

UCLA

UCLA Electronic Theses and Dissertations

Title

Development and Application of Data-Driven Computational Methods for Molecular Toxicology and Environmental Exposure Research

Permalink

<https://escholarship.org/uc/item/8rz9s5jh>

ISBN

9798293826551

Author

Kowal, Thomas Joseph

Publication Date

2025-09-09

Supplemental Material

<https://escholarship.org/uc/item/8rz9s5jh#supplemental>

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Development and Application of Data-Driven
Computational Methods for Molecular Toxicology
and Environmental Exposure Research

A dissertation submitted in partial satisfaction
of the requirements for the degree Doctor of Philosophy
in Molecular Toxicology

by

Thomas Joseph Kowal

2025

© Copyright by
Thomas Joseph Kowal
2025

ABSTRACT OF THE DISSERTATION

Development and Application of Data-Driven
Computational Methods for Molecular Toxicology
and Environmental Exposure Research

by

Thomas Joseph Kowal

Doctor of Philosophy in Molecular Toxicology

University of California, Los Angeles, 2025

Professor Xia Yang, Chair

Molecular and environmental toxicology research is currently heavily reliant on *in vivo* animal models, and the field seeks to develop and implement Novel Alternative Methods (NAMs) such as computational and *in vitro* approaches to drive efficient and translatable discoveries. Leveraging the ample species-, tissue-, and cell type-specific transcriptome datasets to gauge *in vivo* physiology, my dissertation research focused on developing data-driven computational approaches to understand toxicological and environmental exposures at a systems level and to guide *in vitro* model design to better mimic *in vivo* physiology. To establish proof-of-concept, Aim 1 of my

dissertation applied a transcriptomics data-driven approach to elucidate the systematic molecular impacts of the key cannabinoid components cannabidiol (CBD) and delta-9-tetrahydrocannabinol (THC) across species and tissues. Specifically, we curated 105 CBD- and THC-related transcriptomics datasets and analyzed them to detect differentially expressed gene signatures, enriched biological pathways, network regulation, and disease associations. CBD datasets across species and tissues showed relatively higher stability and reproducibility in targeting metabolic, anti-inflammatory, and cell cycle pathways, whereas THC studies showed highly variable genes, pathways, and networks indicative of context-specific molecular activities. Aim 2 of my dissertation expanded this data-driven approach to develop ToxiOmics, an open-access database and web tool to curate and analyze over 1000 transcriptomics datasets for 222 toxicologically relevant compounds across 29 tissues in four model organisms: human, mouse, rat, and rhesus macaque. We identified species-, tissue-, and dose-segregated gene signatures for each compound, allowing within and between chemical comparisons at gene, pathway, and network levels across studies to obtain systematic mechanistic insights and disease associations. Aim 3 of my dissertation compares transcriptomic profiles of cells from *in vivo* tissues with those of *in vitro* cell lines to guide the selection of cell lines that best resemble physiological cells to improve *in vitro* toxicological model paradigms. I curated and analyzed 214 transcriptomics datasets from 16 human and mouse hepatic cell lines, and 7 single cell RNA sequencing *in vivo* human and mouse liver tissue datasets, to determine the physiological relevance of each cell line through gene, pathway, and network analyses. My analysis identified HepaRG as the cell line which best mimics *in vivo* hepatocyte gene expression profiles and provided biological insights into the key differences of some of the widely used hepatic cell lines from *in vivo* physiology. Overall, my dissertation projects leveraging hundreds to thousands of transcriptomic datasets derived

mechanistic insight of the toxicological effects of numerous environmental chemicals and contributed to the advancement of *in silico* and *in vitro* NAMs in toxicology research.

The dissertation of Thomas Joseph Kowal is approved.

Matteo Pellegrini

Oliver Hankinson

Patrick Allard

Xia Yang, Committee Chair

University of California, Los Angeles

2025

DEDICATION

This dissertation is dedicated to Mom; my sisters Allison, Carolyn, and Catherine; my grandparents MaryAnn and Joe; and my pals Charlotte, Olive, and Theo. Thank you for unwaveringly supporting me and believing in me at every step of the way along my academic journey.

TABLE OF CONTENTS

ABSTRACT OF THE DISSERTATION	ii
DEDICATION	vi
LIST OF FIGURES	ix
LIST OF TABLES	xi
LIST OF ABBREVIATIONS	xii
ACKNOWLEDGEMENTS	xv
VITA	xvii
Chapter 1. Introduction	1
1.1 Figures	7
Chapter 2. Comparative analysis of 105 datasets across species and tissues reveals higher variability in transcriptomic responses to THC than CBD	8
2.1 Introduction	8
2.2 Materials and Methods	10
2.3 Results	15
2.4 Discussion	29
2.5 Figures	37
Chapter 3. ToxiOmics: A Species- and Tissue-specific Gene Signature Database and Web Tool to Understand Toxicological Exposures and Disease Associations	51
3.1 Introduction	51
3.2 Methods	53
3.3 Results	59
3.4 Discussion	65

3.5	Figures.....	71
Chapter 4. Computational Integration of <i>In Vivo</i> Single Cell and <i>In Vitro</i> Bulk Transcriptomics Across 221 Human and Mouse Datasets Differentiates Physiological versus Non-Physiological Hepatic Cell Lines for Hepatotoxicity Screening.....		
4.1	Introduction.....	80
4.2	Methods.....	82
4.3	Results.....	87
4.4	Discussion.....	97
4.5	Figures.....	103
4.6	Tables.....	109
Chapter 5. Conclusions and Future Directions		111
REFERENCES		114

LIST OF FIGURES

Figure 1.1 Dissertation overview of chapter contents.	7
Figure 2.1 Overall study design for comparative analysis of THC and CBD.	37
Figure 2.2 UMAP clustering reveals instability in DEG signatures across tissues in THC and CBD.	38
Figure 2.3 Hierarchical clustering reveals clusters of datasets with concordant gene expression signatures in response to THC exposure.	39
Figure 2.4 Hierarchical clustering reveals clusters of datasets with concordant gene expression signatures in response to CBD exposure.	40
Figure 2.5 DEG results indicate CBD induces more DEGs across tissues and identify factors affecting the number of observed DEGs across studies.	42
Figure 2.6 Meta-analysis reveals highly represented gene and pathway perturbations across cannabinoid exposure signatures.	44
Figure 2.7 Key driver analysis identifies key regulatory genes in the perturbation effects of CBD and THC.	46
Figure 2.8 Correlation and overlap analysis reveals the extent of ECS gene involvement in CBD- and THC-induced transcriptome perturbations.	48
Figure 2.9 MSEA reveals associations between cannabinoid exposure and psychiatric, metabolic diseases.	49
Figure 3.1 ToxiOmics analytical workflow.	71
Figure 3.2 Chemical classes included in ToxiOmics.	72
Figure 3.3 ToxiOmics signature coverage summaries.	73

Figure 3.4 An overview of the data contained within ToxiOmics and the functional applications of the web tool.	74
Figure 3.5 Comparison of murine liver meta DEG and pathway signatures for legacy PFAS compounds and GenX.	75
Figure 3.6 Clustering analysis reveals transcriptomic similarities between legacy PFAS compounds Perfluorooctanoic acid (PFOA), Perfluorooctane sulfonic acid (PFOS), Perfluorononanoic acid (PFNA), and Hexafluoropropylene oxide dimer acid (HFPO-DA, “GenX”) across doses, durations of exposure, and developmental stages.	76
Figure 3.7 Pathway enrichment analysis reveals shared biological pathway perturbations elicited by the transcriptomic response to both legacy PFAS and GenX exposure in the murine liver. ...	77
Figure 3.8 Key driver analysis reveals shared key drivers of the transcriptomic response to both legacy PFAS and GenX exposure in the murine liver.	78
Figure 3.9 Network-based comparative analysis highlights potential antagonistic effects between PFOA exposure and pioglitazone treatment.	79
Figure 4.1 UMAP reveals immortalized cell lines that closely resemble hepatic cell types at the level of the basal transcriptome.	103
Figure 4.2 UMAP rules out several likely confounding variables as explanations for the emergence of distinct HepG2 subclusters.	105
Figure 4.3 Pathway enrichment results elucidate perturbed pathways and highlight relevant human diseases among hepatocyte cell lines.	107
Figure 4.4 Key driver analysis reveals key regulators driving the differences in core hepatocyte physiology from immortalized hepatocyte cell lines.	108

LIST OF TABLES

Table 4.1 DEG number and top pathway enrichment results enable comparison between hepatocyte cell lines and <i>in vivo</i> hepatocyte basal gene expression data.	109
Table 4.2 Fisher's Exact Test overlap statistics potentially indicate HepaRG as better capturing the toxicological response to TCDD exposure than HepG2, but highlight the importance of experimental context in cell line selection.	110

LIST OF ABBREVIATIONS

2-AG	2-arachidonoylglycerol
AEA	Anandamide
BH	Benjamini-Hochberg
BPA	Bisphenol A
CAD	Coronary artery disease
CB1	Cannabinoid receptor 1
CB2	Cannabinoid receptor 2
CBD	Cannabidiol
CD	Crohn's disease
CREEDS	Crowd Extracted Expression of Differential Signatures
CTD	Comparative Toxicogenomics Database
CUD	Cannabis use disorder
DEG	Differentially expressed gene
DMEM	Dulbecco's modified essential medium
ECS	Endocannabinoid system
FDR	False discovery rate
GEO	Gene Expression Omnibus
GIP	Glucose-dependent insulintropic peptide
GOBP	Gene Ontology Biological Process
GRN	Gene regulatory network
GSE	GEO series
GWAS	Genome-wide association study

HDL	High-density lipoprotein
HFPO-DA	Hexafluoropropylene oxide dimer acid
HIV-1	Human immunodeficiency virus - 1
HSC	Hepatic stellate cell
ICCVAM	Interagency Coordinating Committee on the Validation of Alternative Methods
KD	Key driver
KDA	Key driver analysis
KEGG	Kyoto Encyclopedia of Genes and Genomes
KS	Komolgorov-Smirnov
LDL	Low-density lipoprotein
LIMMA	Linear models for microarray data
LPS	Lipopolysaccharide
MEM	Minimum essential medium
MSEA	Marker set enrichment analysis
NAFLD	Non-alcoholic fatty liver disease
NAM	Novel alternative method
NASH	Non-alcoholic steatohepatitis
NCBI	National Center for Biotechnology Information
NICEATM	National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods
NTP	National Toxicology Program
PBMC	Peripheral blood mononuclear cell

PFAS	Per- and polyfluoroalkyl substance
PFNA	Perfluorononanoic acid
PFOA	Perfluorooctanoic acid
PFOS	Perfluorooctane sulfonic acid
PMID	PubMed Identifier
RIMBANet	Reconstructing Integrative Bayesian Networks
RNAseq	RNA sequencing
RRA	Robust rank aggregation
scRNAseq	Single-cell RNA sequencing
SIV	Simian immunodeficiency virus
SNP	Single nucleotide polymorphism
SRA	Sequence Read Archive
T2D	Type 2 diabetes
TC	Total cholesterol
TCDD	Tetrachlorodibenzodioxin
TG	Triglycerides
THC	Δ^9 -tetrahydrocannabinol
UMAP	Uniform manifold approximation and projection
VTA	Ventral tegmental area

ACKNOWLEDGEMENTS

To Xia, I am so thankful that you welcomed me into your lab family shortly after my arrival at UCLA. From the very first couple of weeks of my rotation in 2020, I knew that you were the type of person that I would want to guide me through the PhD. You have been such an incredible mentor due not only to your scientific expertise, but also the empathy with which you run your lab. You met me where I was at (which, in terms of bioinformatics experience, was essentially at the starting line) and allowed me to grow and learn to be able to contribute effectively as a researcher. You guided me through scientific challenges and kept me on course when I was facing personal challenges at various points over the years. I will always remember the impact you have had on my career and my life, and be grateful that you have been a part of it.

To Professor Hankinson, Professor Allard, and Professor Pellegrini, thank you all so much for serving on my committee throughout my time as a PhD student. You have all provided invaluable feedback and challenged me to dig deeper and generate more compelling research. It has been a privilege to work with you, learn from you, and be inspired by your challenges to take my work to the next level. My projects would not be what they are without you.

To Oliver and Patrick, thank you for granting me the privilege of being a PhD student in the Environmental and Molecular Toxicology program. All I ever wanted growing up was to get into science at the highest level, and by taking a chance on me right out of my undergraduate studies, you have helped me to live out that dream.

To Mom, my sisters, and all the wonderful friends I have made along the way, thank you all for your support as I have pursued this PhD. You have all inspired me to achieve the best version of myself, and I would not be at this point without you all. You have celebrated the good times with me, and consoled me when things got tough. Thank you all for helping me get here.

Chapter 2 is a version of Liu R⁺, Kowal T⁺, Chow C, Olsen T, Nguyen E, Yang S, Lee J, Yang X*, Blencowe M*. Comparative analysis of 105 datasets across species and tissues reveals higher variability in transcriptomic responses to THC than CBD. This article has been submitted to the Journal of Cannabis Research and is currently under review.

Chapter 4 is a version of Kowal T, Wang S, Cheng M, Liu R, Blencowe M, Yang X. Computational Integration of *In Vivo* Single Cell and *In Vitro* Bulk Transcriptomics Across 221 Human and Mouse Datasets Differentiates Physiological versus Non-Physiological Hepatic Cell Lines for Hepatotoxicity Screening. This article has been submitted for review to Toxicological Sciences.

The work in **Chapters 2, 3, and 4** done by TK is supported by the Eugene and Sallyann Fama fellowship, and the National Institute of Environmental Health Sciences of the National Institutes of Health (T32ES015457).

VITA

EDUCATION

- 2024 PhD Candidate, Molecular Toxicology
University of California, Los Angeles
- 2020 Bachelor of Science, Chemistry (*summa cum laude*) and Biology
Emory University

RESEARCH EXPERIENCE

- 2020 - Present PhD Trainee
Department of Integrative Biology and Physiology
University of California, Los Angeles
- 2016 - 2020 Undergraduate Research Assistant
Department of Chemistry
Emory University

SELECTED MANUSCRIPTS IN REVISION

Liu R⁺, **Kowal T**⁺, Chow C, Olsen T, Nguyen E, Yang S, Lee J, Yang X*, Blencowe M*. Comparative analysis of 105 datasets across species and tissues reveals higher variability in transcriptomic responses to THC than CBD.

Kowal T, Wang S, Cheng M, Liu R, Blencowe M, Yang X. Computational Integration of *In Vivo* Single Cell and *In Vitro* Bulk Transcriptomics Across 221 Human and Mouse Datasets Differentiates Physiological versus Non-Physiological Hepatic Cell Lines for Hepatotoxicity Screening.

Ha S, Ahn IS, **Kowal T**, Yoon J, Diamante G, Cely I, Zhang G, Wang S, Yang X. Cross-tissue multiomics studies reveal gut-brain interactions mediating the effect of *Akkermansia muciphila* in counteracting fructose-induced obesity.

SELECTED MANUSCRIPTS IN PREPARATION

Kowal T⁺, Zamora Z⁺, Ha S, Qi L, Quinto E*, Garcia K*, Narso O, Chen Y, Yang X. ToxiOmics: A Species- and Tissue-specific Gene Signature Database and Web Tool to Understand Toxicological Exposures and Disease Associations.

SELECTED CONFERENCE PRESENTATIONS

Kowal T, Zamora Z, Chen Y-W, Quinto E, Garcia K, Yang X. ToxiOmics - A Transcriptome-Based Database and Web Tool to Query and Understand the Associations Between Environmental Toxicants and Human Diseases. Poster presentation at: *FutureTox V Meeting*, Chapel Hill, NC. 2022.

Kowal T, Zamora Z, Chen Y-W, Quinto E, Garcia K, Yang X. ToxiOmics - A Transcriptome-Based Database and Web Tool to Query and Understand the Associations Between Environmental Toxicants and Human Diseases. Oral presentation at: *Southern California Chapter of the Society of Toxicology Regional Meeting*, San Diego, CA. 2022.

Kowal T, Zamora Z, Chen Y-W, Quinto E, Garcia K, Yang X. ToxiOmics - A Transcriptome-Based Database and Web Tool to Query and Understand the Associations Between Environmental Toxicants and Human Diseases. Poster presentation at: *Society of Toxicology Annual Meeting*, Nashville, TN. 2023.

Kowal T, Cheng M, Liu R, Blencowe M, Yang X. Integration of *In Vivo* Single Cell RNA Sequencing with *In Vitro* Bulk Transcriptomics to Optimize Selection of Immortalized Cell Lines for Hepatotoxicity Screening. Poster presentation at: *Society of Toxicology Annual Meeting*, Orlando, FL. 2025.

Chapter 1. Introduction

At its core, the field of toxicology seeks to understand the adverse effects of chemical exposures on biological systems. Traditionally, studies in the field have been dominated by animal models which have provided valuable insights into toxicological mechanisms, dose-response relationships, and disease associations of toxicological exposures. However, the heavy reliance on *in vivo* animal testing presents important challenges that the field must overcome to continue to deliver accurate and translatable insights into human-relevant toxicological effects. Animal use comes with ethical concerns for animal welfare, high cost, and limited throughput due to the regulatory requirements and length of time required to observe toxicological effects. These concerns have inspired the philosophy of the “3 R’s,” which are the reduction of the number of animals used in an experiment, the refinement of experimental techniques to minimize harm to the animals and make the experiments efficient, and the replacement of *in vivo* animal studies altogether with alternative approaches where possible^{1,2}. The increasing desire to improve the sustainability and ethicality of research through the reduction of animal use has driven the development of Novel Alternative Method (NAM) approaches in toxicology. These methods, which include *in vitro* models and computational *in silico* approaches, aim to accelerate discovery while improving reproducibility, translatability, and sustainability^{3,4}.

