs in a decreasing order based on those quantifications for every parameter combination. Thus, we compared the agreement between these tools at each increment of 5 in the ordered lists under the same conditions. This procedure enabled us to follow the concordance in terms of percentages as we cover the whole list for a given parameter set.

We conducted a similar analysis for the DEMs obtained from different tools under the same parameter set and normalization method (*upper quartile*) in the EdgeR package. However, the number of DEMs was much lower than the detected miRNAs and varied drastically from one treatment to another. For these reasons, we performed this analysis for the treatment points for which we obtained more than 35 DEMs and considered miRNAs with a *p*-value less than 0.05 regardless of their fold changes.

Pairwise Concordance of Tools
In this part of the study, we measured the agreement between tools by computing the Jaccard similarity. In other words, we normalized the overlapping quantity by the union of two miRNA sets that were obtained using different parameters. In the

quantification stage, 24 parameters (2 tools × 6 windowing options × 2 quantification options) were used to achieve this pairwise evaluation. The resulting similarity values were used to generate hierarchical clusters that provide a visual aid to distinguish the impact of parameter selections.

Analysis of Variance for Quantified and Differentially Expressed miRNAs
Two profiling elements (windowing and quantification) have become factors with different levels for the analysis of variance (ANOVA). Due to the available options for windowing, its effect was measured based on six levels that will be detailed later. Quantification, however, was limited to two levels. In this particular design, our dependent variables were the number of DEMs for which we measured the impact of each profiling element in terms of variance. In the following steps, we included the selection of the tool as another factor, which accounted for the effect of any tool on the variance. Then, we incorporated treatment elements as another set of factors. Specifically, we defined three more variables: (i) time, (ii) dose, and (iii) time-dose interaction, which introduced 4, 3, and 12 levels, respectively. In order to see the normalization effect in the number of DEMs, we further added normalization into the equation with three methods [TMM (Robinson and Oshlack, 2010), RLE (Bullard et al., 2010), and upper quartile (Bullard et al., 2010)] in EdgeR while also observing the changes without any normalization (NO).

Time-Dose Response
All the parameters above are expected to have an effect on the downstream. Beside the magnitude of such an effect, we are also interested in whether the trend in time-dose response is affected by the choice of parameters. Therefore, we calculated the correlation in the number of DEMs across different treatments for every pair of parameter combinations.
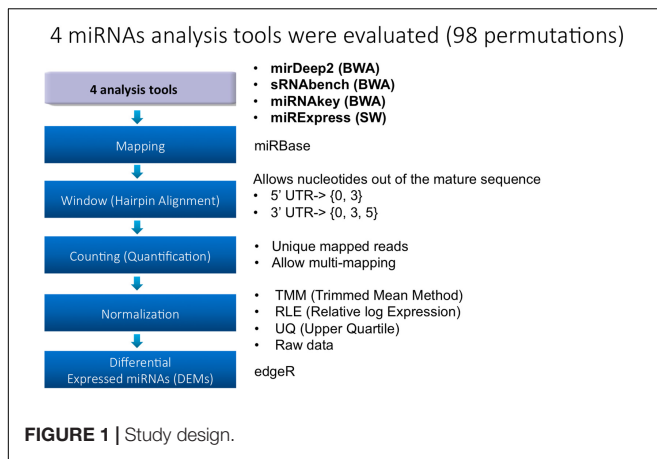
RESULTS

Study Design
Rat liver miRNA data from thioacetamide treatment at multiple doses and time points provided a two dimensional resolution of the treatment effect in both dose and time directions. As such, the number of DEMs across treatments has become a means for assessing the parameter effect and its association with the treatment pattern. As depicted in **Figure 1**, four analysis tools

TABLE 1 | Summary of tool features along with their options.

	Alignment algorithm	Genome mapping	Windowing	Quantification	Total parameter
miRDeep2	Bowtie	Y	Y(6)	Y(2)	12
sRNAbench	Bowtie	Y	Y(6)	Y(2)	12
miRExpress	SW	NA	Y(4)	NA	4
miRNAkey		NA	NA	Y(2)	2

SW, Smith–Waterman algorithm; Y, yes; NA, not applicable.

FIGURE 1 | Study design.

FIGURE 2 | Variance analysis on the number of differentially expressed miRNAs (DEMs) produced by each tool.

were selected. For each tool, parameter choices constituting the next three steps (i.e., mapping, hairpin alignment and counting) introduce further options. For each treatment point, we employ the same parameter combinations for miRNA quantification and DEM detection. Due to the incompatibility issues, 30 parameter combinations shown in **Table 1** are tested based on their applicability through the pipeline in **Figure 1**. Once the miRNAs are quantified, each tool is evaluated within its own parameter space by measuring its sensitivity in the number of DEMs and the impact on the treatment effect. As described below, these evaluations are used as selection criteria for further investigation of miRDeep2 and sRNAbench. Finally, we compare these two tools along with their parameter combinations in terms of miRNA quantification and DEM detection that introduces four normalization choices.

Parameter Setting Introduced a Large Variation in miRExpress and miRNAKey

Selection of a tool is followed by the appropriate option to be employed with parameter manipulations for miRNA profiling as listed in **Figure 1**. For instance, the user needs to decide whether a tolerance value should be set for mismatches in 5′ UTR and 3′ UTR if applicable. Here, we introduce these parameters by discussing the applicability issues.

Each tool comes with own its features, underlying algorithm, and parameter set. Aiming for a comparable analysis of parameter combinations, we considered all parameters that can be manipulated and observed their responses (in terms of number of DEMs) to such changes. For a fair comparison, we performed the DEM calculations under the same normalization method (upper quartile) in EdgeR. Then, we computed the variance in the number of DEMs as we changed the parameters of each tool by which we obtained the coefficient of variation (CoV). It was observed that miRExpress showed higher CoV due to the change in its tolerance (windowing) parameter. Since it was not employing any other parameters, only the tolerance value for windowing was tuned and it was found to be more sensitive to the changes (**Supplementary Figure S1**).

miRNAkey had a comparable CoV with sRNAbench mainly due to its SEQ-EM option when counting the mapped reads.
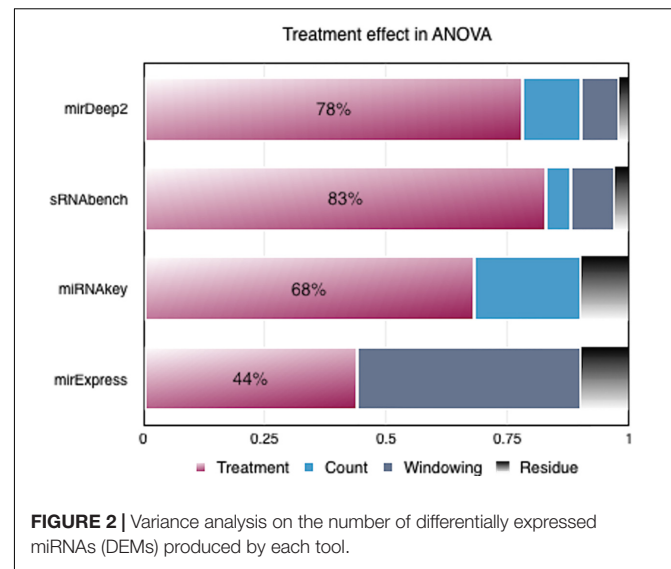
However, when we performed ANOVA analysis on the number of DEMs to compare the impact of tools with their parameters on the treatment effect, we observed that both miRExpress and miRNAkey shadowed the treatment effect more as illustrated in **Figure 2**.

On the other hand, in the case of miRDeep2 and sRNAbench, neither their individual parameters, nor their combined parameters dominated the treatment effect in terms of variance. Therefore, we chose miRDeep2 and sRNAbench for an in-depth investigation of parameter effect on the number of DEMs, which is an essential part of downstream analysis. Compatible options between these two tools also gave us the opportunity to perform further variance analyses.

## miRNA Quantification by sRNAbench and miRDeep2
Concordance for sRNAbench and miRDeep2

In this study, we considered a miRNA that has a non-zero expression value as detected. Then, we compared the total number of detected miRNAs for both tools and computed the overlaps in the ranked lists based on expression values. For varying profiling parameters, we counted the number of miRNAs that have an expression value greater than zero (**Supplementary Table S1**). This analysis showed that sRNAbench mostly reported more miRNAs than miRDeep2 (**Figure 3A**). Nevertheless, they had a high overlap of around 95% for each parameter combination. For each treatment point, we compared their sorted lists which were in decreasing order w.r.t. expression values. Excepting their top positions, which still showed an agreement around 80%, both tools were able to capture the same miRNAs within the top 500 candidates (**Figure 3B**).

Variance Analysis

In the analyses above, we presented the agreement across different bioinformatics choices pictorially. However, a holistic view is still needed to quantify the level of such impact and to
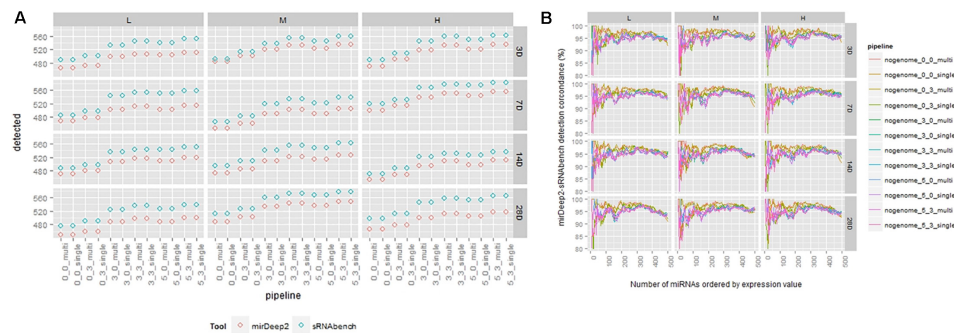
FIGURE 3 | (A) Number of detected miRNAs for each tool under the same pipeline parameters. **(B)** Detection comparison in sorted lists from sRNAbench and miRDeep2 under the same parameter sets.

further assess the significance of a choice against treatment effect or toxic exposure. Therefore, we performed ANOVA for various scenarios.

For the detection task, we performed ANOVA for two cases based on: (i) profiling parameters, (ii) profiling parameters with treatment components such as dose, time and dose-time interaction. In the first case, the choice of tool dominated the windowing effect that was also very effective in detecting the miRNAs (**Figure 4A**). Once we incorporated the treatment elements, their shares decreased, but their impact was still more than treatment factors (**Figure 4B**).

Differentially Expressed miRNAs by sRNAbench and miRDeep2
Concordance for sRNAbench and miRDeep2
The number of DEMs was much smaller than the detected miRNAs and, for some cases, such as 3 days with medium dose, the number dropped down to 1. In order to get reasonable statistics, we excluded those cases having less than 35 DEMs and considered only *p*-values less than 0.05 as a DEM criterion as indicated in Section "Materials and Methods." Thus, we represented 11 treatments out of 12 in **Figure 5** in which the level of agreement between tools under the same parameters was illustrated in terms of percentages. Due to the difference in prioritization in the DEM lists acquired through miRDeep2 and sRNAbench pipelines, the overlapping ratio started at 40% in top 5 DEMs (2 DEMs were in common in top 5). As we went through the rest of the lists, we observed an increasing agreement which even reached 100% and converged around 80%.