The rise of big data in the biomedical sciences has presented the field with a unique opportunity to rapidly accelerate NAM development and adoption. My dissertation focuses on leveraging transcriptomics data, which captures perturbations in gene expression, to guide *in silico* and *in vitro* approaches to accelerate toxicology and environmental exposure research. The transcriptome has long been used as a gauge of cellular physiology and function, and as a first-line measurement of toxicity response, in toxicology research⁵⁻⁷. Thanks to the advent of big

transcriptomics data in toxicology, we are able to view a holistic picture of the effects of toxicological perturbations. Big transcriptomics data has the power to elucidate direct molecular perturbations and higher-order regulatory processes through the capture of global gene expression patterns. Genetic research in the late 2010s led to a paradigm shift from viewing traits on a single-gene basis to viewing them as the result of complex interactions in gene regulatory networks (GRNs). This model of phenotypic traits arising from complex interactions across an interconnected network of genes is known as the omnigenic model^{8,9}. With unbiased transcriptomics data, we are able to assess toxicologically-induced perturbations in gene expression across the genome and leverage network biology to fully piece together the complex interactions leading to the observed phenotypic endpoints of toxicological exposures.

My dissertation seeks to contribute to the development and adoption of NAMs through the development and application of transcriptomics-based computational methods to study toxicological exposures and to guide *in vitro* model selection to deliver translatable results (Fig. 1.1). The overarching goal of this work is to leverage existing big transcriptomics data to derive mechanistic insight into toxicological effects and advance *in silico* and *in vitro* NAM approaches.

As a case study for proof of concept of this transcriptomics data-driven approach to toxicological research, we conducted a study of cannabis to understand the beneficial versus adverse mechanistic actions. *Cannabis sativa* contains more than 100 biologically active cannabinoids¹⁰, of which cannabidiol (CBD) and Δ^9 -tetrahydrocannabinol (THC) have emerged as the most prominent due to their increasing medicinal and recreational uses in recent years^{11,12}. THC is the primary psychoactive component and has demonstrated therapeutic benefits for pain management and cancer treatment^{13,14}, while CBD is non-psychoactive and has been added to a variety of products advertised to induce calming sensations and reduce inflammation^{15,16}.

Widespread legalization initiatives in the United States have fueled a dramatic increase in cannabis use in both adolescents and adults^{17,18}. However, the effects of cannabinoids at the molecular level across tissues and exposure contexts remain incompletely understood despite their increasing use and exposure in the general population. To address this challenge, **Aim 1** of this dissertation systematically analyzed more than 100 transcriptomic datasets related to THC and CBD across species and tissue types. By identifying differentially expressed genes (DEGs), enriched biological pathways, and regulatory networks, and integrating disease association analyses, this work provides a comprehensive overview of the molecular effects of cannabinoids. Importantly, it highlights how CBD exhibits relatively consistent modulation of metabolic, anti-inflammatory, and cell cycle pathways across contexts, whereas THC produces more variable signatures suggestive of highly context-specific actions. This analysis serves to clarify cannabinoid biology and illustrate the power of transcriptomics data-driven approaches to elucidate complex effects of toxicological exposures.

While the cannabinoid case study is an instructive case study of the transcriptomics data-driven approach and addresses an area of particular contemporary relevance, the scope of toxicology extends much further. Worldwide populations are consistently exposed to an array of toxicologically relevant substances, including industrial chemicals, metals, persistent organic pollutants, endocrine disruptors, and psychoactive agents. Understanding their toxicological effects is critical in the regulatory space to develop guidelines for safe levels of exposure. To increase research throughput while addressing concerns about the traditional reliance on animal use, regulatory agencies such as the US National Toxicology Program (NTP) and the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) have called for the development of NAM approaches in toxicity screening⁴. Several resources, including the

Comparative Toxicogenomics Database (CTD)¹⁹ and Crowd Extracted Expression of Differential Signatures (CREEDS)²⁰, have attempted to catalog toxicant-induced transcriptomics signatures. However, these resources are limited in that their reliance on literature curation induces bias, underlying raw data may be inaccessible, and community-submitted analyses may vary in quality and methodology. To overcome these gaps, **Aim 2** of this dissertation developed ToxiOmics, a transcriptome-based database and web tool to query and understand the associations between toxicological exposures and human diseases. ToxiOmics applies a standardized, transcriptomics-driven pipeline to over 1,000 transcriptomics datasets spanning 222 toxicologically relevant compounds across 29 tissues in humans, mice, rats, and rhesus macaques to generate species-, tissue-, dose-, and time-specific DEG and pathway signatures. ToxiOmics leverages pathway enrichment and network biology approaches, and integration with our existing Mergeomics^{21,22} and PharmOmics²³ tools to derive mechanistic insights, disease associations, and potential drug interactions from toxicological exposures. As demonstrated through our case study comparing legacy per- and polyfluoroalkyl substance (PFAS) compounds and the emerging replacement GenX, ToxiOmics allows researchers to identify conserved or divergent mechanisms of action between toxicants, infer concordant or discordant effects with pharmaceuticals, and explore regulatory subnetworks perturbed by exposures. We expect that ToxiOmics will offer a powerful, unbiased data-driven platform to advance *in silico* NAM-based toxicological research across a broad spectrum of environmental exposures.

NAM approaches in toxicology are not limited to *in silico* computational methodologies. *In vitro* cell culture experiments also offer a rapid and relatively inexpensive way to predict *in vivo* toxicological outcomes. However, *in vitro* systems vary widely in their ability to replicate physiological responses, due to both the short-lived and donor-specific nature of primary cells and

the genetic and epigenetic modifications acquired during the immortalization of cell lines. Thus, it is critical to select cell lines that balance experimental practicality and physiological relevance. Traditional approaches to cell line selection have focused on criteria like known genetic mutations, compatibility with specific assays, and surface markers²⁴⁻²⁷. However, these criteria do not necessarily capture global functional similarity to *in vivo* tissues. The transcriptome can be leveraged to provide a holistic measure of baseline cellular physiology and assess the physiological resemblance of cell lines to their corresponding *in vivo* cell types in both baseline physiology and toxicological response. To this end, **Aim 3** of this dissertation developed a transcriptomics-based framework to evaluate the physiological relevance of *in vitro* cell lines. We focused here on the liver, a central organ in xenobiotic metabolism and toxicological response, and curated and analyzed 214 transcriptomics datasets from 16 human and mouse hepatic cell lines along with 7 single-cell RNA sequencing (scRNAseq) datasets from human and mouse liver tissues. Through gene-, pathway-, and network-level analyses at physiological baseline, we identified HepaRG as the cell line most closely resembling *in vivo* hepatocytes and documented key differences in other commonly used hepatic cell lines, providing insight into their strengths and limitations. This transcriptomics data-driving approach offers a generalizable strategy to guide *in vitro* cell line selection to model other tissues. This approach grounds cell line selection in data rather than arbitrary criteria, and in so doing promotes more translatable *in vitro* approaches to toxicity studies to advance the adoption of NAMs.

In summary, this dissertation illustrates how big transcriptomics data can be harnessed to advance NAM development and adoption in toxicology research. In **Aim 1**, we elucidate cannabinoid mechanisms by systematically analyzing gene expression effects of THC and CBD across species and tissues in order to differentiate between therapeutic versus potentially adverse

molecular pathways. In **Aim 2**, we build a community-serving infrastructure with ToxiOmics, a scalable open access resource which enables mechanistic comparisons across chemicals and studies. In **Aim 3**, we guide *in vitro* cell line selection to improve *in vitro* modeling by introducing a transcriptomics-based framework for evaluating the physiological relevance of immortalized cell lines which is generalizable from our liver case to other tissues. Together, this work advances the field of toxicology toward more efficient, ethical, and sustainable research in line with the philosophy of the “3 R’s.” Through the leverage of big transcriptomics data, we promote the development and adoption of NAMs which will reduce reliance on animal studies, enhance reproducibility, and provide tools to promote further research inquiry and guide regulatory and experimental decision-making.

1.1 Figures

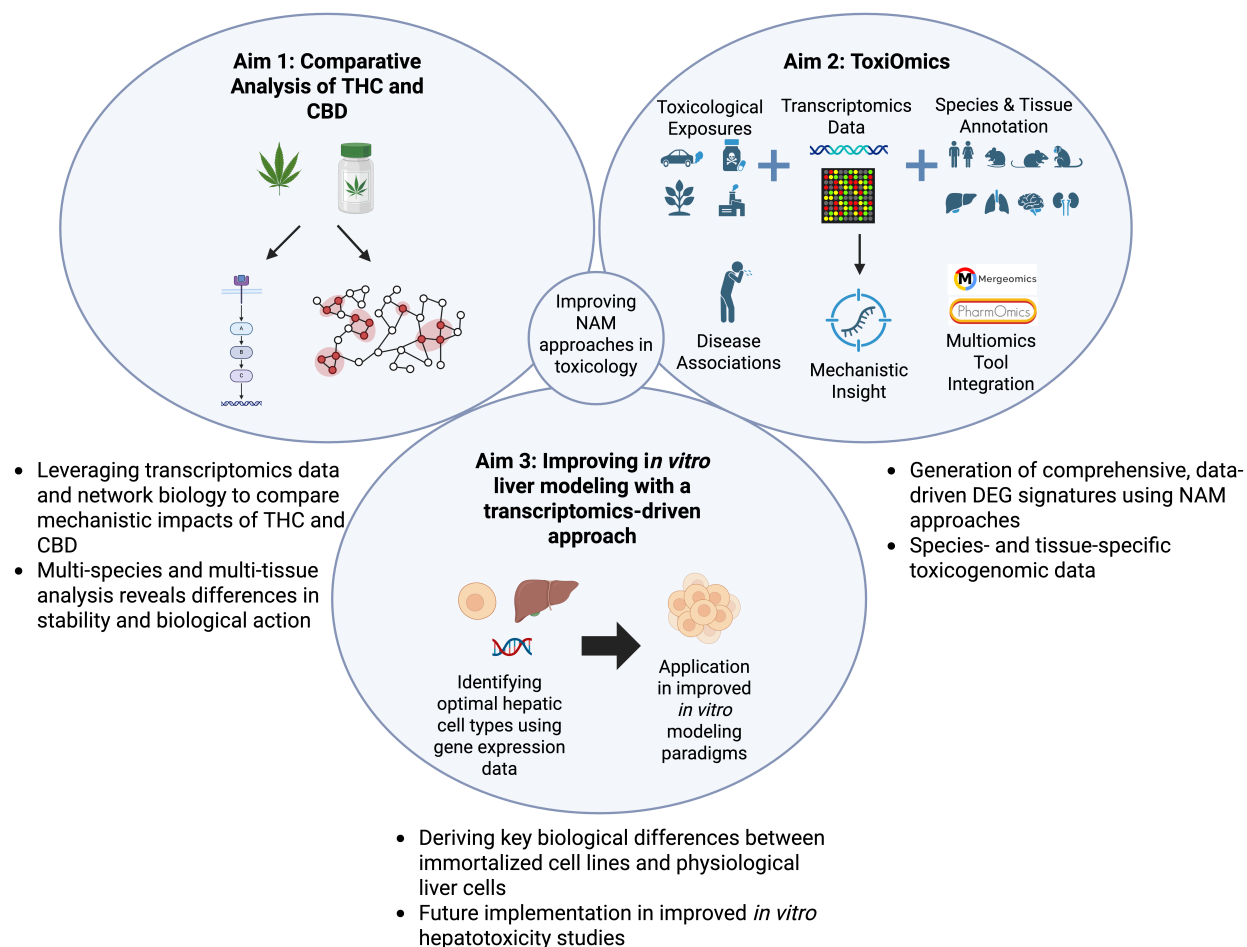


Figure 1.1. Dissertation overview of chapter contents.

Chapter 2. Comparative analysis of 105 datasets across species and tissues reveals higher variability in transcriptomic responses to THC than CBD

2.1 Introduction

Cannabis use for recreational and medicinal purposes has increased substantially over the past several years in the United States, likely due to broadening legalization of cannabis. By December 2018, medical cannabis was legalized in 33 states and D.C., while recreational use was permitted in 10 states and D.C.²⁸. In 2022, 30.7% of US high school seniors reported cannabis use in the past year, with 6.3% using it daily²⁹. Among adults, cannabis use increased from 7.59% to 15.11% in 2013-2022³⁰. Delta-9-tetrahydrocannabinol (THC) and cannabidiol (CBD) are the two most prominent cannabinoids in *Cannabis sativa*, comprising up to 40% of the plant's extract^{31,32}. These compounds interact with the endocannabinoid system, a complex network of receptors and signaling molecules that regulate pain, mood, inflammation, and immune responses³³. THC, the primary psychoactive component, exhibits therapeutic effects as an analgesic for cancer-related chronic pain and has demonstrated anti-invasive and anti-metastatic properties in cancer treatment³⁴. CBD is non-psychoactive but still influences the brain and nervous system^{35,36}, is recognized for its potential in managing epilepsy, anxiety, and neurodevelopmental disorders, and is valued for its anti-inflammatory and antioxidant properties³⁷.

While cannabinoids have demonstrated therapeutic potential, concerns regarding their safety in terms of physical health, mental health, public safety, and the side effects have tempered their widespread acceptance^{37,38}. For instance, while some studies suggest that chronic daily use of up to 1500 mg/day is well tolerated in humans, others report both physical and mental side effects such as cognitive impairment, cardiovascular complications, and respiratory issues³⁸. Additionally, cannabinoids have been linked to immune suppression, resulting in increased

susceptibility to human immunodeficiency virus - 1 (HIV-1) infections and disease progression. Mental health effects are particularly concerning, as cannabinoids may exacerbate bipolar disorder in predisposed individuals and increase the risk of temporary psychosis. In addition, cannabinoids also pose broader public health and safety concerns. Driving under the influence is associated with impaired motor function and a higher incidence of motor vehicle accidents. Frequent and heavy cannabis use during adolescence increases the risk of developing cannabis use disorder (CUD)³⁹, while addiction and cannabis dependence may contribute to lower income, unemployment, and reduced life satisfaction³⁸.

Therefore, understanding the precise molecular and biological mechanisms underlying the beneficial and adverse effects of THC and CBD is crucial for guiding safe use and developing strategies to mitigate health concerns. At the transcriptomic level, studies have shown that cannabinoid-induced gene expression changes vary depending on tissue type, sex, age, and genetic backgrounds, such as gene mutations³⁹⁻⁴². This variability has made it challenging to establish precision targets that differentiate the therapeutic vs adverse effects. This study systematically analyzes over 100 transcriptomic datasets related to THC and CBD across species and tissue types from the Gene Expression Omnibus (GEO)⁴³ (Fig. 2.1). Using transcriptomic data, we identified differentially expressed genes (DEGs) for each dataset. We further investigated global regulatory patterns across datasets through clustering and correlation assessment as well as cannabinoid-specific effects through pathway and disease association analysis, and network modeling. We also investigated the effects of THC and CBD on the endocannabinoid system across the datasets. These integrative approaches allowed us to uncover consistent and unique patterns of gene regulation across studies to partition potential targets underlying beneficial vs. adverse effects of THC and CBD.

2.2 Materials and Methods

Data curation

Transcriptomic data from previous studies were curated from the National Center for Biotechnology Information's Gene Expression Omnibus (NCBI GEO)^{43,44}. To identify THC-specific datasets, we queried GEO using the keywords "THC" and "Delta-9-tetrahydrocannabinol," while CBD-specific datasets were retrieved using "CBD" and "Cannabidiol." Studies were included if they met the following criteria: 1) the model organism was human, mouse, rat, or rhesus macaque to focus on mammalian species with higher translational potential; 2) the dataset was generated using cDNA microarrays or RNA sequencing; 3) a minimum sample size threshold of $n = 3$ per group for *in vivo* studies and $n = 2$ per group for *in vitro* studies was applied, allowing us to include most available studies while balancing data quality and comprehensiveness; 4) the study was not part of a subseries or superseries on GEO to avoid duplicate data. For studies testing multiple conditions, samples were grouped based on their matching physiological and pathological background conditions to ensure accurate differential expression analysis. Specifically, only control and treatment samples with shared conditions, such as prior chemical exposures, genetic mutations, and preexisting diseases, were analyzed together. This approach allowed for the assessment of treatment effects within comparable biological contexts while minimizing confounding factors. A total of 47 datasets for THC and 58 datasets for CBD met the criteria and were curated (Table S.2.1).