As both tools prioritize different miRNAs either in the detection step or at the DEM computation stage, their pipelines produce nearly the same amount of DEMs whereas there was a bigger gap in their detection levels. More interestingly, we also observed treatments where sRNAbench suggests fewer DEMs than miRDeep2, although sRNAbench consistently detected more miRNAs. In **Figure 6**, we present DEM set sizes for upper quartile normalization in which DEM set sizes are close, but they do not necessarily overlap 100%. We have similar observations for other normalization choices (**Supplementary Figure S2**).

In **Figure 3A**, both tools show an increasing number of detected miRNAs as choices involve wider windows which tolerate more mismatches beyond the mature sequence. This trend is not preserved for the number of DEMs as illustrated in **Figure 6**. Indeed, there are cases, such as 14 days with medium dose, where DEM set size is almost constant even though profiling parameters change. This demonstrates the fact that miRNAs detected incrementally due to parameter change may not be differentially expressed.
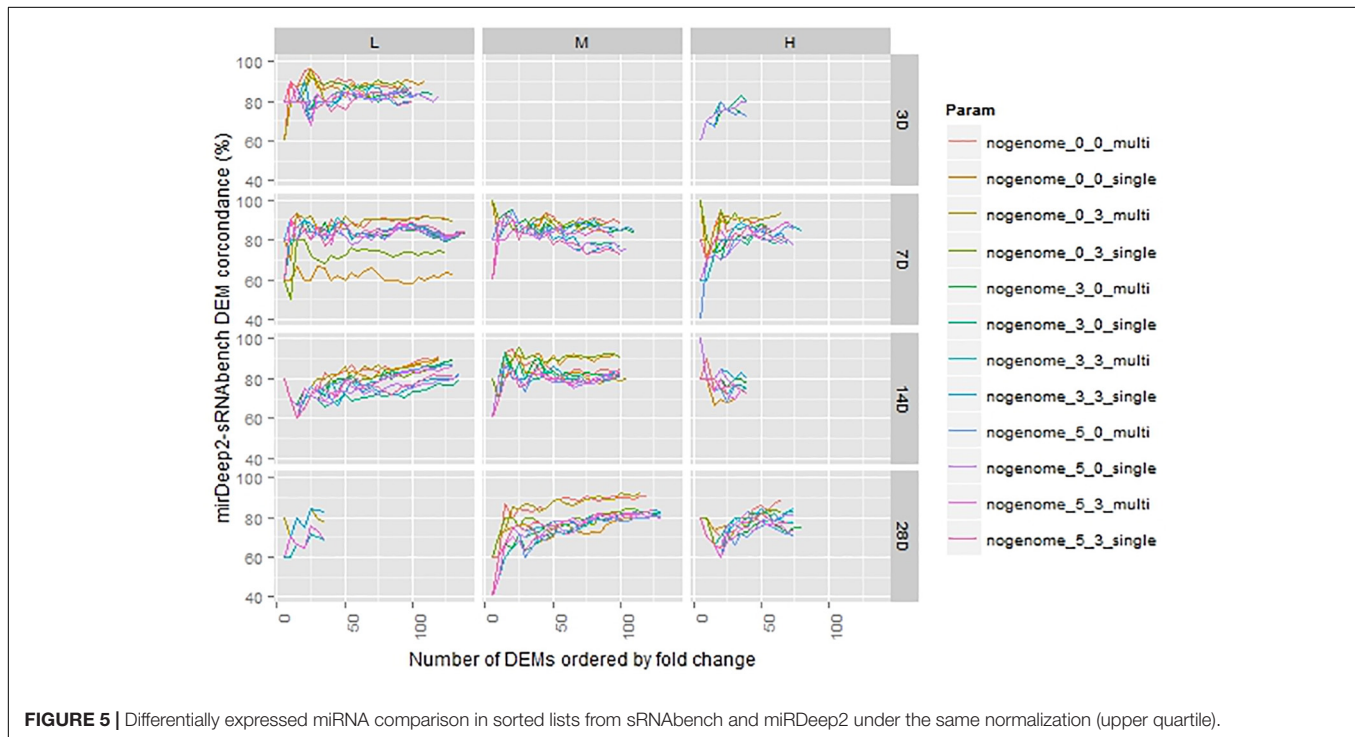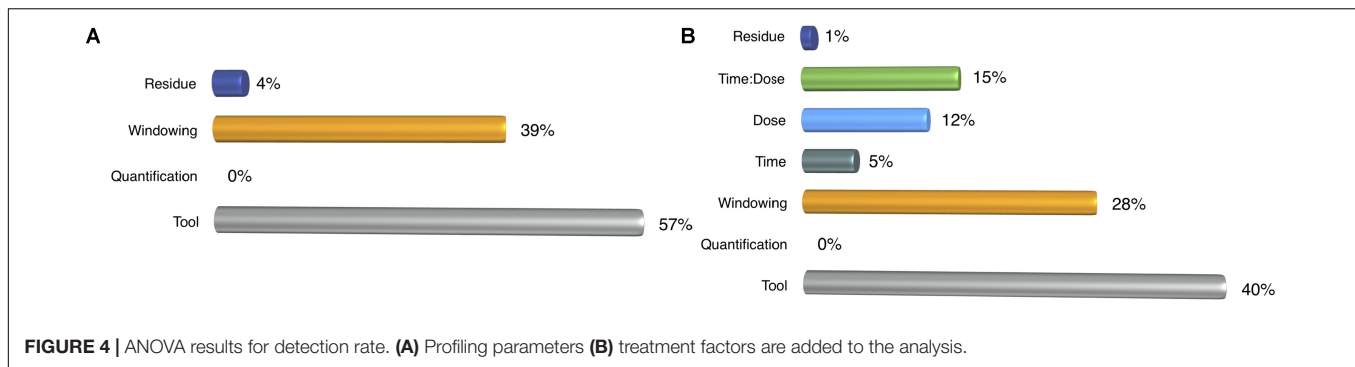
Variance Analysis
We conducted ANOVA on the number of DEMs to have a better understanding of the parameter effect on downstream analysis. Similar to the ANOVA in the detection step, we incrementally carried out our analysis by incorporating additional parameters. In the first analysis, we only measured the impact of profiling parameters on the number of DEMs. It resulted that selection of the tool and the windowing option sustained their effects as they had in the detection rate, but ANOVA resulted in high residue (**Figure 7A**), which might be due to some other factors such as dose and time. Therefore, in the next step, we wanted to determine the effect of profiling parameters and treatment factors under a single normalization method, which was randomly chosen to be upper quartile. In **Figure 7B**, it is clearly observed that the share of the profiling parameters in the variance does not exceed 8% which indicates that the treatment has a much stronger impact on the number of DEMs and implies freedom of choice in the selection of pipeline. However, normalization introduces a huge variance that competes with that of treatment effect (**Figure 7C**). Even though profiling elements still have a minor effect on the number of DEMs, normalization drastically changes the number of DEMs as much as the dose level. Furthermore, it dominates the impact of treatment duration and therefore implies that it is a crucial step for downstream studies. When we excluded un-normalized data from our analysis, we still observed the same outcome (**Supplementary Figure S3**).

Pairwise Concordance of Possible Bioinformatics Choices
After comparing each tool under the same parameter combinations, we also made a comparison that took into

FIGURE 4 | ANOVA results for detection rate. **(A)** Profiling parameters **(B)** treatment factors are added to the analysis.

FIGURE 5 | Differentially expressed miRNA comparison in sorted lists from sRNAbench and miRDeep2 under the same normalization (upper quartile).

account all possible choices including choice of the tool itself as well as the normalizations for the DEM step. More specifically, we performed three types of analyses that rely on Jaccard similarities for detection rate, DEM calculation under upper quartile normalization, and DEM calculation with additional normalization choices in EdgeR.

Initially, we had 24 choices that included the selection of any tool with windowing and quantification options. For each treatment, we performed hierarchical clustering which illustrates agreement across different pipelines in terms of detection concordance. It was found that windowing was the major factor in determining the clusters of bioinformatics choices (**Supplementary Figure S4**). While bioinformatics choices having a tolerance value of 5 in the downstream were consistently clustered together regardless of their upstream end, those with exact match (0, 0) formed another cluster. Depending on the dose, bioinformatics choices with intermediate tolerance values joined one of the clusters determined by those extreme

parameters. In all cases, neither sRNAbench nor miRDeep2 were split under the same parameters, showing the power of windowing parameter over all other parameters from the concordance point of view.

For the same set of parameters, we obtained clusters of bioinformatics choices based on DEM concordance for each treatment. Under the same normalization method, i.e., UQ (upper quantile), we observed that clusters were mainly built by the choice of tool and windowing (**Supplementary Figure S5**). Including 3 days with medium dose, in which we did not have a high enough number of DEMs for such analysis, pipelines belonging to the same tool form the bigger clusters and windowing determines the subclusters.

In the third step, we increased the number of bioinformatics choices by incorporating normalization methods that resulted in four times larger dendrograms than the earlier analyses. In specific, we either worked with un-normalized data (NO) or used
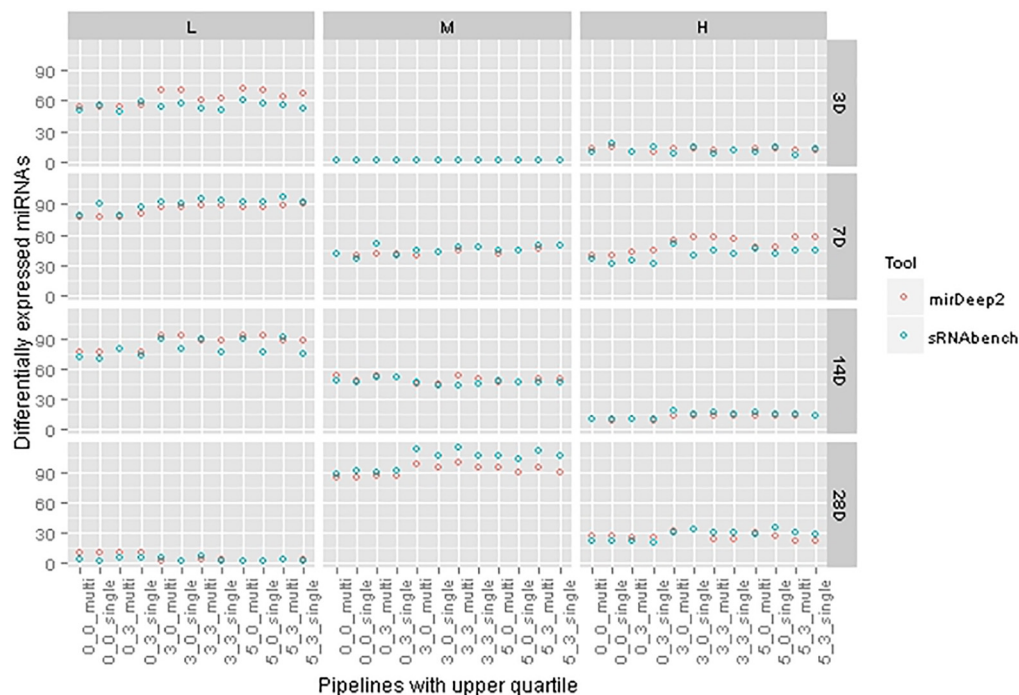
FIGURE 6 | Differentially expressed miRNA set sizes for each tool under the same pipelines with upper quartile normalization.
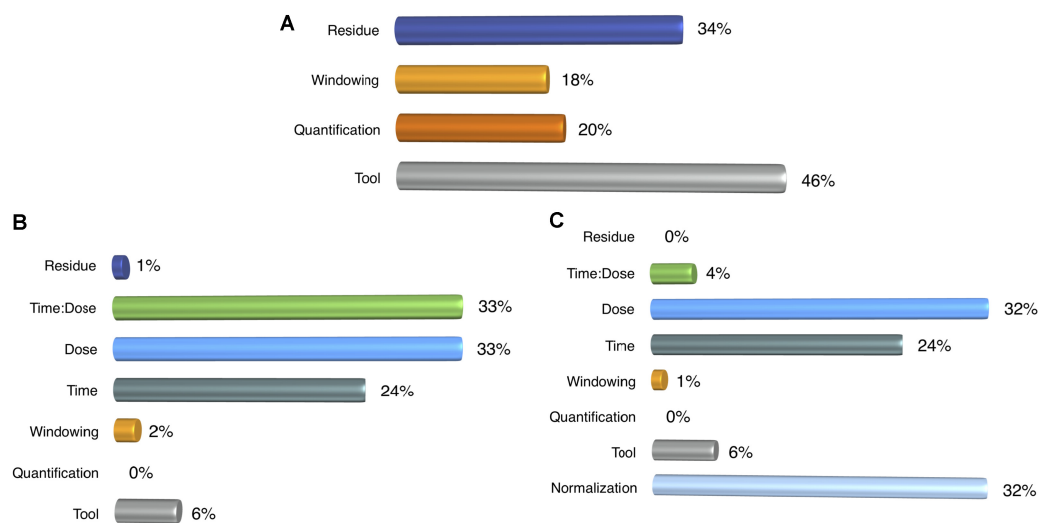