To validate that our data curation and analysis procedures can identify THC and CBD-specific transcriptomic effects, we also included additional well-studied, non-cannabinoids chemicals, including bisphenol A (BPA), perfluorooctanoic acid (PFOA), and estradiol, for comparison. Our previous studies of BPA and PFOA have revealed stable and consistent gene

signatures for PFOA but more variable gene expression changes in response to BPA across studies, making them potential reference points to assess gene signature stability vs variability⁴⁵⁻⁴⁷. We also included estradiol, an endogenous bioactive molecule with well-known biology, as a control to assess whether our analytical pipeline retrieves known biology. Gene expression data for these chemicals were obtained from GEO and processed using the same methods as described for THC and CBD. A total of 50 BPA datasets, 39 estradiol datasets, and 14 PFOA datasets were curated (Table S.2.2).

Data download and preprocessing

Microarray data were downloaded from GEO using the *GEOquery* package⁴⁸. As microarray data submitted to GEO are pre-processed and quality-controlled, we verified normalization and applied log2 transformation before downstream analysis.

Raw RNA sequencing (RNA-seq) datasets were retrieved from NCBI's Sequence Read Archive (SRA), quality-checked, and processed⁴⁹. FASTQ files were downloaded using the parallel-fastq-dump wrapper, followed by quality control and preprocessing. Trim Galore (v0.6.6) was used to remove low-quality bases from the 3' end of reads⁵⁰. Cutadapt (v2.1.0)⁵¹ was employed to remove adapter sequences. Reads shorter than 20 base pairs after trimming and adapter removal were filtered out to ensure data quality.

Reads were mapped to appropriate species-specific reference genomes using Salmon (v1.10.0)⁵². The reference genome assemblies used were GRCh38.108 for *Homo sapiens*, GRCm39.108 for *Mus musculus*, mRat.BN7.2.108 for *Rattus norvegicus*, and Mmul_10.108 for *Macaca mulatta*. Reference genomes for each species were indexed using the Salmon index tool

to optimize alignment efficiency. Finally, transcript-level quantification results from Salmon were imported and summarized using the tximport package (v1.32.0)⁵³.

Differentially expressed gene (DEG) analysis

For each dataset, the treatment group (THC or CBD) was compared to its corresponding control group to identify DEGs using Linear Models for Microarray Data (LIMMA) for cDNA microarray datasets⁵⁴ and DESeq2 for RNA-seq datasets⁵⁵. Multiple testing was corrected using the Benjamini-Hochberg method to obtain false discovery rate (FDR). For studies with multiple doses, baseline conditions (e.g. prior chemical exposures, genetic mutations, and preexisting diseases), or time points, each variation was treated as an independent dataset to extract distinct DEG signatures for each dose, condition, and timepoint. DEGs were considered significant at an FDR < 5%. All gene labels were converted to their human orthologs to facilitate cross-species comparisons.

Clustering and correlation analysis of DEG gene signatures across datasets

To compare DEGs across studies, we combined the gene expression log fold changes from individual studies from the differential gene expression analysis. Since RNAseq and microarray experiments can produce log fold changes on different scales, we applied a rank-based normalization method to standardize values within a range of -1 (downregulated in the treatment group) to 1 (upregulated in the treatment group) for each dataset. Specifically, we separately ranked positive and negative log fold changes within each dataset. Positive values were ranked in ascending order and scaled between 0 and 1 (no change to most upregulated), while negative values were ranked in descending order and scaled between -1 and 0 (most downregulated to no change).

This approach preserves the relative magnitude of gene expression changes among genes while making the datasets comparable across studies and transcriptome platforms. To assess similarity and differences across datasets for THC and CBD through cluster analysis, we used the top 2,500 most variable genes. For the analysis across all chemicals (THC, CBD, PFOA, BPA, estradiol), given the large number of datasets, we first selected genes present in at least 70% of the 208 datasets before selecting the top 2,500 most variable genes. Missing values in the DEG fold change table were imputed using the missForest R package⁵⁶. We then applied UMAP for dimensionality reduction using Euclidean distance to visualize the clustering patterns across chemicals and datasets.

As an alternative approach to assess similarity and differences across datasets, we used Spearman's rank correlation analysis of the normalized gene expression fold change data to compute pairwise correlations among datasets. The pairwise correlation coefficients were further used to group datasets into clusters with similar gene regulation patterns. To identify consistent DEGs across datasets within each cluster in response to THC or CBD, we performed a meta-analysis using the Robust Rank Aggregation (RRA) package (v1.2.1)⁵⁷. Rank aggregation was conducted separately for up-regulated and down-regulated DEGs in each dataset cluster. Genes with an RRA score < 0.05 were considered robust, consistent DEGs, reflecting consistent differential expression across datasets within each cluster. The RRA score measures the probability of a gene achieving its observed ranking pattern across datasets by chance, with lower scores indicating greater consistency.

Pathway enrichment analysis

Pathway enrichment analysis was performed on the identified DEGs using EnrichR⁵⁸. DEGs were compared against the Gene Ontology Biological Process (GOBP)⁵⁹ databases to identify significantly enriched pathways. Pathways with an FDR < 5% and at least 5 overlapping DEGs were considered statistically significant.

Weighted Key Driver Analysis (wKDA) for brain network modeling of THC and CBD DEGs

To identify potential gene regulatory networks and network key drivers (KDs) underlying cannabis-induced brain effects, we applied Weighted Key Driver Analysis from the Mergeomics pipeline^{21,22}. We conducted the analysis separately from THC- and CBD-treated brain datasets. Using a previously constructed brain Bayesian network based on large human and animal model omics studies in Mergeomics, we identified nodes (genes) whose immediate subnetworks were enriched for the THC or CBD DEG sets at FDR < 0.05 as statistically significant KDs. Network visualizations were performed using Cytoscape⁶⁰.

Marker Set Enrichment Analysis (MSEA) for disease/trait association assessment

MSEA in the Mergeomics package was used to identify the enrichment of THC or CBD DEGs for genetic associations with 101 genome-wide association studies (GWAS) of diseases and phenotypic traits^{21,22}. Disease-associated genes were mapped using full summary statistics from the GWAS Catalog⁶¹, with single nucleotide polymorphisms (SNPs) assigned to genes within a 50 kb distance. MSEA employs a chi-square-like statistic with multiple quantile thresholds to assess whether a DEG set shows enrichment of disease SNPs compared to random chance. 10,000 permuted gene sets were generated for each DEG set. As detailed in Shu et al.²¹, the enrichment statistics from the permutations were used to approximate a Gaussian distribution from which enrichment p-values were determined. FDR was estimated using the Benjamini-Hochberg (BH)

correction. DEG sets were determined to be statistically significant for a given disease or trait if $FDR < 5\%$.

2.3 Results

Curation of transcriptomic datasets on cannabinoids across species and tissues

To investigate the impact of cannabinoids on gene expression, we obtained 108 transcriptomic datasets from GEO^{43,48} (Table S.2.1). Specifically, we obtained 3 datasets for overall cannabis use, 58 for CBD, and 47 for THC, spanning 32 human (*Homo sapiens*) studies, 8 non-human primate (*Macaca mulatta*) studies, 50 mouse (*Mus musculus*) studies, and 18 rat (*Rattus norvegicus*) studies across 15 broad tissue categories (e.g. blood, brain, cancer, digestive, heart, immune, kidney, liver, lung, muscle, oral, placenta, skin, stem cell, and vasculature). Since the “overall cannabis use” datasets involved either cannabis smoke exposure or observational studies in which other botanical compounds from the cannabis plant were not controlled for, we excluded them from our analysis to maintain our specific focus on the compounds THC and CBD. Notably, reflecting the focus on neurological effects of cannabinoids, brain-related datasets accounted for nearly half of all studies.

Many studies included various physiological and pathological background conditions in both control and treatment groups, such as genetic mutations (e.g., *Cox15*, *Ndufs4*), preexisting diseases (e.g., SARS-CoV-2, Simian Immunodeficiency Virus (SIV) infection), and prior chemical exposures (e.g., 2'-3'-cGAMP, formoterol/budesonide treatment). Additionally, the datasets covered variations in study design, including differences in sex distribution, dose, route and duration of administration. The broad coverage of diverse conditions allows us to not only determine specific transcriptomic signatures for each condition but also assess the consistency and

generalizability of cannabinoid-specific effects across datasets. After downloading and preprocessing the gene expression data, we conducted differential gene expression analysis. We then performed pathway enrichment and disease association analyses to characterize the broader transcriptomic effects of cannabinoids (Fig. 2.1)

THC and CBD DEGs are distinct from other chemicals

To evaluate whether cannabinoids THC and CBD show distinct transcriptomic signatures compared to other chemicals, we incorporated 103 datasets from three well-characterized chemicals, Bisphenol A (BPA), perfluorooctanoic acid (PFOA), and estradiol, which have been associated with metabolic and reproductive functions, respectively. Both BPA and PFOA are endocrine-disrupting chemicals that influence cardiometabolic disease-related pathways^{21,60,61}. Estradiol, a primary estrogen hormone, plays a role in reproductive and sexual function by regulating gene expression through estrogen receptors^{62,63}. We selected these as comparative datasets because we expect them to be functionally distinct from THC and CBD, which can be reflected by different downstream genes and pathways. In total, we included 50 BPA datasets, 39 estradiol datasets, and 14 PFOA datasets from GEO (Table S.2.2) and processed them using the same methodology as cannabinoids.

We first compared the chemical-induced expression changes for all genes across all THC, CBD, BPA, estradiol, and PFOA datasets based on the similarities in the normalized log fold changes in gene expression changes between treatment and control groups using clustering analysis and visualized the datasets with UMAP. Agreeing with our hypothesis, most BPA, estradiol, and PFOA datasets each formed distinct clusters, supporting biological coherence within each chemical and differences between chemicals (Fig. 2.2A). These results suggest that

comparative transcriptomics analysis captures gene expression patterns that likely reflect the different molecular mechanisms of different classes of chemicals. Agreeing with our previous observation on higher stability of PFOA gene signatures⁴⁵, most PFOA datasets clustered tightly in UMAP. In comparison, while BPA and estradiol datasets also formed distinct clusters in UMAP, the distance between datasets for BPA or estradiol was larger than that for PFOA. Compared to these reference chemicals, THC and CBD datasets were more spread out in UMAP. These results suggest that the molecular mechanisms of the reference chemicals are more coherent and consistent across datasets and species, with PFOA showing the highest stability, whereas THC and CBD datasets showed less stable gene signatures in comparison. Additionally, we did not observe a distinct separation between datasets from RNA sequencing and microarray methods, indicating a minimal effect of the technical platform on gene signatures (Figure S.2.1A).

We also observed that many tightly clustered datasets for the reference chemicals originated from the same study. For example, several BPA and estradiol microarray datasets were derived from GSE50705, where datasets shared the same experimental setup but differed in dosage. These datasets are clustered in the upper left corner of the UMAP plot (Fig. 2.2A). To mitigate the contribution of study-specific batch effects to the tight clustering of the chemicals, we averaged the log fold changes of each gene across datasets for each chemical from the same study, then performed dimensionality reduction again (Fig. S.2.1B). The clustering patterns remained consistent with the original analysis, confirming that study-related effects did not drive the observed trends in Fig. 2.2A. Specifically, when accounting for study effects, we found a stability ranking among the chemicals: PFOA exhibited the highest stability, followed by estradiol and BPA, with CBD and THC being highlighted as having the greatest variability.

CBD gene signatures demonstrate tighter clustering and higher stability across species and routes of administration compared to THC

To better assess the similarities and differences between THC and CBD datasets, we further clustered THC and CBD datasets excluding BPA, estradiol, and PFOA datasets. There was a subtle separation between THC and CBD datasets along the first UMAP dimension (UMAP1), indicating that their global transcriptional effects exhibit certain differences although not strongly distinguishable (Fig. 2.2B). Notably, datasets from both RNAseq and microarray platforms were mixed, again supporting that the technological platform did not induce major differences in transcriptomic signatures (Fig. S.2.1C, S.2.1D, S.2.1E).

Within the THC datasets, we found no distinct clustering patterns based on species (Fig. 2.2C), route of administration (Fig. S.2.1F), or tissue type (Fig. S.2.1G), suggesting that these factors did not significantly influence global transcriptional profiles.

In contrast, CBD datasets exhibited weak species-specific clustering. Along the second UMAP dimension (UMAP2), *Homo sapiens* and *Mus musculus* datasets showed subtle shifts, whereas *Rattus norvegicus* datasets clustered more tightly (Fig. 2.2D). However, these trends were not strong enough to indicate robust species-driven transcriptional signatures. Regarding the route of administration, *in vitro* (in medium) datasets tended to cluster at higher UMAP1 values, whereas *in vivo* (injection or ingestion) datasets were mainly positioned at lower UMAP1 values (Fig. S.2.1H). In the tissue-based UMAP (Fig. S.2.1I), no clear clustering trends emerged across tissue types. However, datasets from the same study, such as the two heart datasets and two kidney datasets, clustered more closely, suggesting that study-specific effects also contributed to their proximity.

Overall, the weak clustering patterns across species, tissues, and routes of administration suggest that these factors did not play a dominant role in shaping the global transcriptomic shifts induced by THC or CBD. However, CBD datasets displayed slightly greater clustering and stability compared to THC, indicating more consistent transcriptional responses across different conditions.

CBD studies demonstrate different correlation patterns than THC datasets

In addition to analyzing dataset clustering based on global gene expression changes through dimensional reduction using UMAP, we assessed correlation patterns in gene expression changes across all THC, CBD, BPA, estradiol, and PFOA datasets. A positive correlation (red) indicates that most genes are upregulated or downregulated concordantly between two datasets, whereas a negative correlation (blue) suggests discordance or opposite directions in gene regulation patterns. Results from this analysis are consistent with our UMAP clustering results in that the molecular patterns of THC and CBD are less consistent compared to the reference chemicals (Fig. S.2.2).

Next, we carried out clustering analyses separately for THC and CBD datasets based on the correlation patterns. In the THC correlation heatmaps, two major clusters of positively correlated datasets were observed: THC Cluster 1 in the upper left quadrant (11 datasets) and THC Cluster 2 in the lower right quadrant (6 datasets), both largely composed of brain-related datasets spanning multiple species. However, these two clusters exhibited moderate negative correlations with each other (Fig. 2.3). This result suggests that THC gene signatures from the brain could be partitioned into distinct patterns with each pattern showing consistency across studies. Notably, some datasets with the same experimental designs but differing in sex were assigned to different clusters. For instance, in GSE273695, which analyzed rat blood samples with daily intraperitoneal

injections escalating from 2.5 mg/kg to 10 mg/kg, the male dataset clustered in THC Cluster 1, while the female dataset appeared in THC Cluster 2. Additionally, multiple mouse brain datasets from GSE189821 (daily 10 mg/kg intraperitoneal injection) were distributed across both clusters: specifically, the male amygdala dataset was placed in THC Cluster 2, while the female amygdala dataset fell into THC Cluster 1. However, this separation of male and female samples into different clusters was not observed in all datasets. For example, rat prefrontal cortex datasets of both sexes in GSE273695 were all assigned to THC Cluster 1. While this finding suggests that sex may influence THC-induced transcriptomic responses, this sex-based separation was not consistently observed across all THC datasets and hence any potential sex-specific effects are likely more nuanced and depending on tissue type or other conditions to be elucidated.

CBD datasets formed four distinct clusters, with the two largest being CBD Cluster 1 in the upper left quadrant (17 datasets) and CBD Cluster 4 in the lower right quadrant (15 datasets) (Fig. 2.4). These clusters were primarily composed of *in vitro* experiments and included datasets from three species: human, mouse, and rat. CBD Cluster 1 featured a mixture of tissues, with prominent representation from immune, cancer, and brain. In contrast, approximately half of the datasets in CBD Cluster 4 were from the skin. Two smaller clusters were observed in the central region of the plot, representing study-specific clusters: CBD Cluster 2 included 7 mouse hypothalamus cell line datasets from GSE270378, and CBD Cluster 3 contained 5 mouse liver datasets from GSE261716, where datasets varied by CBD dosage and exposure time.

CBD induces more transcriptional changes than THC

To go beyond global transcriptomic patterns and focus on individual genes exhibiting significant changes in response to THC and CBD, we derived DEGs at FDR < 5%. Overall, CBD induced

more significant DEGs than THC (Fig. 2.5A, 2.5B). Most CBD datasets contain hundreds or thousands of DEGs, whereas THC datasets predominantly have fewer than 100 DEGs (Table S.2.3). Kolmogorov-Smirnov (KS) tests of the distributions of DEG counts across datasets confirmed that both upregulated and downregulated gene counts differed significantly between THC and CBD.

Other factors, including transcriptome platform, species, tissue, exposure type, duration, dosage, and sex, also influenced DEG counts. RNA-seq datasets generally yield more DEGs than microarray datasets (Fig. 2.5C). *In vitro* exposure studies of CBD exhibited higher numbers of DEGs than THC, while *in vivo* datasets show similar DEG distributions between CBD and THC (Fig. 2.5D). Examining species effects in CBD datasets, human *in vitro* cell line studies had the highest number of DEGs (Fig. 2.5E). For THC datasets, where the most DEG-rich studies are *in vivo*, mouse studies generate the largest number of significant DEGs, followed by non-human primate and rat studies. At the tissue level, CBD studies show the highest median number of DEGs in lung, muscle, cancer, immune, and kidney tissues (Fig. 2.5F). For THC, while no tissue exhibits a notably high median DEG count, the brain had the greatest variation in DEG numbers across studies, likely due to brain region-specific effects (Fig. 2.5G). Additionally, as expected, the number of DEGs in CBD studies increases with both exposure time and dosage (Fig. S.2.3A, Fig. S.2.3B, Fig. S.2.3C for different studies). For THC, no study specifically examined both time points and dosage variations. Unlike time and dosage, sex had a notable but inconsistent effect on DEG counts. Among 10 THC datasets with identical experimental designs differing only by sex, females exhibited more differentially expressed genes (DEGs) in five datasets, including the mouse brain amygdala, dorsolateral striatum, ventral tegmental area (VTA), prefrontal cortex, and rat peripheral blood mononuclear cells (PBMCs) under LPS stimulation. In contrast, males showed

higher DEG counts in four datasets: the rat brain orbitofrontal cortex (under both LPS and saline treatments), rat PBMCs with saline, and the mouse brain nucleus accumbens (Fig. S.2.3D). This variability highlights the complexity of sex-specific factors in cannabinoid responses.