FIGURE 7 | ANOVA results for the number of DEMs. **(A)** Impact of profiling parameters, **(B)** Profiling parameters and treatment components were compared and **(C)** Normalization methods were added on the top of earlier ANOVA designs.

one of the normalization methods that come with EdgeR, i.e., UQ, LRE, and TMM. The selection of a tool was still found to be a primary reason to form clusters, which implies that DEMs obtained from a tool are mostly shared by bioinformatics choices (**Supplementary Figure S6**). Even though normalization methods determine the second level of clustering for most of the cases, windowing still preserves its potential to affect the outcome of DEM calculations. Therefore, there are also clusters

in which pipelines using different normalization methods are grouped together.

Parameter Sensitivity for sRNAbench and miRDeep2

Through hierarchical clustering, we presented the Jaccard similarities of all possible parameter combinations including the tools. We further utilized these similarities for the DEM sets
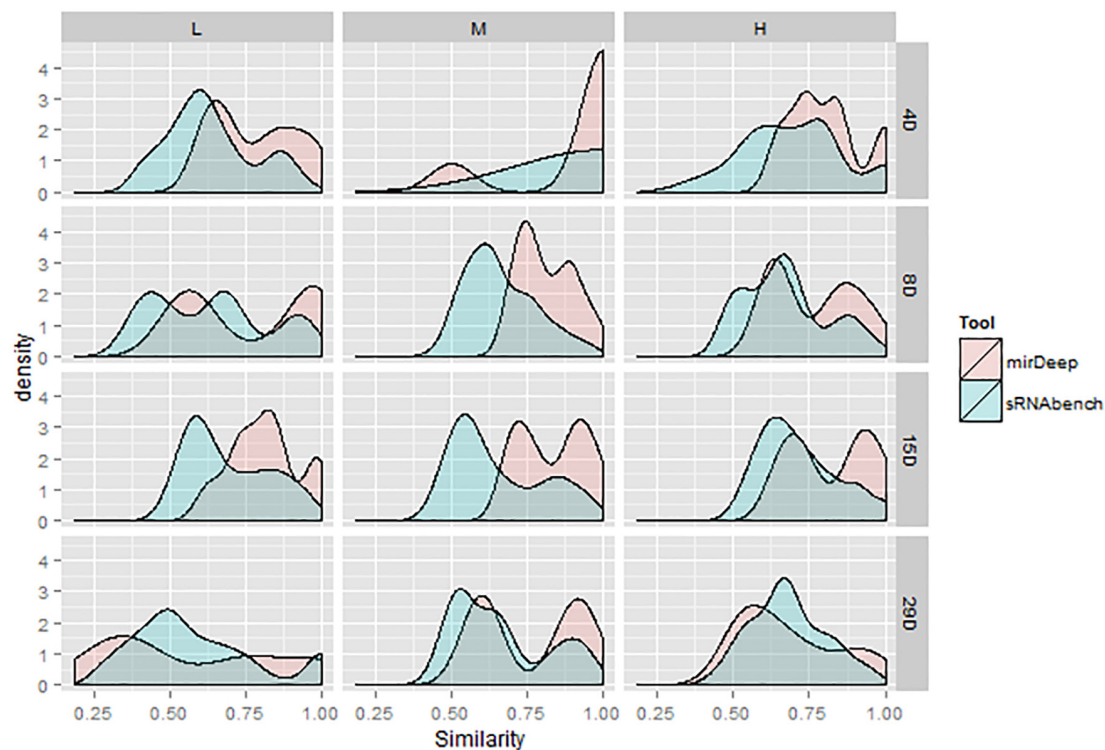
FIGURE 8 | Density curves for similarity values obtained from the pipelines within each tool.

under upper quartile normalization to illustrate the sensitivity levels of tools for given parameter changes. We then plotted the density curves for the similarity values, which effectively outline the accumulation points. A peak around a high similarity value means that there are many parameter combinations giving very similar outcomes under the same tool. As illustrated in **Figure 8**, similarities for miRDeep2 mostly start at higher values. In addition, peaks are shifted toward the right, which might be an indication that miRDeep2 is less sensitive to parameter changes. However, when it comes to the 28th day with high dose, which is the strongest treatment, we observed similar behavior within both tools.

Consistency in the Time-Dose Response

Both the concordance and the ANOVA results showed that downstream analysis was significantly affected by the choice of normalization. Nevertheless, the pattern in the number of DEMs across treatments needed to be investigated to elucidate the treatment effect (which we found to be competing with the normalization effect). We already discovered that number of DEMs fluctuated due to the normalization choice, but we wanted to determine whether this fluctuation would follow a pattern as we proceeded in time and dose. In order to capture a possible pattern, we computed the correlation in the number of DEMs across different time and dose combinations. In other words, we picked every possible pair of parameter combinations and compared the number of DEMs resulting from these pipelines for 12 treatment points. We found that

correlation scores start from 0.92 even for the bioinformatics choices that not only differ in profiling elements, but also in their normalization step. These high correlations imply that this trend is sustained due to the treatment effect, although the numbers are drastically affected by the change in the normalization method.

DISCUSSION

Techniques such as qPCR and microarray have been widely used to assess the role of miRNAs in toxicogenomics research. With the advance of sequencing technologies, NGS has become a popular means for miRNA profiling and has led to novel discoveries. Thus, miRNA-seq also stands as a promising tool for toxicogenomics experiments where linking toxicants to mRNAs through miRNAs helps us understand genome-wide alterations. While miRNAs target a considerable portion of mRNAs, it is important to investigate miRNA profiling tools and their parameters to obtain a robust number of DEMs, which will potentially affect protein translation.

In a toxicological context, we examined different tools which can be run as stand-alone applications so that we could study a data set which provided a two-dimensional resolution in time and dose. Having the advantage of 12 treatment points in this particular data set, we pursued a study in which treatment effect can be a measure to analyze the miRNA-seq profiling tools. Initially, we started with four tools (miRNAkey,

miRExpress, sRNAbench, and miRDeep2) whose stability was observed in terms of the number of DEMs by manipulating their own parameters. We found that miRExpress showed the most fluctuating number of DEMs as we changed its windowing (tolerance) parameter. sRNAbench and miRNAkey showed close variances, but the treatment effect was shadowed in terms of variance in the choice of miRNAkey. Thus, sRNAbench and miRDeep2 were selected to inspect the parameter effect on downstream analysis.

In our experimental setting, the number of DEMs was considered to be the indicator of the downstream effect of a toxicant. Therefore, our measurements or observations mostly relied on the change across time and dose, which was assumed to be an indicator of the treatment signal or toxic effect of thioacetamide. We then interpreted such changes to assess the magnitude of the pipeline effect on downstream analysis.

Since the detection rate is the only measure that does not need the DEMs, our first analysis relied on the number of miRNAs which had non-zero expression values. Those miRNAs, regardless of their expression levels, were considered as detected. Since the expression level was not our focus, miRNA sets from different bioinformatics choices showed high overlaps. Furthermore, ANOVA did not attribute much variance to the treatment effect. In fact, the choice of the tool played the most significant role. This observation is expected as the presence of a miRNA is not necessarily due to the toxic effect. Rather, differences in the expression levels between groups can statistically account for such an effect. For this reason, we delved deeper into DEM analysis.

In the subsequent DEM analyses, we observed that highly overlapping bioinformatics choices in the detection stage started to show differences which in turn led to decreasing agreement between bioinformatics choices. These discrepancies might be due to the selection of a profiling tool or its interaction with the parameters. Therefore, further research is warranted to investigate which tool can lead to more accurate DEMs for downstream analysis by examining the tool-specific DEMs. Current measurements, which were illustrated with the density curves in **Figure 8**, favored miRDeep2 by indicating that it is less sensitive to parameter changes. Namely, different parameter combinations resulted in very similar DEM sets for mirDeep2. Nevertheless, the same findings also emphasized that tool selection may not be that significant in the presence of a strong treatment signal. For instance, at the high dose on the 28th day, all the pipelines within each tool produce highly overlapping DEMs.

When we inspected the windowing option, we noticed an increasing number of detected miRNAs as both tools allow more mismatches. However, the number of DEMs remained almost constant and did not show considerable change beyond 3 nts. This may indicate that tolerated mismatches helped in quantifying more miRNAs with low expression values, which were not differentially expressed most of the time.

We performed ANOVA on the number of DEMs without the treatment factors, which indicated that change in DEM was not only due to the selection of the tool. Once the time and dose factors are included in the variance analysis, the treatment effect explained not only the remaining variance, but also demonstrated that the profiling elements were not that effective in downstream analysis. This finding suggests more flexibility in bioinformatics choice under the same normalization method. However, the choice of normalization can be as dominant as the treatment effect as it was shown by another ANOVA. Thus, the normalization step may require more attention than the earlier steps in miRNA-seq profiling.

The similarity between parameter combinations decreased gradually as we proceeded toward downstream analysis, especially for those that employ different normalization choices. Whereas normalization causes fluctuations in the numbers, the treatment effect enforces a pattern on these fluctuations. Namely, high correlations between parameter combinations prove the persistence of the toxic effect of thioacetamide in terms of a fluctuation pattern. Finally, the DEM sets present a diverging characteristic with lowering overlaps, but this characteristic does not attenuate the treatment signal, which is the most valuable information to be observed in an experiment.

Even though we were able to reach ~80% overlap between DEM sets from different pipelines, one of the limitations of our study remains the lacking ground truth for toxic implications of discovered DEMs. Under this constraint, we developed our study design to evaluate different miRNA profiling tools by solely relying on the number of DEMs, which are an essential part of the downstream analysis. While different levels of discrepancies were observed between tools and protocols, focusing on the size of DEM sets and their concordances provided the opportunity to assess consistency. Namely, we not only measured the robustness of each tool, but also monitored whether the dose- and time-dependent pattern of DEM sizes would remain the same regardless of the protocol. In the absence of a ground truth set, we believe our methodology and findings provide an insight in pipeline selections not only for toxicogenomics studies, but also for other research endeavors in the community.