CBD demonstrates higher consistency at gene and pathway levels across datasets

To evaluate the consistency of differentially expressed genes (DEGs) across datasets, we applied two complementary approaches: 1. identifying consistent DEGs within correlation-based clusters, and 2. assessing individual DEG recurrence and directionality consistency.

Cluster-Level Consistency:

We used rank aggregation across datasets and identified consistently upregulated and downregulated DEGs within each correlation cluster (Fig. 2.3, 2.4)⁵⁷. THC Cluster 1 contained relatively few DEGs, with 29 consistently upregulated genes, 18 consistently downregulated genes, and no significantly enriched pathways (Fig. 2.3, Table S.2.4A). Many consistent DEGs are related to neurodevelopment and brain functions. For example, upregulated genes include *TBRI*, a key regulator of cortical development⁶²; *NPTX2*, neuronal pentraxin for synaptic plasticity⁶³; neuropeptide S receptor *NPSRI* in neuroendocrine cells⁶⁴; *ADRA2B*, which regulates neurotransmitter norepinephrine⁶⁵; *ADORA2A*, linked to anxiety, arousal, and sleep regulation⁶⁶. Downregulated genes include *FEZF1* for neuronal differentiation⁶⁷; *LHX5* associated with mammillary body development⁶⁸; *GPR50* for neural progenitor cell differentiation⁶⁹; *CACNA1B*, a calcium channel subunit for neurotransmitter release⁷⁰. THC Cluster 2 showed even lower consistency, with only two consistently downregulated genes, B Cell receptor *CD72*⁷¹ and presynaptic receptor *GRM8*⁷², and no significant pathways (Fig. 2.3, Table S.2.4B).

In contrast, CBD Cluster 1 exhibited robust transcriptional changes, with 1,005 consistently upregulated DEGs and 997 consistently downregulated DEGs (Fig. 2.4, Table S.2.5A). The upregulated pathways included negative regulation of growth, cellular response to zinc ions, response to ER stress, and inflammatory response. In contrast, the downregulated pathways were primarily related to development, ECM organization, and sprouting angiogenesis, and skin development. CBD Cluster 4 had 251 consistently upregulated DEGs and 283 downregulated DEGs (Fig 2.4, Table S.2.5B), but only the upregulated DEGs showed significantly enriched pathways, including negative regulation of growth and cellular response to zinc ions. The absence of strong downregulated pathways in Cluster 4 may explain the separation between the two large clusters and their weak negative correlation. CBD Clusters 2 and 3 displayed more limited consistency. CBD Cluster 2 only had 2 consistent DEGs: cellular zinc sensor *MTF1* upregulated⁷³ and *SENP3* in *Wnt* signaling downregulated⁷⁴ (Fig. 2.4, Table S.2.5C). CBD Cluster 3 had 73 consistently upregulated DEGs, 77 downregulated DEGs, and two enriched upregulated pathways: cellular response to lipopolysaccharide and inflammatory response (Fig. 2.4, Table S.2.5D).

Consistency in DEGs across Datasets:

When evaluating consistency in DEGs across datasets, again we found CBD induced broader and more reproducible gene expression changes compared to THC (Fig. 2.6A, Table S.2.6A, Table S.2.6B). Among the 30 most frequently detected DEGs, 22 exhibited (>70%) a consistent direction of regulation, meaning that they were detected as either consistently upregulated or downregulated across datasets (Fig. 2.6A, Table S.2.6A). For example, *SLC30A1*, a zinc transporter⁷⁵, was significantly upregulated in 23 CBD datasets and not downregulated in any. *HMOX1* (heme oxygenase 1), a gene involved in oxidative stress response⁷⁶, was significantly

upregulated in 19 datasets while downregulated in only 4 CBD datasets. Some other frequent DEGs play a role in ion metabolism, such as downregulation of *SLC39A10* for zinc transport⁷⁷ and upregulation of *MT2A* and *MT1E* for metal detoxification and protection against oxidative stress⁷⁸.

In contrast, DEGs associated with THC were less frequently replicated across datasets, and the most recurrent genes appeared only three times across 46 datasets (Fig. 2.6A, Table S.2.6B). *PRXL2A* (peroxiredoxin-like 2A), an antioxidant protein that protects cells from oxidative stress⁷⁹, was downregulated in three datasets. Other THC-associated upregulated genes included *BPGM* in red blood cell metabolism, *COL1A2* for collagen synthesis, microtubule-binding protein *FRY*, a potassium channel gene *KCNQ5*, circadian regulator *PER1*, *RALGAP2* for intracellular signaling, and *XRCC2* for DNA repair.

We also performed pathway enrichment analysis of the significant DEGs for each dataset and summarized the frequency of significant pathways across datasets. Again, THC studies exhibited minimal pathway-level consistency (Fig. 2.6B, 2.6C, Table S.2.7B), but CBD datasets showed recurring enriched pathways (Table S.2.7A). Consistent CBD pathways include the upregulation of the “metabolic pathway” across 17 studies (Fig. 2.6B) and the “cellular response to zinc ion” pathway across nine studies (consistent with the reproducibility of DEGs *SLC30A1* and *MT2A*), and the downregulation of cell cycle, mRNA export, cancer, and mismatch repair pathways (Fig. 2.6C).

Gene and pathway-level evidence for the endocannabinoid system in CBD and THC transcriptomic signatures

Given the well-established interaction between cannabinoids and the endocannabinoid system (ECS)^{80,81}, we anticipated enrichment of ECS-associated genes in the transcriptional signatures of

both CBD and THC. To systematically evaluate this, we curated a list of 135 ECS-associated genes using prior literature and established databases, including Gene Ontology Biological Process (GOBP) annotations⁵⁹ and the DisGeNET database^{82,83} (Table S.2.9). This gene set includes metabolic enzymes involved in endocannabinoid degradation (i.e. *FAAH*, *MGLL*, *ABHD6*, *ABHD12*, *PTGS2*) and biosynthesis (i.e. *NAPEPLD*, *DAGLA*, *DAGLB*) of key ligands, such as 2-arachidonoylglycerol (2-AG) and anandamide (AEA). It also encompasses cannabinoid receptors (i.e. *CRN1*, *CRN2*, *GPR55*, *TRPV1*), neurotransmitter receptors (i.e. *DRD2*, *DRD4*, *HTR1A*, *SSTR4*), or nicotinic acetylcholine receptors (i.e. *CHRNA2*, *CHRNA3*, *CHRNA5*, *CHRNA6*, *CHRNA7*, *CHRNAB3*), endocannabinoid transporters (i.e. *FABP1*, *FABP3*, *FABP5*, *FABP7*), and transcriptional regulators (i.e. *PPARA*, *EP300*, *FOXP2*).

We examined the regulation of ECS genes identified as DEGs and found that more CBD datasets had significantly regulated ECS genes compared to THC (Fig. 2.8A, Table S.2.10). In addition, THC datasets generally exhibited fewer ECS DEGs than CBD datasets. Notably, most THC datasets with significant ECS gene regulation were brain-derived, whereas CBD datasets predominantly originated from peripheral tissues. A cluster on the left side of Fig 2.8A, composed mainly of peripheral CBD datasets, exhibited some consistency in regulatory direction. For instance, consistently upregulated genes included *SQSTM1*, *SLC6A9*, and *BEST1*, while consistently downregulated genes included *CALM2*, *KIAA2013*, and *SMC2*. This limited degree of coherence suggests weak gene-level evidence for consistent ECS involvement in the transcriptomic effects of CBD and THC, although context-dependent, tissue-specific regulation of individual genes remains possible.

To evaluate whether ECS pathway genes were overrepresented (enriched) among DEGs, we performed Fisher's Exact Tests comparing the DEG lists to the ECS gene set. Significance was

defined as $p < 0.05$, with $p < 0.10$ considered suggestive (Fig. 2.8B, 2.8C, Table S.2.10). Overall, CBD showed more datasets with significant or suggestive enrichment compared to THC. Notably, all THC datasets showing significant and suggestive ECS gene set enrichment originated from brain regions, including the rat orbitofrontal cortex, mouse amygdala, mouse dorsolateral striatum, and mouse medial prefrontal cortex (Table S.2.10). In contrast, approximately half of the significantly or suggestively enriched CBD datasets came from peripheral tissues and cell lines, including the mouse liver, human liver cell line, monocyte cell line, and microglial cell line. These findings support that THC's regulatory effects are brain-specific, whereas CBD exhibits more pleiotropic effects across central and peripheral systems.

In addition, we investigated pathways related to pain and inflammation, which are regulated by the endocannabinoid system³³. However, no significant enrichment was found for the GOBP “sensory perception of pain pathway” and “response to pain pathway”⁵⁹. However, the inflammatory response pathway was enriched among DEGs from two CBD datasets, both derived from monocyte-derived dendritic cell lines.

THC- and CBD-induced brain network modeling identifies unique key drivers

We next focused on the brain datasets on THC and CBD to identify key similarities and differences in the regulation of cannabis-induced transcriptomic alterations in the brain. We applied weighted key driver analysis (wKDA) to DEG sets from THC- and CBD- treated brain datasets to identify potential key drivers (KDs) using a Bayesian gene regulatory network previously constructed from dozens of human and mouse brain datasets^{21,22}. KDs with a false discovery rate (FDR) < 0.05 were considered statistically significant.

For CBD-treated brain datasets, 14 significant KDs were identified (top 5 shown in Fig. 2.7A; complete list in Table S.2.8). The KDs in the CBD network were derived from multiple brain cell lines, including mouse brain microglia, rat primary hippocampal neurons, and human neuroblastoma cells (Fig. 2.7A). Most KDs were involved in neural development and synaptic functions, such as *SH3GL2*, *DPYSL5*, *JPH3*, *SLC17A7*, and *GRIN3B* (*MMT00076709*). Other KDs were associated with signal transduction and kinase activity (*PRKCA*, *TAOK1*, and *ADRBK1*), membrane transport (*SLC9A3R1*), metabolism (*PPP1R3B*), gene regulation (*USF1*), and immune responses and cancer (*SLAMF8*, *SRC*, and *KIAA0100*).

Despite the lack of consistent DEGs and pathways across THC-treated brain datasets, we identified a total of 176 significant KDs based on DEGs from individual datasets (top 5 shown in Fig. 2.7B; complete list in Table S.2.8), among which *SLC17A7* and *TAOK1* were shared KDs between THC and CBD. Notably, most THC KDs were found to be significant in only a single DEG dataset rather than across multiple datasets (Fig. 2.7B), suggesting that THC may exert brain region-specific effects on gene regulation. KDs in the THC network were identified from multiple animal models, including the mouse hippocampus, mouse amygdala, mouse microglia, and rat orbitofrontal cortex. A majority of THC-associated KDs were involved in neuronal functions, signal transduction, structural integrity, and metabolic functions. Notably, some KDs with known neuronal roles include *SLC17A7*, *RTN4R*, *RTN4RL1*, *RTN4RL2*, *NEUROD2*, *NEUROD6*, *NGB*, *CAMK2A*, *ADCY1*, *SYT1*, *DLGAP1*, *DLGAP3*, *NTS*, *RGS14*, *GAD2*, and *CNIH3*.

CBD and THC DEGs are associated with neuropsychiatric disorders while CBD is associated with more metabolic traits

The endocannabinoid system plays a role in various health outcomes, including mood disorders, cardiovascular disease, stroke, cancer, diabetes, autoimmune conditions, and neurological disorders⁸⁴. To further investigate the connection between cannabinoids and human health, we analyzed publicly available summary statistics from large-scale Genome-Wide Association Studies (GWAS) for more than 100 diseases or phenotypic traits using Marker Set Enrichment Analysis (MSEA) via the Mergenomics package^{21,22}. This approach enabled us to assess whether signature genes from THC and CBD datasets were enriched for human disease/phenotype variants.

DEGs from both cannabinoids showed significant associations with Schizophrenia, type 2 diabetes (T2D), and depressive symptoms (Fig. 2.9). In addition, CBD transcriptomic signatures exhibited broader associations with lipid metabolism, including low-density lipoprotein (LDL), high-density lipoprotein (HDL), total cholesterol (TC), and triglycerides (TG), as well as with body composition traits such as waist circumference (WCadjBMI), hip circumference (HIPadjBMI), waist-to-hip ratio (WHRadjBMI), and height. In contrast, THC was linked only to WCadjBMI. Beyond metabolic and anthropometric traits, CBD was associated with Crohn's disease (CD) and coronary artery disease (CAD).

To identify the genes most strongly linked to disease-associated SNPs, we examined the top genes from the significantly enriched datasets (Figure S.2.4A, S.2.4B, S.2.4C, S.2.4D). In the CBD datasets, key associations for CD included tumor suppressor *CYLD*⁸⁵ and cytosolic receptor *NOD2* (Fig. S.2.4A). Mutations or altered expression of *NOD2* have been observed in CD patients⁸⁶, and genetic studies have identified *NOD2* locus polymorphisms and an independent involvement of the neighboring gene *CYLD* in CD⁸⁷. For T2D, a key association for CBD DEGs was the Wnt signaling pathway transcription factor *TCF7L2* (Fig S.2.4C). Carrying two copies of a common variant in this gene is associated with an approximately twofold increase in T2D

risk^{88,89}. CBD DEGs CDKN2A and CDKN2B were also among the top CAD-associated GWAS candidates, which are located adjacent to the lead CAD-linked SNP on the 9p21 locus⁹⁰ (Fig. S.2.4D). In contrast, THC showed more moderate levels of association between DEGs and disease-related SNPs. Of note, this analysis mainly focuses on overlaps between CBD/THC DEGs with genetic associations of diseases and does not implicate increases or decreases in disease risks.

2.4 Discussion

In this study, we systematically analyzed over 100 publicly available transcriptomic datasets to understand the molecular effects of THC and CBD across species and tissue types. By examining both global and significant gene expression changes from microarray and RNA-sequencing data, we found highly variable gene signatures for these two cannabinoids compared to other exposures such as PFOA, BPA, and estradiol. We also identified DEGs, pathways, network key drivers (KDs), and diseases/phenotypes associated with the significant DEGs. Our findings highlight significant variability in THC and CBD-induced transcriptional responses, with CBD demonstrating better consistency across studies, species, and central and peripheral tissue types compared to THC, which shows more dataset-specific effects in brain regions.

Our broad dimensionality reduction and correlation analyses revealed that CBD and THC datasets were more scattered in contrast with other well-studied chemicals PFOA, BPA, and estradiol, which displayed stronger internal consistency across datasets. In addition, neither THC nor CBD datasets exhibited strong clustering in the UMAP or formed larger clusters in the correlation analyses across species and tissues. This finding suggests that cannabinoid-induced gene expression changes are highly variable depending on combinations of experimental conditions, genetic or disease backgrounds, and prior chemical exposures. These results highlight

the challenge of defining a universal and predictable transcriptional signature for cannabinoids and align with previous literature indicating that cannabinoid effects depend on tissue, developmental stage, sex, and exposure conditions^{39,41,42,91}.

Our analyses on the DEGs and enriched pathways across datasets indicate that CBD has a more stable gene regulatory signature than THC. This regulatory stability agrees with CBD's well-characterized antioxidant, anti-inflammatory, and neuroprotective effects^{92,93}. In particular, two major CBD-associated gene clusters (CBD Cluster 1 and CBD Cluster 4) exhibited hundreds to thousands of consistent DEGs, with both CBD clusters exhibiting upregulation in two pathways: negative regulation of growth and cellular response to zinc ions. CBD was associated with higher frequency of consistent DEGs across datasets, such as *HMOX1*, *SLC30A1*, *MT2A*, and *SLC39A10*. Previous *in vitro* studies have identified *HMOX1* as the most upregulated gene and protein following CBD treatment. The proposed mechanism involves CBD-induced nuclear export and proteasomal degradation of transcriptional repressor BACH1⁷⁶. Another study suggests that CBD may influence *HMOX1* levels through the Nrf2 pathway, which regulates antioxidant defenses⁹⁴. These indicate that *HMOX1* is linked to antioxidant and anti-inflammatory properties under the impact of CBD⁹⁵. Notably, pathways related to metabolic regulation and cellular response to zinc ion were recurrently upregulated. Supporting this finding, a previous study in BV-2 microglial cells reported that CBD upregulated zinc-related genes (*Mt2*, *Ndr1*, *Mmp23*) and zinc transporters *SLC30A1* and *SLC39A4* while downregulating *SLC39A10* and *Zfp472*⁹⁶. Another study in autoimmune T cells found that CBD suppressed pro-inflammatory genes and enhanced oxidative stress-related genes, including *SLC30A1*⁹⁷. We now expand this finding to a much broader range of tissues and experimental conditions. Together, these results suggest that modulation of zinc homeostasis may be a key mechanism through which CBD exerts its antioxidant and anti-

inflammatory effects. The cell cycle and cancer pathway were top downregulated pathways across CBD-associated datasets. Prior studies have shown that CBD can arrest the cell cycle in the G0/G1 phase and promote cell death in gastric cancer cell lines, such as SGC-7901⁹⁸. Our study of around 50 datasets highlighting this pathway across studies further highlight the potential anti-cancer role of CBD in addition to its metabolic and immunomodulatory effects.

In contrast, THC datasets exhibited minimal gene-level and pathway-level consistency in dataset clusters and globally, reinforcing its more variable and context-dependent effects. The weaker transcriptional stability of THC may stem from its complex pharmacodynamics, particularly its biphasic effects on neural and immune signaling. Our analysis revealed several diverse central and peripheral nervous system targets of THC that may ultimately lead to these downstream differences in pathway enrichment. For example, among the most consistently enriched DEGs induced by THC was *EGR2*, which is associated with neuropathy and congenital hypomyelination of peripheral neurons⁹⁹. A number of THC target DEGs were also associated with synaptic signaling in the brain, and the onset of epilepsy and intellectual deficits, including *RPH3A*¹⁰⁰, *KCNQ5*¹⁰¹, and *FRY*¹⁰². Further, we also found that *PER1* was among our most consistently detected DEGs, which is a primary circadian pacemaker which has implications on human behavior and cognition¹⁰³. Some of these genes, such as *FRY* and *EGR2*, are associated with neuronal irregularities in region-specific targets (e.g. the cerebellum for *FRY* and peripheral neurons for *EGR2*). One possible explanation for the highly variable effects of THC lies in its high lipid solubility¹⁰⁴, which may influence its distribution across datasets depending on adipose content and tissue composition. Additionally, physiological barriers such as the blood-brain and blood-testicular barrier restrict THC accumulation in the brain and testes during acute exposure,

and similar protective mechanisms may exist in other tissues as well¹⁰⁵. These factors together may contribute to the heterogeneous effects of THC across individuals and biological contexts.

We also observed heterogeneity in the brain network through KDA. While the CBD network based on DEGs from individual datasets had only 14 KDs, the THC network contained over 100 KDs. However, most of these THC KDs were significant in only one DEG dataset, further supporting the idea that THC induces region-specific effects in the brain. Additionally, KDA revealed that the molecular mechanisms activated in the brain differ between THC and CBD. Although both networks featured KDs involved in neuronal function and signal transduction, only two KDs, *SLC17A7* and *TAOK1*, were shared between them. *SLC17A7* encodes vesicular glutamate transporter 1 (*VGLUT1*), which facilitates glutamatergic neurotransmission¹⁰⁶. Cannabinoids have been shown to influence glutamatergic signaling: CBD reduces neuronal activation in VGLUT+ neurons¹⁰⁷, while cannabinol upregulates genes associated with glutamatergic synaptic function¹⁰⁸. *TAOK1* encodes a serine/threonine-protein that functions as a MAP kinase kinase kinase (MAP3K) and regulates MAPK signaling cascade¹⁰⁹. Although direct evidence linking cannabis treatment to *TAOK1* regulation is lacking, THC has been shown to modulate microRNAs that target mRNAs of proteins involved in MAPK signaling, including *TAOK1*¹¹⁰. The limited overlap in KDs between the two networks highlights the distinct molecular and cellular mechanisms through which THC and CBD exert their effects in the brain. This divergence is consistent with our broader findings that THC and CBD regulate largely non-overlapping gene sets and pathways across different tissue types.

While the endocannabinoid system (ECS) is widely recognized as a primary target of phytocannabinoids such as THC and CBD⁸⁰, our multi-dataset analysis revealed that ECS-associated genes were regulated inconsistently across studies, both in direction and magnitude.

These inconsistencies might reflect several underlying factors, including the tissue-specific expression of ECS components as well as distinct pharmacological profiles of THC and CBD. For example, in terms of binding affinity, THC acts as a partial agonist at both cannabinoid receptors, while CBD exhibits negligible affinity for cannabinoid receptor 1 (CB1) and functions as a partial agonist at cannabinoid receptor 2 (CB2)^{111,112}. Despite this gene-level variability, pathway-level enrichment analysis demonstrated that the ECS-related gene set were significantly enriched in several datasets, comparable in frequency to some of the top-ranking Gene Ontology Biological Process (GOBP) and KEGG pathways (Fig. 2.6B, 2.6C). Importantly, a clearer divergence between THC and CBD emerged when ECS enrichment was stratified by tissue type. ECS enrichment associated with THC was exclusively observed in brain-derived datasets, where CBD-associated ECS enrichment was distributed more evenly across both central and peripheral systems, including immune-related cell types and liver tissue. This divergence likely reflects differences in receptor affinity and expression patterns. CB1 is abundantly expressed in the brain and to a lesser extent in select peripheral tissues, whereas CB2 is predominantly found in immune cells and exhibits limited expression in the central nervous system¹¹³. Thus, THC's higher binding affinity to CB1 compared to CB2 explains its brain-focused transcriptomic effects¹¹³, while CBD's lack of CB1 binding and moderate activity at CB2 likely underlies its broader effects across peripheral and immune contexts. The observed significant ECS enrichment in CBD-treated immune and liver datasets aligns with this pharmacological profile.

To further investigate the potential health implications of cannabis exposure, we performed Marker Set Enrichment Analysis (MSEA) to assess whether gene expression profiles associated with THC and CBD were enriched for genetic markers of human diseases. CBD-responsive genes showed significant enrichment for markers related to cholesterol and lipid metabolism traits,

aligning with prior studies indicating that CBD can modulate lipid profiles and holds therapeutic potential for managing lipid disorders¹¹⁴. CBD was also significantly associated with anthropometric traits, such as waist circumference, hip circumference, and height. A systematic review across multiple databases and registries reported that while cannabis use is generally associated with reductions in weight, waist circumference, and BMI, CBD specifically has been linked to increased body fat¹¹⁵.

Both CBD and THC show association with multiple diseases, such as Schizophrenia, type 2 diabetes (T2D), and depressive symptoms. The relationship between cannabis use and depression is particularly complex and potentially bidirectional. Studies relying on self-reported depressive symptoms suggest mixed short-term effects of cannabis use: a minority (20%) with increased depression and a majority (64%) with decreased depression¹¹⁶. However, a meta-analysis of 15 studies found that even a single THC administration induces depression and anxiety with large effect sizes¹¹⁷. Conversely, extended cannabis abstinence has been associated with significant improvements of depressive symptoms¹¹⁸. Longitudinal data also suggest a bidirectional or reinforcing relationship, as baseline depression has been significantly associated with increased THC use in e-cigarettes 12 months later¹¹⁹.

CBD and other non-psychoactive cannabinoids have been investigated in human clinical trials for their potential therapeutic benefits in T2D and Schizophrenia. Indeed, a randomized clinical trial showed that CBD reduced circulating resistin, a hormone associated with insulin resistance, and increased glucose-dependent insulintropic peptide (GIP), which plays a role in preserving pancreatic β -cell function in T2D patients¹²⁰. CBD has also demonstrated beneficial effects and a safety profile in Schizophrenia, where patients treated with CBD for six weeks demonstrated lower levels of positive psychotic symptoms¹²¹. In another randomized clinical trial

comparing CBD with the antipsychotic amisulpride, both treatments led to significant clinical improvement; however, CBD had a markedly superior side-effect profile¹²². Thus, our MSEA results, supported by human clinical evidence, highlight the broad therapeutic potential of cannabinoids, especially CBD, in metabolic and psychiatric diseases and underscore the relevance of cannabinoid-responsive genes in the genetic architecture of complex traits.

Here, we have conducted a comprehensive, multispecies, multitissue investigation of the metabolic effects of THC and CBD. One of the primary strengths of our study was the use of highly heterogeneous datasets, with variations in experimental design, dosing regimens, and sample characteristics. This allowed us to consider cannabinoid exposure in a number of biological contexts, including in co-occurrence with diseases, on top of exposure to other relevant chemical agents, and in different tissues or regions within those tissues (e.g. several different brain regions were covered by our datasets). Meta-analysis across species allows us to consider the maximal amount of data from translatable model organisms so we can draw human-relevant conclusions. Despite the comprehensive nature of our analysis, however, some important limitations should be acknowledged. First, while we applied normalization techniques to minimize batch effects, residual variability may have influenced our findings. However, the same procedures were applied to the datasets from the reference chemicals, where coherent clustering and coherence across datasets were found. Therefore, our findings are less likely due to technical artifacts but are more likely the results of intrinsic activities of the compounds examined. Second, the sample sizes for cannabinoid studies tend to be small, with most groups consisting of only around three replicates. This limitation reflects the current state of the field and underscores the need for larger, more coordinated genomic studies. To mitigate the impact of limited sample size, we adopted a meta-analytic strategy to integrate signatures across studies, enhancing robustness through aggregated

evidence. Third, the use of bulk transcriptomic data precludes cell type-specific resolution, which is critical for understanding the differential effects of cannabinoids on distinct cell populations. Future studies employing single-cell RNA sequencing or spatial transcriptomics, when datasets are available, could provide deeper insights into the cellular specificity of cannabinoid-induced gene expression changes. Additionally, while we focused on gene expression data, post-transcriptional and epigenetic modifications likely play a crucial role in cannabinoid-mediated effects. Integrating multi-omics approaches, including proteomics and metabolomics, could enhance our understanding of cannabinoid biology. Lastly, our study highlights the need for controlled, well-designed longitudinal studies to assess the long-term impact of THC and CBD exposure on human health.

In summary, this study provides a comprehensive transcriptomic analysis of THC and CBD across species and tissue types. Our findings demonstrate that CBD exhibits greater transcriptional stability compared to THC, with more consistent effects on metabolic, cell cycle, and inflammatory pathways. In contrast, THC-induced gene expression changes are highly variable, complicating the identification of robust molecular signatures. These insights have important implications for the therapeutic use of cannabinoids and highlight the less predictive nature of the long-term biological effects of THC. As the legalization and medicinal use of cannabis continues to expand, a deeper understanding of its molecular mechanisms will be critical for optimizing its clinical applications while minimizing potential risks.

2.5 Figures

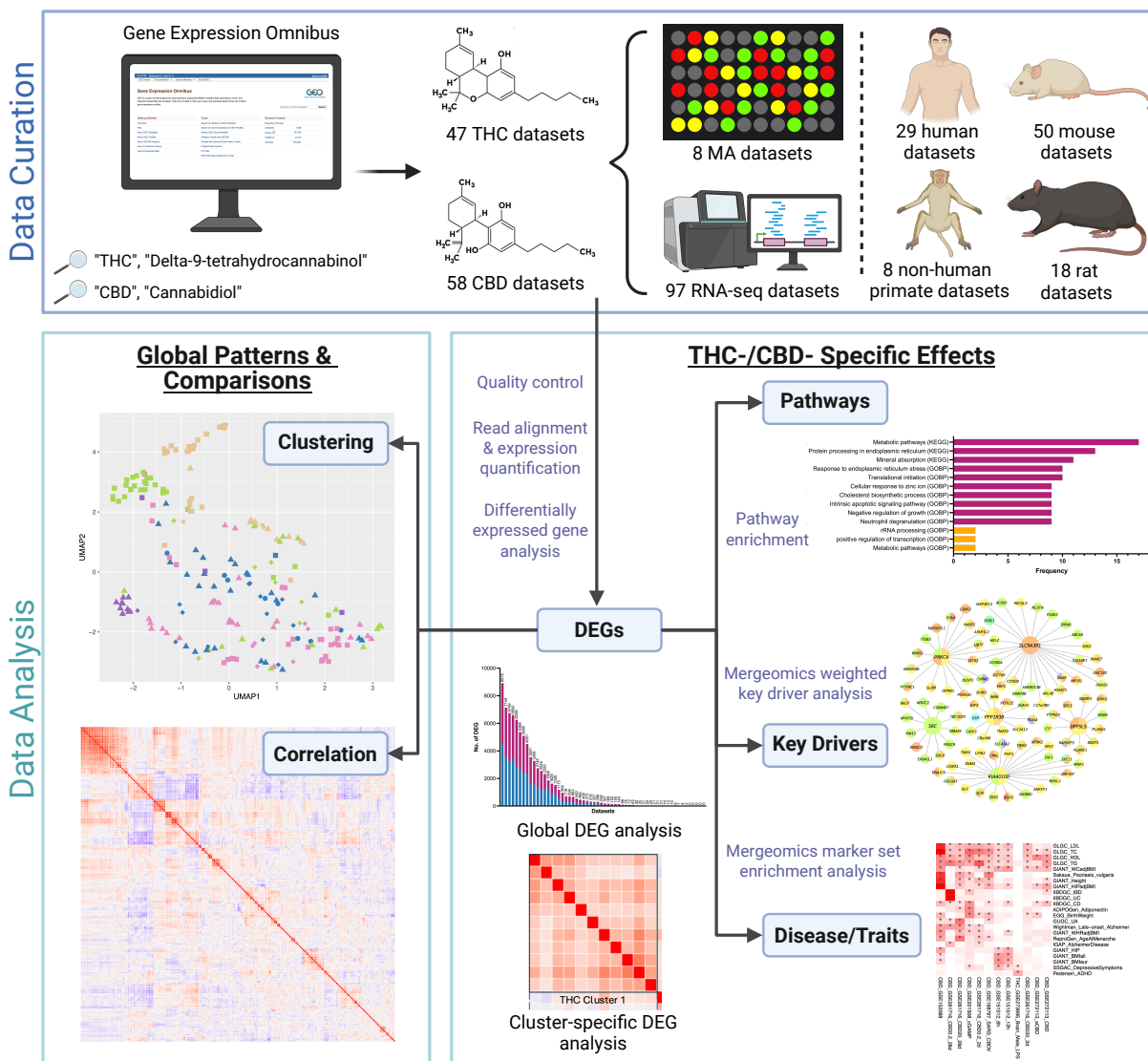


Figure 2.1. Overall study design. Datasets from THC and CBD exposure studies were curated from GEO and analyzed according to our established pipelines depending on whether they were cDNA microarray or RNAseq studies to detect differentially expressed genes (DEGs) across 4 mammalian species. Control chemicals, including PFOA, BPA, and estradiol, were processed similarly. Gene expression signatures were clustered using dimensionality reduction analysis and correlated using heatmaps derived from Spearman's rank correlation coefficients. DEGs were then analyzed for biological pathway enrichment, key regulators in brain gene regulatory networks, and disease enrichment using Mergeomics.