CONCLUSION

miRNAs are important regulatory elements which are also sensitive to toxicants and require well-established bioinformatics choices for expression analysis. With a toxicogenomics focus, we studied different tools with their parameter combinations and their impact on the downstream analysis in terms of DEM sets. Our results indicated that the selection of tools and parameter manipulations could cause a limited difference in DEM sets and their variability, whereas the choice of normalization method might have more impact on concordances. We further demonstrated that the treatment effect was highly preserved despite the discordances in DEM sets, concluding that biological variance overcomes other factors in parameter choices, especially in the presence of a strong treatment signal. With different detection sensitivities, sRNAbench and miRDeep2 produced a close number of DEMs. However, mirDeep2 demonstrated lower sensitivity to parameter changes, which makes miRDeep2 preferable over other choices when windowing allows up to three nucleotides from both ends.

AUTHOR CONTRIBUTIONS

HB conducted all the analysis and wrote the first draft. BG contributed to the part of analysis. YW provided the dataset for this study. WT conceived the study design and finalized the manuscript. All authors read and approved the final manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2018.00022/full#supplementary-material

FIGURE S1 | Basic statistics for number of DEMs for each tool.

FIGURE S2 | Number of detected miRNAs and DEMs for each normalization method from sRNAbench and mirDeep2 pipelines.

FIGURE S3 | ANOVA on number of DEMs without normalization.

FIGURE S4 | Hierarchical clusterings based on detected miRNAs from sRNAbench and mirDeep2 pipelines for 3 and 7 days.

FIGURE S5 | Hierarchical clusterings based on DEMs from sRNAbench and mirDeep2 pipelines using upper quartile normalization for 3 and 7 days.

FIGURE S6 | Hierarchical clusterings based on the DEMs from sRNAbench and mirDeep2 pipelines using different normalization methods.

TABLE S1 | List of all pipelines with the number of detected and differentially expressed miRNAs.

REFERENCES

An, Y. R., and Hwang, S. Y. (2014). Toxicology study with microRNA. *Mol. Cell. Toxicol.* 10, 127–134. doi: 10.1007/s13273-014-0014-y

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biol.* 11:R106. doi: 10.1186/gb-2010-11-10-r106

Bullard, J. H., Purdom, E., Hansen, K. D., and Dudoit, S. (2010). Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. *BMC Bioinformatics* 11:94. doi: 10.1186/1471-2105-11-94

Chen, Y., and Verfaillie, C. M. (2014). MicroRNAs: the fine modulators of liver development and function. *Liver Int.* 34, 976–990. doi: 10.1111/liv.12496

Consortium, S. M.-I. (2014). A comprehensive assessment of RNA-seq accuracy, reproducibility and information content by the Sequencing Quality Control Consortium. *Nat. Biotechnol.* 32, 903–914. doi: 10.1038/nbt.2957

Cortez, M. A., Bueso-Ramos, C., Ferdin, J., Lopez-Berestein, G., Sood, A. K., and Calin, G. A. (2011). MicroRNAs in body fluids–the mix of hormones and biomarkers. *Nat. Rev. Clin. Oncol.* 8, 467–477. doi: 10.1038/nrclinonc.2011.76

Dweep, H., Morikawa, Y., Gong, B., Yan, J., Liu, Z., Chen, T., et al. (2017). Mechanistic roles of microRNAs in hepatocarcinogenesis: a study of thioacetamide with multiple doses and time-points of rats. *Sci. Rep.* 7:3054. doi: 10.1038/s41598-017-02798-7

Finch, M. L., Marquardt, J. U., Yeoh, G. C., and Callus, B. A. (2014). Regulation of microRNAs and their role in liver development, regeneration and disease. *Int. J. Biochem. Cell Biol.* 54, 288–303. doi: 10.1016/j.biocel.2014.04.002

Fitzhugh, O. G., and Nelson, A. A. (1948). Liver tumors in rats fed thiourea or thioacetamide. *Science* 108, 626–628. doi: 10.1126/science.108.2814.626

Friedlander, M. R., Chen, W., Adamidi, C., Maaskola, J., Einspanier, R., Knespel, S., et al. (2008). Discovering microRNAs from deep sequencing data using miRDeep. *Nat. Biotechnol.* 26, 407–415. doi: 10.1038/nbt1394

Friedländer, M. R., Mackowiak, S. D., Li, N., Chen, W., and Rajewsky, N. (2012). miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. *Nucleic Acids Res.* 40, 37–52. doi: 10.1093/nar/gkr688

Fukushima, T., Hamada, Y., Yamada, H., and Horii, I. (2007). Changes of micro-RNA expression in rat liver treated by acetaminophen or carbon tetrachloride-regulating role of micro-RNA for RNA expression. *J. Toxicol. Sci.* 32, 401–409. doi: 10.2131/jts.32.401

Hackenberg, M., Sturm, M., Langenberger, D., Falcon-Perez, J. M., and Aransay, A. M. (2009). miRanalyzer: a microRNA detection and analysis tool for next-generation sequencing experiments. *Nucleic Acids Res.* 37, W68–W76. doi: 10.1093/nar/gkp347

Hendrix, D., Levine, M., and Shi, W. (2010). miRTRAP, a computational method for the systematic identification of miRNAs from high throughput sequencing data. *Genome Biol.* 11:R39. doi: 10.1186/gb-2010-11-4-r39

Huang, P. J., Liu, Y. C., Lee, C. C., Lin, W. C., Gan, R. R., Lyu, P. C., et al. (2010). DSAP: deep-sequencing small RNA analysis pipeline. *Nucleic Acids Res.* 38, W385–W391. doi: 10.1093/nar/gkq392

Kozomara, A., and Griffiths-Jones, S. (2014). miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res.* 42, D68–D73. doi: 10.1093/nar/gkt1181