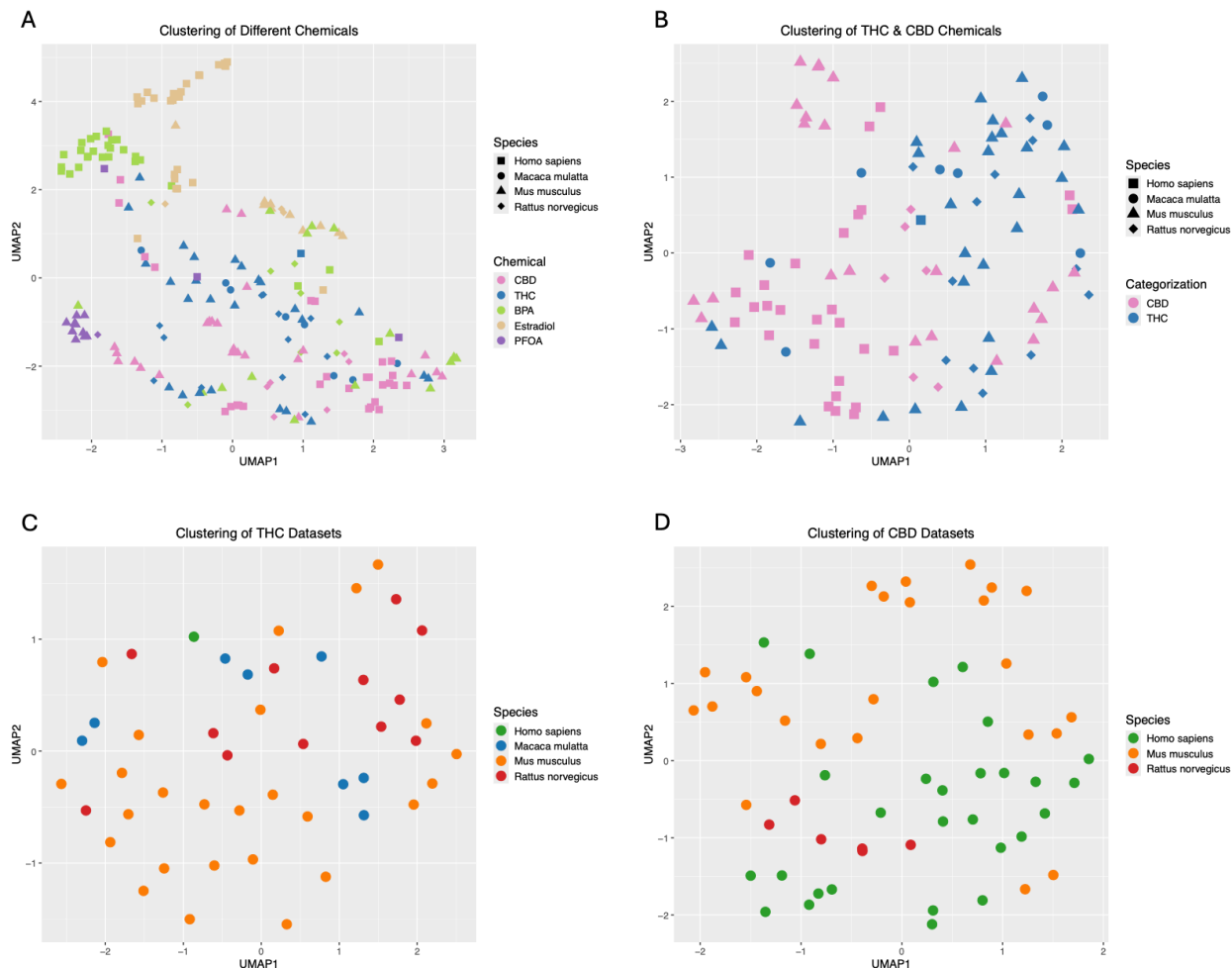


Figure 2.2. UMAP clustering reveals instability in DEG signatures across tissues in THC and CBD. Gene expression signatures were clustered using UMAP by the normalized log2 fold change induced in each gene between the treatment and control groups for THC and CBD, as well as the reference endocrine disrupting chemicals PFOA, BPA, and estradiol. Across all these chemicals, the top 2,500 most variable genes by log2 fold change were included in this analysis. Dots in the plot correspond to individual DEG signatures. In (A-B), the color of the dot corresponds to the chemical, and the shape of the dot corresponds to the species. **A)** PFOA (purple), BPA (green), and estradiol (yellow) all form relatively tight clusters on the left-hand side of the plot. THC (blue) and CBD (pink) signatures are more widely spread, indicating a higher degree of variability in the DEG signatures curated for this study. **B)** THC (blue) and CBD (pink) exhibit subtle separation along the UMAP1 dimension, indicating weak differences between these two chemicals in their DEG signatures. In (C-D), the color of the dot corresponds to the species. **C)** Within THC signatures, signatures derived from experiments on all four model organisms clustered together with no sharp distinction among any of them, revealing that none of the species has a species-specific response in their THC-associated DEG signatures. **D)** Within CBD signatures, mouse datasets and human datasets show subtle shifts along the UMAP2 dimension, and rat datasets cluster tightly, revealing that CBD exhibits weak species-specific effects on gene signatures.

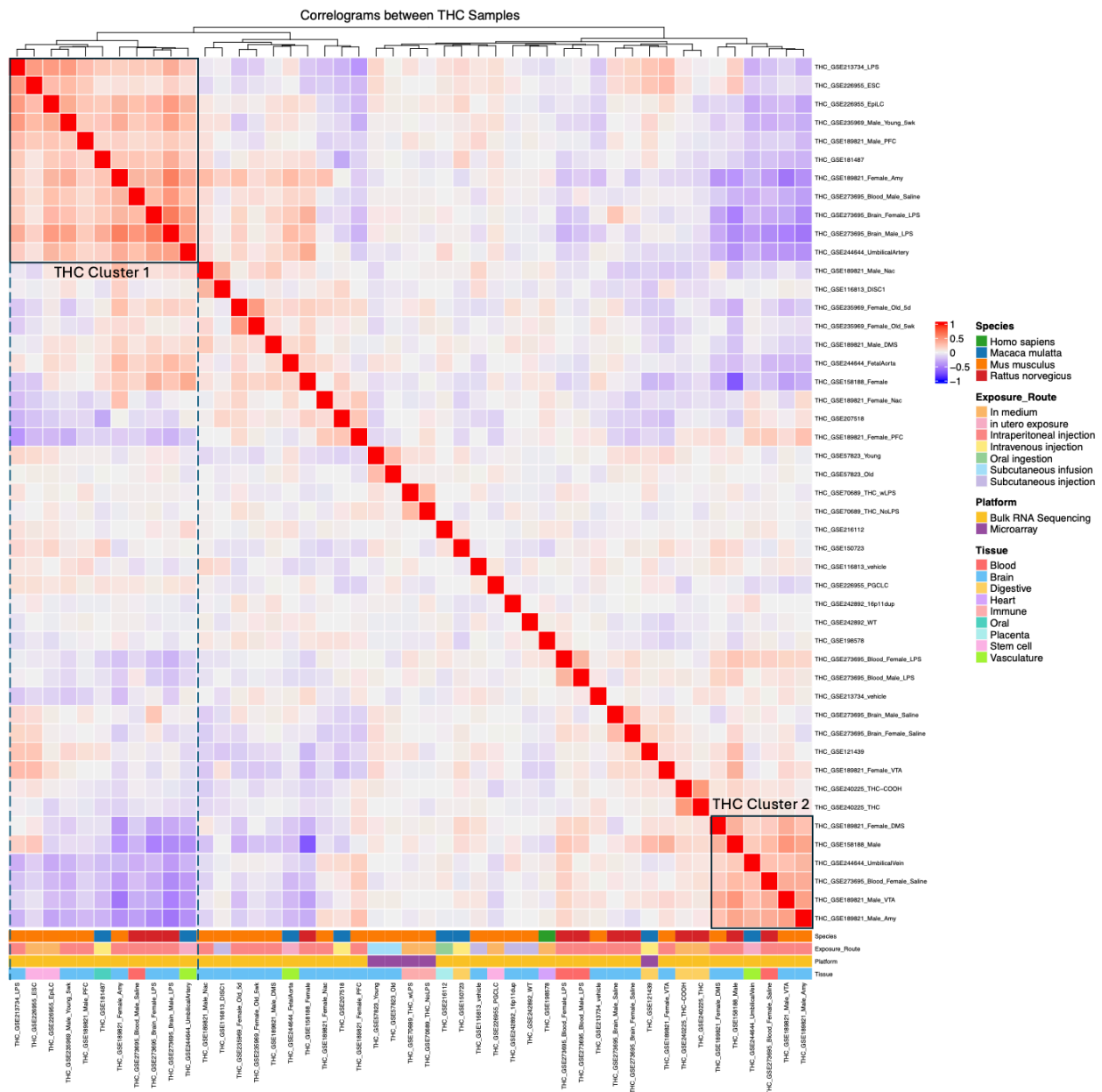


Figure 2.3. Hierarchical clustering reveals clusters of datasets with concordant gene expression signatures in response to THC exposure. The log2 fold changes across the genome for each THC dataset across were used to calculate Spearman's correlation coefficients and perform hierarchical clustering analysis. The results are shown in the heatmap. Red hues indicate concordant gene signatures (e.g. both datasets have their genes changing in the same direction and by a similar magnitude) while blue hues indicate discordant gene signatures. Two distinct clusters of concordant THC datasets are indicated, with cluster 1 containing 11 datasets in the upper left quadrant and cluster 2 containing 6 datasets in the lower right quadrant. Both clusters are composed mostly of brain-related datasets and contain datasets derived from multiple species. These two clusters exhibited moderate negative correlations with each other.

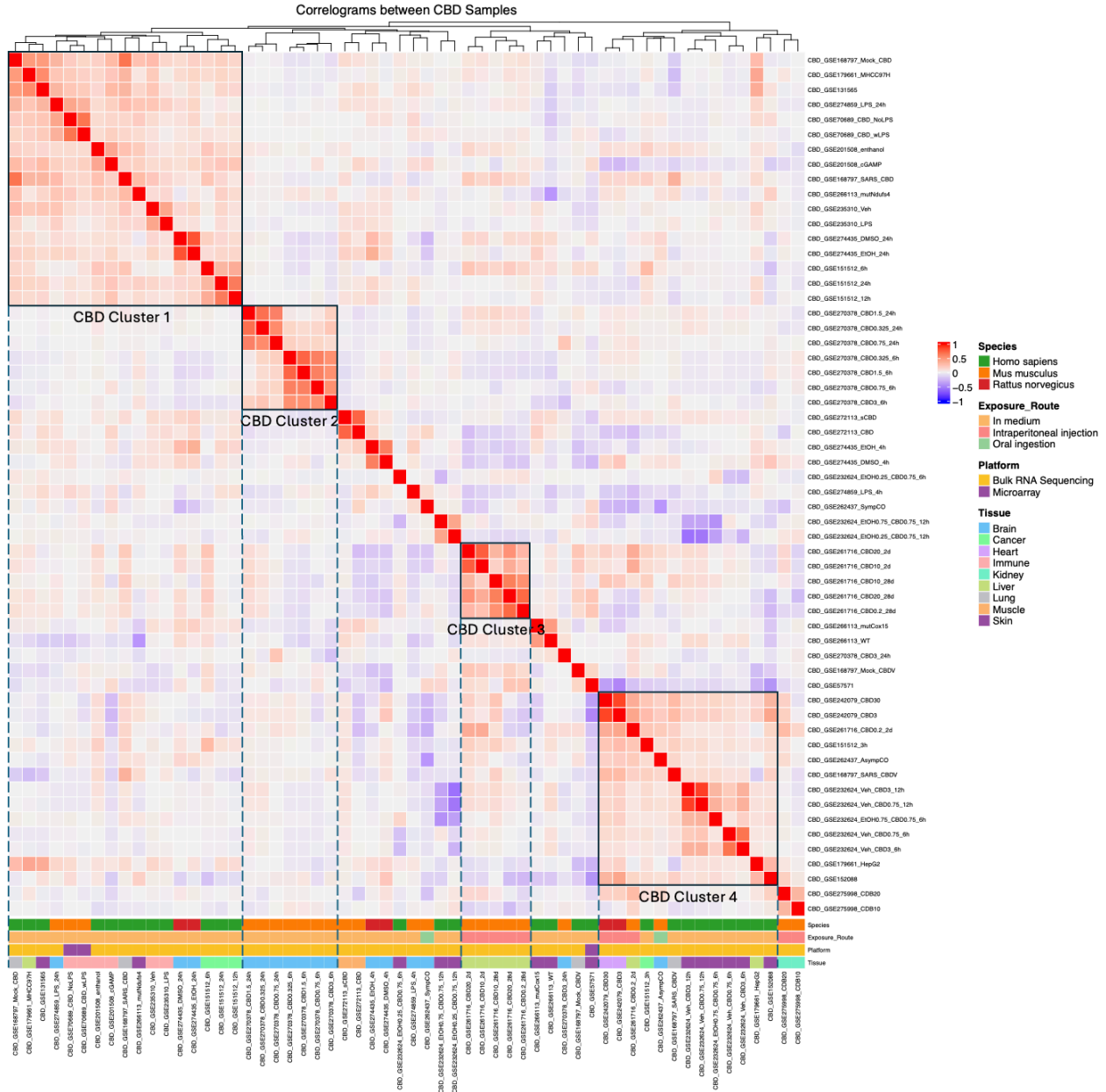


Figure 2.4. Hierarchical clustering reveals clusters of datasets with concordant gene expression signatures in response to CBD exposure. The log₂ fold changes across the genome for each CBD dataset across were used to calculate Spearman's correlation coefficients and perform hierarchical clustering analysis. The results are shown in the heatmap. Red hues indicate concordant gene signatures (e.g. both datasets have their genes changing in the same direction and by a similar magnitude) while blue hues indicate discordant gene signatures. Four distinct clusters of concordant CBD datasets are indicated, with cluster 1 containing 17 datasets in the upper left quadrant, cluster 2 containing 7 datasets in the center-left region, cluster 3 containing 5 datasets in the central right region, and cluster 4 containing 15 datasets in the lower right quadrant. Both

cluster 1 and cluster 4 contained datasets from multiple species and tissue types. Cluster 1 is purely *in vitro* datasets where cluster 4 has a mixture of multiple exposure routes and a high representation of skin-derived datasets. Clusters 2 and 3 were composed of datasets derived from one murine hypothalamus study and one murine liver study, respectively.

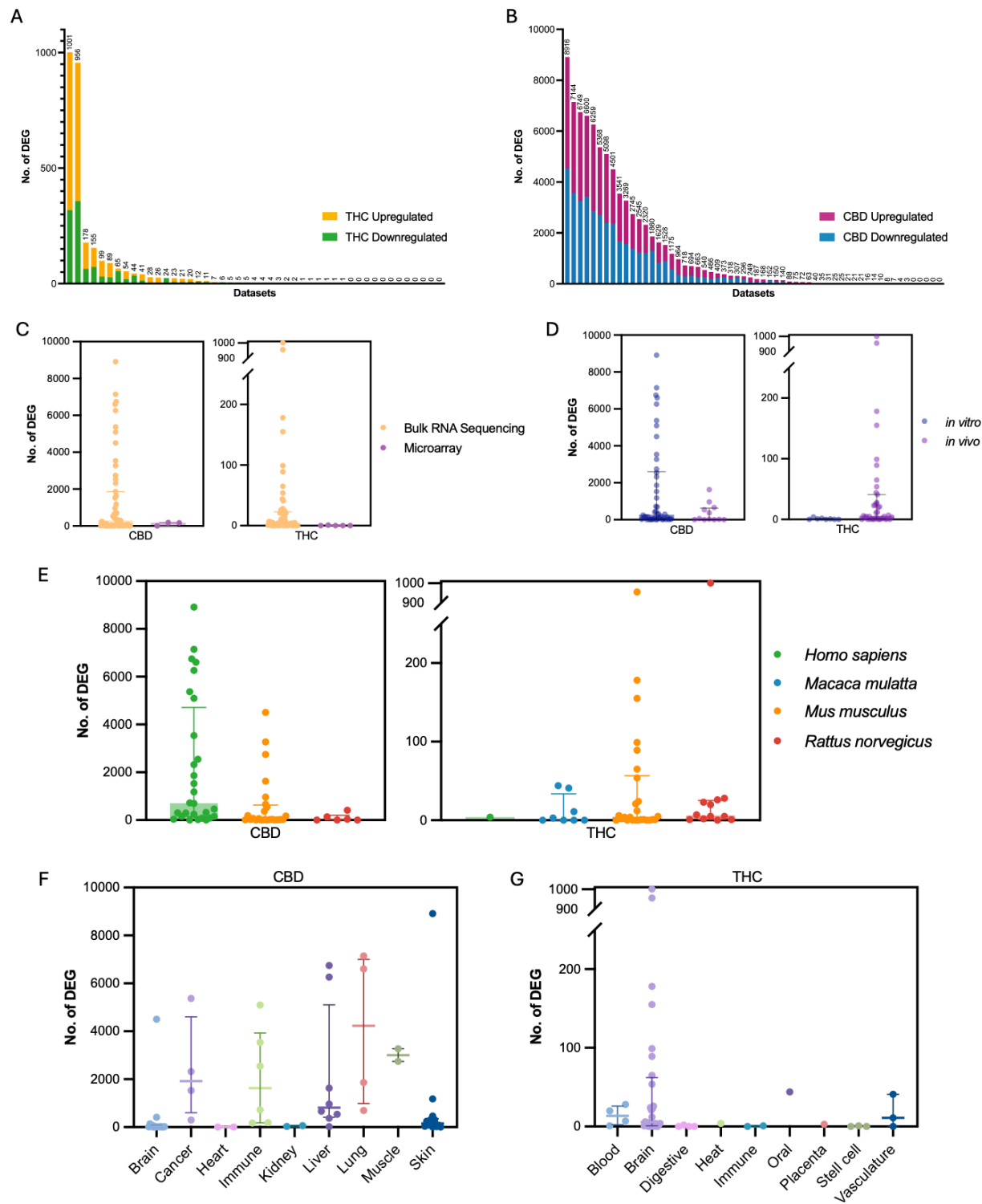


Figure 2.5. DEG results indicate CBD induces more DEGs across tissues and identify factors affecting the number of observed DEGs across studies. The number of DEGs observed across datasets was quantified according to chemical exposure and the potential confounding variables of sequencing platform, route of exposure, species, and tissue. **A)** A bar plot showing the number of DEGs detected in each THC dataset analyzed. Four datasets had hundreds of DEGs and

approximately half of the datasets had at least ten, while very low or zero DEG counts were observed in approximately half of the studies. **B)** A bar plot showing the number of DEGs detected in each CBD dataset analyzed. Nearly one third of datasets demonstrated more than 1,000 DEGs, and a significant portion of datasets also resulted in hundreds of DEGs. A small number of datasets with ten or fewer DEGs was also observed. **C)** A dot plot showing the amount of DEGs observed in each study for CBD (left) and THC (right), colored according to the sequencing platform used to derive each dataset. In general, RNAseq datasets are more sensitive to DEGs and more are observed in these studies than in cDNA microarray studies of the same chemical exposures. **D)** A dot plot showing the amount of DEGs observed in each study for CBD (left) and THC (right), colored according to whether the study design used an *in vivo* or *in vitro* exposure model. CBD datasets with *in vitro* exposure paradigms exhibit more DEGs while THC datasets with *in vivo* exposures exhibit more DEGs. **E)** A dot plot showing the amount of DEGs observed in each study of CBD (left) and THC (right), colored according to the species used in each study. CBD human and mouse datasets exhibited higher numbers of DEGs than rat datasets, while THC mouse datasets exhibited the most DEGs with slightly fewer in rat and macaque models. **F-G)** Dot plots showing the number of DEGs for CBD (**F**) and THC (**G**) datasets, colored according to the tissue examined in each dataset. CBD exhibited high numbers of DEGs in all tissues except the heart and kidney, with consistently high numbers of DEGs observed in the liver, lung, immune cells, and cancer cell line models across tissues. THC consistently exhibited high numbers of DEGs in brain datasets with far fewer DEGs in the rest of the tissues, suggesting brain specific gene perturbation effects.