Lema, C., and Cunningham, M. J. (2010). MicroRNAs and their implications in toxicological research. *Toxicol. Lett.* 198, 100–105. doi: 10.1016/j.toxlet.2010.06.019

Lewis, B. P., Burge, C. B., and Bartel, D. P. (2005). Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* 120, 15–20. doi: 10.1016/j.cell.2004.12.035

Li, Y., Zhang, Z., Liu, F., Vongsangnak, W., Jing, Q., Shen, B., et al. (2012). Performance comparison and evaluation of software tools for microRNA deep-sequencing data analysis. *Nucleic Acids Res.* 40, 4298–4305. doi: 10.1093/nar/gks043

Marsit, C. J., Eddy, K., and Kelsey, K. T. (2006). MicroRNA responses to cellular stress. *Cancer Res.* 66, 10843–10848. doi: 10.1158/0008-5472.CAN-06-1894

Mathelier, A., and Carbone, A. (2010). MIReNA: finding microRNAs with high accuracy and no learning at genome scale and from deep sequencing data. *Bioinformatics* 26, 2226–2234. doi: 10.1093/bioinformatics/btq329

Mestdagh, P., Hartmann, N., Baeriswyl, L., Andreasen, D., Bernard, N., Chen, C., et al. (2014). Evaluation of quantitative miRNA expression platforms in the microRNA quality control (miRQC) study. *Nat. Methods* 11, 809–815. doi: 10.1038/nmeth.3014

Metpally, R. P., Nasser, S., Malenica, I., Courtright, A., Carlson, E., Ghaffari, L., et al. (2013). Comparison of analysis tools for miRNA high throughput sequencing using nerve crush as a model. *Front. Genet.* 4:20. doi: 10.3389/fgene.2013.00020

Pasaniuc, B., Zaitlen, N., and Halperin, E. (2011). Accurate estimation of expression levels of homologous genes in RNA-seq experiments. *J. Comput. Biol.* 18, 459–468. doi: 10.1089/cmb.2010.0259

Pogribny, I. P., Tryndyak, V. P., Boyko, A., Rodriguez-Juarez, R., Beland, F. A., and Kovalchuk, O. (2007). Induction of microRNAome deregulation in rat liver by long-term tamoxifen exposure. *Mutat. Res.* 619, 30–37. doi: 10.1016/j.mrfmmm.2006.12.006

Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140. doi: 10.1093/bioinformatics/btp616

Robinson, M. D., and Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* 11:R25. doi: 10.1186/gb-2010-11-3-r25

Ronen, R., Gan, I., Modai, S., Sukacheov, A., Dror, G., Halperin, E., et al. (2010). miRNAkey: a software for microRNA deep sequencing analysis. *Bioinformatics* 26, 2615–2616. doi: 10.1093/bioinformatics/btq493

Sathyan, P., Golden, H. B., and Miranda, R. C. (2007). Competing interactions between micro-RNAs determine neural progenitor survival and proliferation after ethanol exposure: evidence from an ex vivo model of the fetal cerebral cortical neuroepithelium. *J. Neurosci.* 27, 8546–8557. doi: 10.1523/JNEUROSCI.1269-07.2007

Shah, Y. M., Morimura, K., Yang, Q., Tanabe, T., Takagi, M., and Gonzalez, F. J. (2007). Peroxisome proliferator-activated receptor α regulates a microRNA-mediated signaling cascade responsible for hepatocellular proliferation. *Mol. Cell. Biol.* 27, 4238–4247. doi: 10.1128/MCB.00317-07

Smith, T. F., and Waterman, M. S. (1981). Identification of common molecular subsequences. *J. Mol. Biol.* 147, 195–197. doi: 10.1016/0022-2836(81)90087-5

Sourceforge (2015). *MicroRNA Discovery by Deep Sequencing*. Available at: https://sourceforge.net/projects/mireap/ [accessed August 5, 2015].

Vosa, U., Vooder, T., Kolde, R., Fischer, K., Valk, K., Tonisson, N., et al. (2011). Identification of miR-374a as a prognostic marker for survival in patients with early-stage nonsmall cell lung cancer. *Genes Chromosomes Cancer* 50, 812–822. doi: 10.1002/gcc.20902

Wang, W. C., Lin, F. M., Chang, W. C., Lin, K. Y., Huang, H. D., and Lin, N. S. (2009). miRExpress: analyzing high-throughput sequencing data for profiling microRNA expression. *BMC Bioinformatics* 10:328. doi: 10.1186/1471-2105-10-328

Zhang, R., and Su, B. (2009). Small but influential: the role of microRNAs on gene regulatory network and 3'UTR evolution. *J. Genet. Genomics* 36, 1–6. doi: 10.1016/S1673-8527(09)60001-1

Zhao, S., Fung-Leung, W. P., Bittner, A., Ngo, K., and Liu, X. (2014). Comparison of RNA-Seq and microarray in transcriptome profiling of activated T cells. *PLOS ONE* 9:e78644. doi: 10.1371/journal.pone.0078644

Zhu, E., Zhao, F., Xu, G., Hou, H., Zhou, L., Li, X., et al. (2010). mirTools: microRNA profiling and discovery based on high-throughput sequencing. *Nucleic Acids Res.* 38, W392–W397. doi: 10.1093/nar/gkq393

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read
for greatest visibility
and readership

FAST PUBLICATION
Around 90 days
from submission
to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative,
and constructive
peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers
acknowledged by name
on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: info@frontiersin.org | +41 21 510 17 00

REPRODUCIBILITY OF RESEARCH
Support open data
and methods to enhance
research reproducibility

DIGITAL PUBLISHING
Articles designed
for optimal readership
across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics
track visibility across
digital media

EXTENSIVE PROMOTION
Marketing
and promotion
of impactful research

LOOP RESEARCH NETWORK
Our network
increases your
article's readership