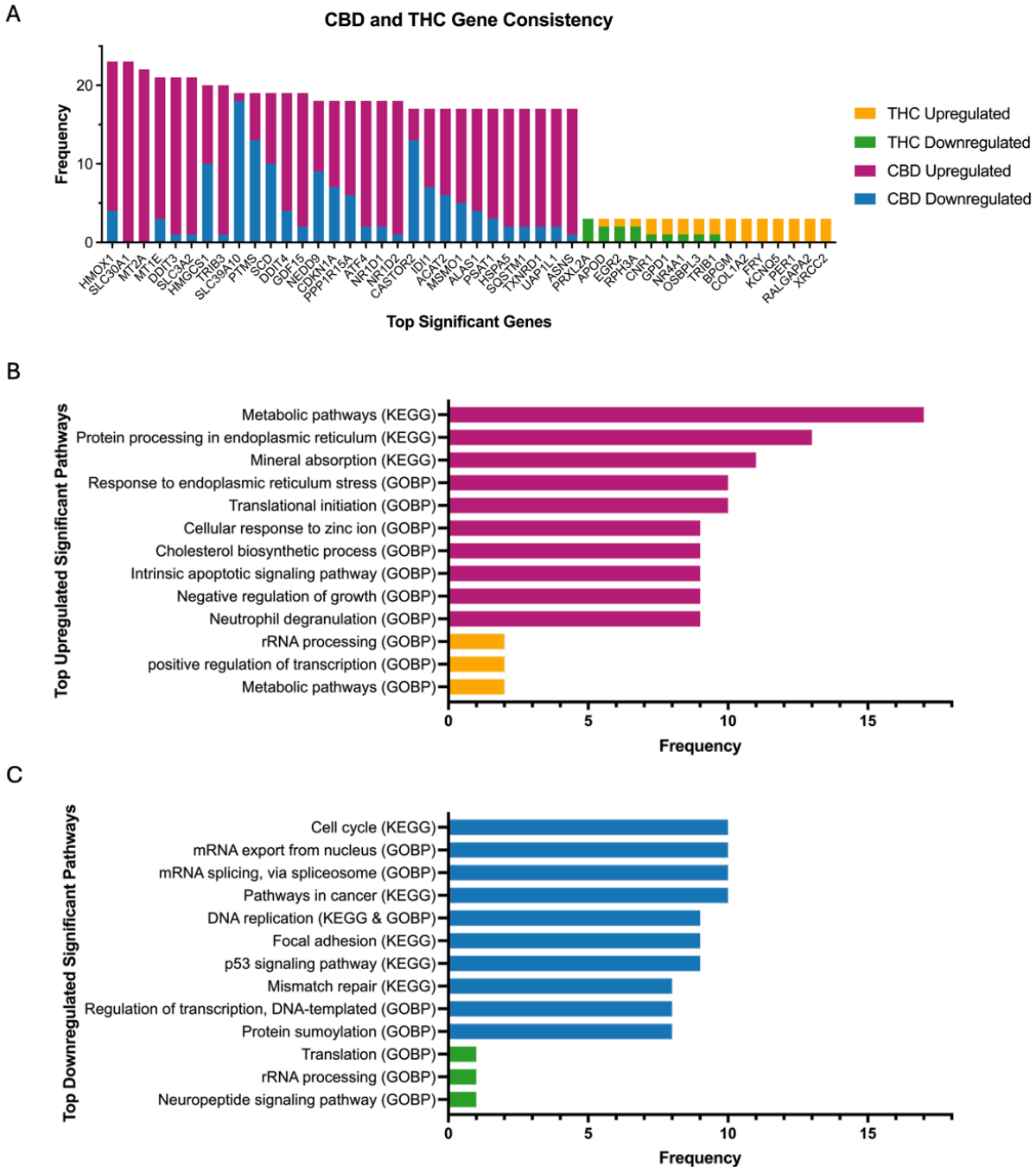
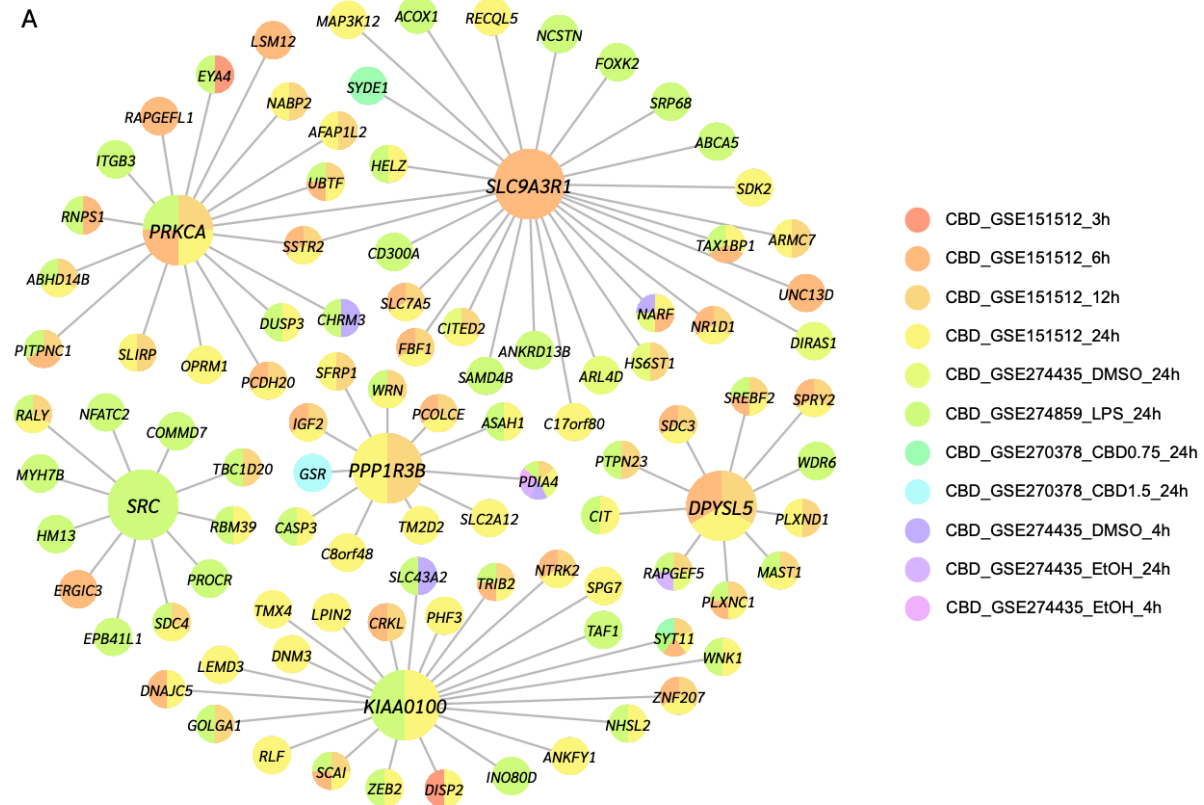


Figure 2.6. Meta-analysis reveals highly represented gene and pathway perturbations across cannabinoid exposure signatures. A) Recurrent genes that were detected as DEGs across CBD and THC exposure datasets are plotted on this bar graph. CBD tended to induce more DEGs and thus has recurrent genes that appear more often across datasets. THC had more variable gene perturbation effects and thus fewer genes were identified across multiple studies. Numbers of studies in which the genes appear as up- or downregulated DEGs are shown in each bar according to the color legend. **B-C)** Recurrent pathways that were detected based on the enrichment analysis of the DEGs for CBD (**B**) and THC (**C**). The bars are colored based on whether the pathway was

up- or downregulated according to the color legend, and the length of the bars indicates the number of exposure datasets in which the pathway was significantly enriched.

A



B

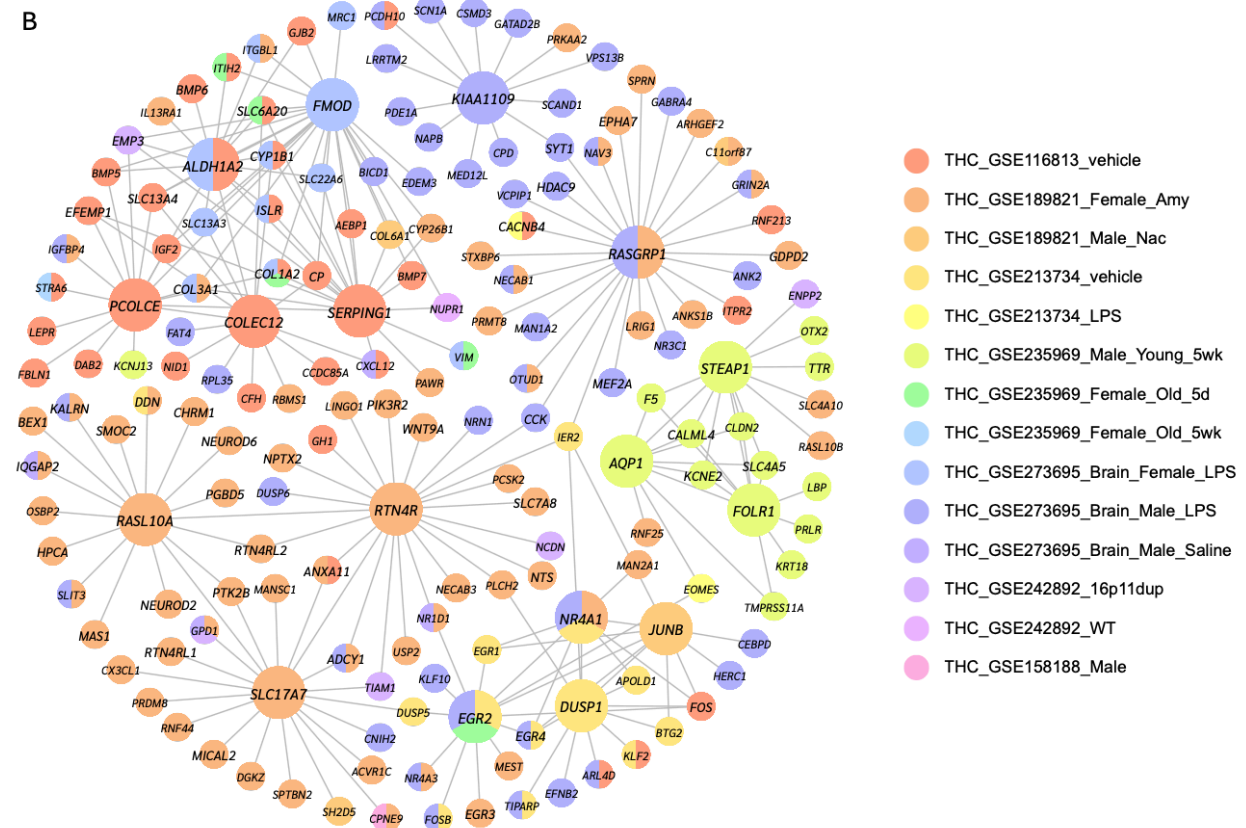


Figure 2.7. Key driver analysis identifies key regulatory genes in the perturbation effects of CBD and THC. Gene regulatory networks were constructed using key driver analysis from Mergeomics for CBD (**A**) and THC (**B**) brain exposure datasets. The colors of the network nodes indicate datasets in which the gene was a DEG, as described by the color legends. **A)** Brain dataset key drivers for CBD were primarily related to neural development and synaptic functions and were consistently detected across datasets from multiple model organisms and brain regions. **B)** Brain dataset key drivers for THC were primarily related to neuronal functions, signal transduction, structural integrity, and metabolic functions but were less consistently detected across different datasets, as most were uniquely DEGs in specific studies.

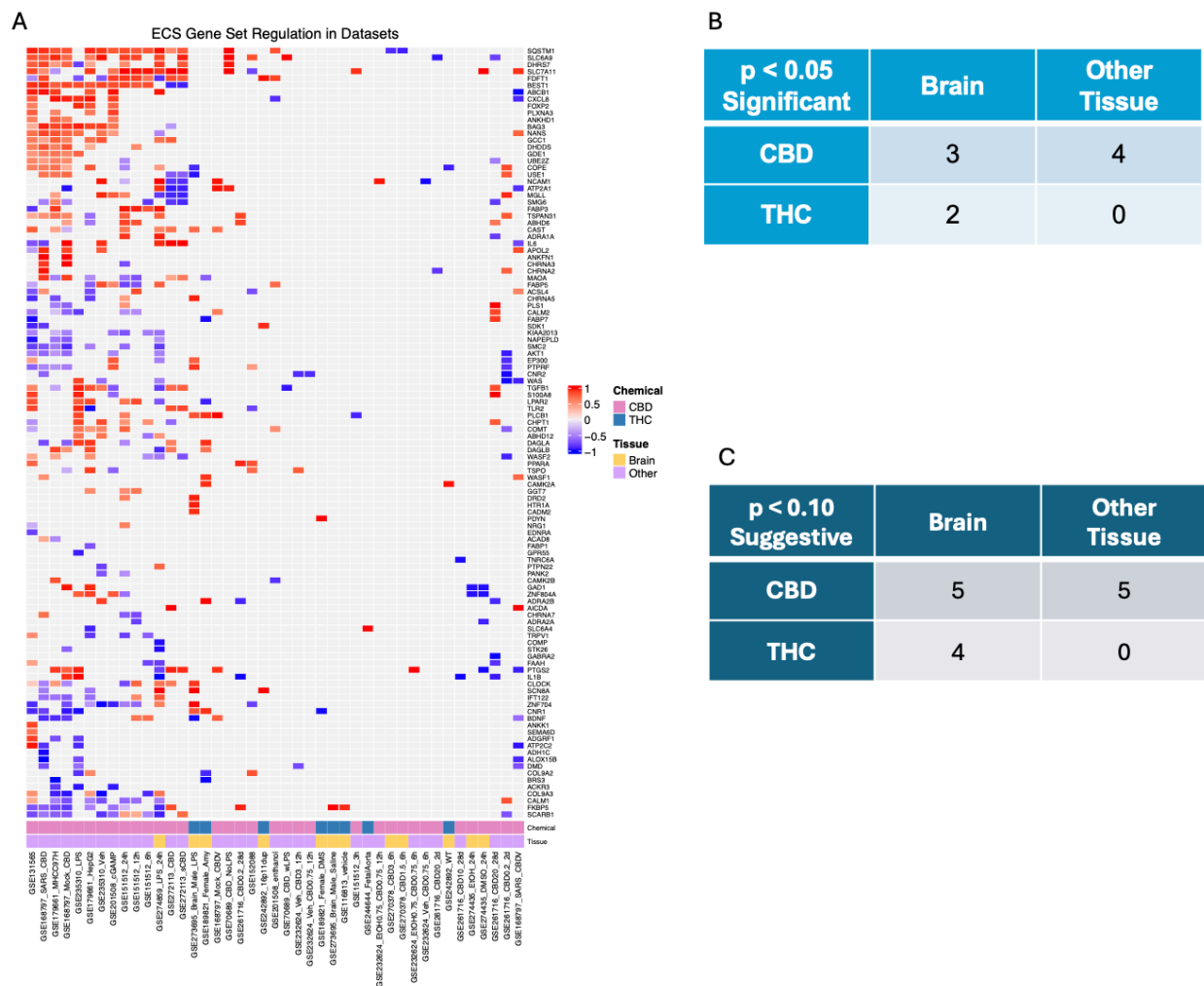


Figure 2.8. Correlation and overlap analysis reveals the extent of ECS gene involvement in CBD- and THC-induced transcriptome perturbations. **A)** The normalized log2 fold changes of the ECS genes detected as significant DEGs in each study are shown in this heatmap, with more highly upregulated genes shown in red hues and more highly downregulated genes shown in blue hues. Clusters of studies with similar directional changes of each gene indicate that the genes are consistently affected by CBD and THC exposure across datasets, and these results indicate ECS gene involvement in CBD- and THC-induced gene perturbations across studies. **B-C)** Fisher's exact test was performed to determine if the ECS gene set is significantly enriched in each CBD and THC dataset. The number of datasets derived from brain and peripheral tissue datasets that showed significant enrichment of the ECS gene set at a significant ($p < 0.05$, **B**) and suggestive ($p < 0.10$, **C**) threshold is shown in each table.

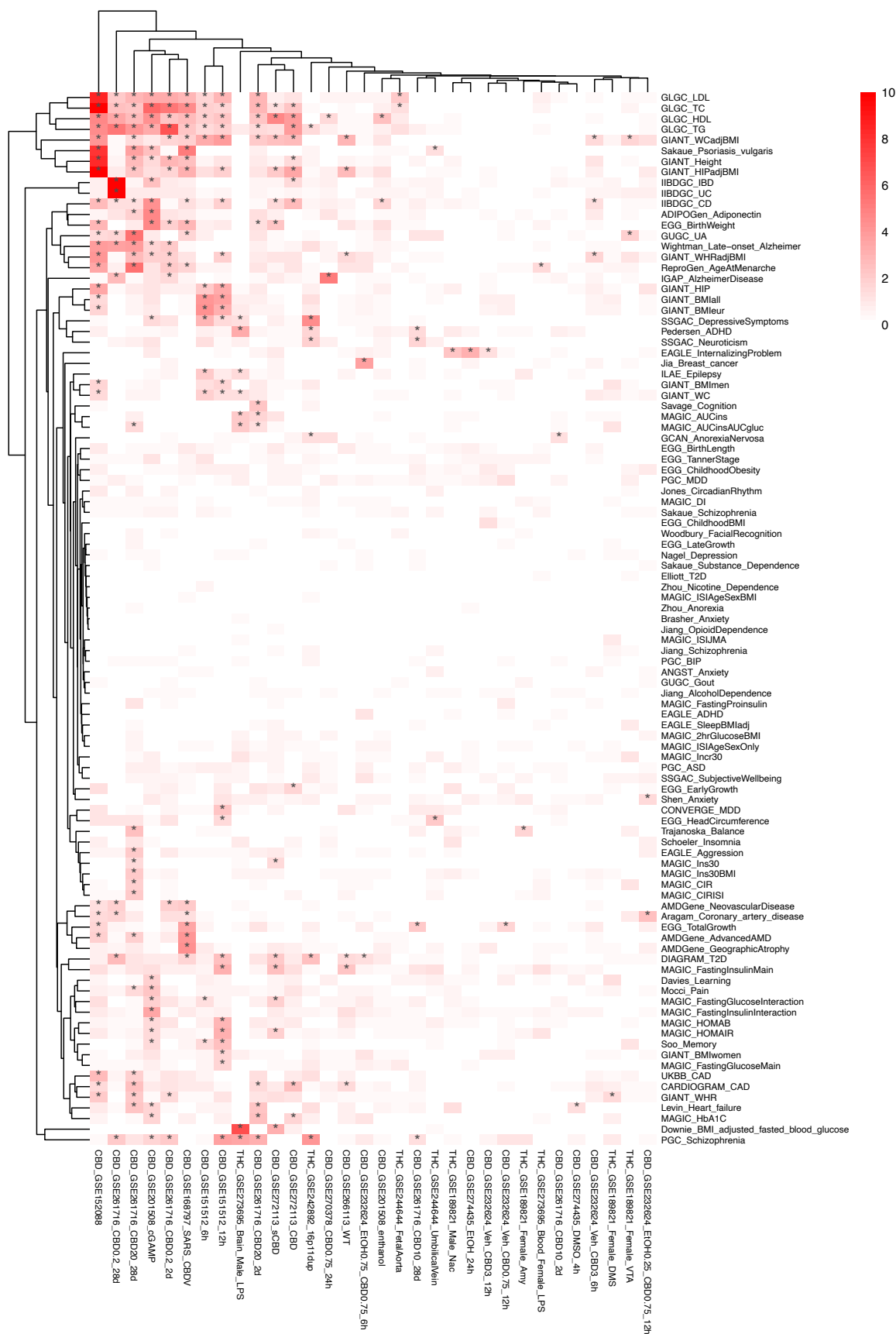


Figure 2.9. MSEA reveals associations between cannabinoid exposure and psychiatric, metabolic diseases. Marker set enrichment analysis (MSEA) from Mergeomics was performed on the DEGs from each dataset to identify disease associations based on marker genes for each disease. Only diseases or traits with at least one significantly associated dataset are shown. More highly associated diseases are shown with redder hues, and significant associations are starred. Psychiatric conditions, including schizophrenia and depression, and metabolic conditions, such as type 2 diabetes, were detected as significant associations with both CBD and THC exposure.

Chapter 3. ToxiOmics: A Species- and Tissue-specific Gene Signature Database and Web Tool to Understand Toxicological Exposures and Disease Associations

3.1 Introduction

Decades of industrialization has led to increased exposures of disease-inducing, toxicologically relevant substances such as industrial chemicals, metals, hormones, and psychoactive agents. However, the extent of adverse health effects as well as the underlying mechanisms remain poorly understood. Although animal studies have been indispensable in advancing our understanding of the exposome, the growing concerns for animal welfare and sustainability in chemical testing have led to numerous initiatives to develop and assess alternative methodologies to traditional animal studies by the National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM) and the National Institute of Environmental Health Sciences' Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM)^{123,124}. The overall goal of such initiatives is to develop and adopt novel alternative method (NAM) approaches, such as *in vitro* and *in silico* approaches to toxicological exposure studies, to reduce the contemporary reliance on animal studies, and relegate them to use on a much more selective basis^{1,123,124}.

The rise of big data and systems biology approaches has underscored the utility of gene expression (transcriptomics) data to inform toxicological mechanisms and toxicants' relations to diseases^{7,125–128}. Early efforts to catalog toxicant-induced transcriptomic effects in response to a variety of toxicant exposures include the Comparative Toxicogenomics Database (CTD)¹⁹, which aggregates the results of toxicant-induced differential gene expression analyses on a per-study basis in a centralized fashion, and the Crowd Extracted Expression of Differential Signatures (CREEDS)²⁰, which is a collection of community-supplied differential expression analyses in

response to perturbations with various drugs and toxicants. However, the CTD is constructed based on mining the literature for differentially expressed genes (DEGs) in response to toxicological exposures^{19,129}. This leads to an inherent bias and selective reporting for genes of interest in particular studies and the under-reporting of all significant DEGs in response to toxicological exposures. Likewise, CREEDS is a crowdsourcing effort²⁰ with inconsistent analytical methodologies, biases from individual analysts, and the lack of explicit focus on and limited data coverage for toxicological exposures. Due to these limitations that potentially lead to the under-coverage of toxicologically relevant genes and pathways, a new analytical methodology is needed to comprehensively identify and catalog toxicant-induced DEGs. Building on our recently established PharmOmics²³, which utilized a data-driven analytical pipeline to comprehensively curate and reanalyze high throughput transcriptomics data and catalog DEGs across species and tissues in response to pharmaceutical drug exposures. We have further tested the platform in pilot toxicogenomics studies for various endocrine disruptors⁴⁵ and cannabinoid exposures (Chapter 1). As an extension of the previous work, we now establish ToxiOmics, to utilize a data-driven analytical pipeline to determine toxicologically relevant DEGs and catalog them for query and downstream analysis by community users. The application of this data-driven pipeline has allowed the comprehensive cataloging and analysis across existing studies of various species, tissue types, dosage, and sample sizes to maximize power, evaluate reproducibility, and assess biological variability²³. This pipeline allows the segregation of toxicologically-relevant gene expression signatures by species, dose, and tissue type, enabling the compilation of highly refined and organized information to accelerate toxicology research.

We also construct an open access interrogable web tool which allows cross-species and cross-tissue comparisons of transcriptomics signatures in response to toxicological exposures on

both a per-study and inter-study basis, as well as the integration of network biology to elucidate toxicological effects, represents a significant advance in the field and a possible building block for further NAMs in toxicology.

3.2 Methods

Overall study design

The current version of ToxiOmics curated 304 studies of 222 toxicologically relevant compounds, whose raw sequencing data were posted to the Gene Expression Omnibus¹³⁰ 1,075 species-, tissue-, dose-, and time-segregated transcriptomics and pathway enrichment signatures. These data cover 29 tissues across 4 species (humans, mice, rats, and rhesus macaques). From these data, we also derived species- and tissue-specific meta transcriptomics signatures across doses and time points to understand the generalizable effects of toxicological exposures. We have compiled these data into ToxiOmics and constructed an analytical web tool to leverage network biology through integration with our existing Mergeomics^{21,22} (designed for disease network modeling) and PharmOmics²³ (designed for drug repositioning and drug toxicity prediction) tools. Integration of these tools derives gene regulatory networks (GRNs) perturbed by toxicological exposures, infers concordant and discordant mechanisms of action between toxicologically relevant compounds, and infers potential adverse effects between common toxicological exposures and pharmaceutical therapeutics. We demonstrate these utilities through a comparative analysis of legacy perfluoroalkyl substances (PFAS) and emerging “GenX” PFAS.

Data curation and inclusion

We developed a semi-automated pipeline combining automated search with manual checking to identify relevant datasets with appropriate chemical treatments. The automated process first extracts datasets containing chemical names or abbreviations in CTD with associated PubMed Identifier (PMID) and GEO Series (GSE) accession numbers to confirm dataset availability in GEO. Additionally, some chemicals of interest were individually queried in GEO according to nomenclature in the CDC's Toxic Substance Portal. GSE accession numbers identified by the automated process were manually checked to validate labels and exclude unusable false positive detected datasets. The resulting curated datasets met the following criteria for inclusion: 1) transcriptome data type (either cDNA microarray or RNA sequencing (RNASeq)); 2) derived from humans, mice, rats, or rhesus macaques to focus on mammalian species with translational value; 3) sample size of at least two replicates in each of the control and treatment groups (most *in vitro* studies involve duplicates per group and most *in vivo* studies involve 3 biological replicates per group); 4) not a duplicate entry on GEO (e.g. not part of a superseries or subseries). Studies with multiple species, tissues, doses, and time points were stratified into individual signatures if they satisfied the prior criteria.

Transcriptome download and processing

Microarray data were obtained directly from GEO using the GEOquery R package⁴⁸. Dataset metadata were processed, and cell lines and tissues were reannotated using the BRENDA Tissue Ontology¹³¹. Term processing followed the methodology described by Chen et al.²³ As microarray data submitted to GEO are pre-processed and quality-controlled, the resulting expression matrices were only checked for normalization and log2-transformed if necessary for downstream analysis.

RNAseq data were downloaded from the NCBI Sequence Read Archive (SRA)¹³² using the parallel-fastq-dump wrapper¹³³ to retrieve FASTQ files. Quality control was performed using Trim Galore (v 0.6.6)⁵⁰ and Cutadapt (v 2.1.0)⁵¹ in the Anaconda¹³⁴ environment. Briefly, Trim Galore was used to trim low-quality base calls at the 3' end of reads, followed by adapter sequence removal using Cutadapt. Short read sequences (< 20 bp) were filtered out.

The Salmon package (v 1.10.0)⁵² was used to map sequencing reads to appropriate reference genomes (human: GRCh38.p13¹³⁵, mouse: GRCm39¹³⁶, rat: Rnor_6.0¹³⁷, rhesus macaque: Mmul_10¹³⁸) and quantify mapped reads. The R package *Tximport* (v 1.22.0)⁵³ was used to import and summarize Salmon quantification results. Transcripts with fewer than one mean count across replicates were filtered out.

Differential gene expression analysis of stratified and meta transcriptomics signatures

For studies with multiple doses and/or time points, each dose and time was treated as a separate dataset to derive dose- and time-specific transcriptomics signatures. For single signatures, DEG analysis was performed by high performing gold-standard DEG analysis tools appropriate for different data types¹³⁹. Linear Models for Microarray Data (LIMMA)⁵⁴ was used for microarrays and DESeq2⁵⁵ was used for RNAseq data to infer DEGs at FDR < 0.05.

We further conducted meta-analysis to derive consensus DEGs for each chemical by species and tissue (e.g. across doses and time points), which likely represent generalizable targets of exposure across conditions, using the RobustRankAggreg R package⁵⁷ to perform robust rank aggregation. Briefly, genes were ranked by the magnitude of their log2 fold changes to determine the most significantly altered genes in each signature. Then, robust rank aggregation was applied

and genes at $FDR < 0.05$ were taken as meta-DEGs. The overall workflow of the data curation, signature generation, and integration into ToxiOmics is shown in Fig. 3.1.

Pathway enrichment analysis

To better understand the biological processes represented by the tissue-specific DEGs, pathway enrichment analysis of DEGs was performed on DEGs derived from single and meta signatures. Kyoto Encyclopedia of Genes and Genomes (KEGG)¹⁴⁰ and Gene Ontology Biological Processes (GOBP)¹⁴¹, and DisGeNET⁸² pathway databases were used to functionally annotate the DEGs. First, to ensure consistency in pathway annotation across species, the gene symbols of mouse, rat, and rhesus macaque DEGs were converted to human ortholog symbols using ortholog information from the HGNC consortium¹⁴². Enrichment analysis was then performed on all significant DEGs using the EnrichR R package⁵⁸. Biological pathways meeting the $FDR < 0.05$ threshold were incorporated into the ToxiOmics database.

Assessment of concordance and discordance in transcriptomic response to exposures across studies

To compare the species-, tissue-, dose-, and time-stratified transcriptomic responses in ToxiOmics, we assessed the correlative relationship in transcriptomic changes between studies and generated correlation-based clustering heatmaps. Briefly, the normalized log2FC values for all genes from each stratified transcriptome dataset are compiled into a dataframe in R, during which process the genes of each signature are converted to human orthologs using HCOP gene symbols for each species¹⁴³. The resulting dataframe, containing the normalized log2FC values across studies, is then used to calculate non-parametric Spearman's rank correlation coefficients between pairs of

transcriptomics datasets. These correlation coefficients are used to generate unsupervised clusters and heatmaps using the R package ComplexHeatmap¹⁴⁴, where datasets with similar transcriptomic patterns are clustered closer to each other.

Gene regulatory network key driver analysis

Previously, we constructed species- and tissue-specific GRNs from gene expression data, eQTLs from GTEx, and tissue-specific transcription factor and target gene pairs utilizing the Reconstructing Integrative Bayesian Networks (RIMBANet) tool¹⁴⁵. Following the selection of the appropriate species- and tissue-specific Bayesian GRN that match the exposure study design, unweighted key driver analysis (KDA) is performed on DEG signatures. KDA is performed using the associated function from Mergeomics^{21,22}. KDA detects potential key drivers (KDs) of toxicological DEG signatures based on the topology of the constructed Bayesian GRNs. A KD is defined as a gene connected to a significantly larger number of input signature genes compared to the expected number for a randomly selected network gene within a given GRN based on a Chi-square-like statistic²¹. KDs and input DEG signatures within the tissue-specific GRN are visualized using Cytoscape⁶⁰.

Measurement of DEG signature similarity between exposures

We used two different approaches to assess the similarity between the transcriptomic signatures of toxicologically relevant compounds, as developed in PharmOmics²³. The first approach is based on signature overlaps and uses a signed Jaccard score. The signed Jaccard score takes as input the upregulated genes in the first chemical signature ($a1$), the upregulated genes in the second chemical signature ($b1$), the downregulated genes in the first chemical signature ($a2$), and the

downregulated genes in the second chemical signature ($b2$), and is calculated according to the following formula:

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}$$

$$\text{signed Jaccard score} = J(a1, b1) + J(a2, b2) - J(a1, b2) - J(a2, b1)$$

The second method, as also established in PharmOmics²³, uses a distance measure derived from the mean of the shortest path lengths between network key drivers of the first chemical gene signature (A) and the second chemical gene signature (B) in a given Bayesian GRN based on a key driver analysis from Mergeomics^{21,22} as described above. The distance measure is calculated according to:

$$\text{distance}(B, A) = \frac{1}{\|A\|} \sum_{a \in A} \min_{b \in B} \text{distance}(b, a)$$

This distance measure, described in PharmOmics, is adapted from a previous study using protein interaction networks¹⁴⁶.

Web server implementation

To allow easy data access and use of ToxiOmics, we have created a freely accessible web tool deployed on the same Apache server used to host Mergeomics^{21,22} and PharmOmics²³. Integration with these resources allows for the use of the Mergeomics computational pipeline for integrative analysis of multi-omics datasets to derive disease-associated pathways, networks, and network regulators, and previously constructed Bayesian networks to conduct Key Driver Analysis (KDA) using our ToxiOmics DEG signatures.

The ToxiOmics tool contains functions to query and view exposure DEGs and pathways, assess disease associations, and compare chemical DEG signatures using gene overlap-based and

network-based approaches. The gene overlap-based approach using a signed Jaccard score, and the network-based approach using network repositioning, apply the same methodology as was used in the respective analytical methods described above and in PharmOmics²³. The web server allows for the selection of toxicologically relevant compounds to compare, and for the uploading of custom gene sets of interest to carry out these analyses to determine toxicologically relevant compounds that are enriched for gene sets of interest to the end user.

3.3 Results

Constructing the ToxiOmics database from species-, tissue-, dose-, and time-stratified DEG signatures of toxicologically relevant compounds

As illustrated in Fig. 3.2, we began with a list of 222 toxicologically relevant compounds cross-referenced between CTD and GEO. Our curated compounds encompass an array of toxicologically relevant exposures, including but not limited to persistent organic pollutants like perfluoroalkyl substances (PFAS), endocrine disrupting chemicals, air pollutants, metals, pesticides, and psychoactive compounds. These compounds were selected based on their presence in CTD with citations that also had associated GSE accession numbers in GEO for raw data access. The resulting GSE accession numbers were checked for suitability for inclusion based on our criteria, and subsequently analyzed to detect DEGs and enriched pathways in a species-, tissue-, dose-, and time-stratified fashion. Numbers of datasets, species and tissue coverage information, and platform and exposure type information are shown in Fig. 3.3. The stratified DEG signatures were also used to compute species- and tissue-specific meta-DEG signatures across doses and times of exposure to catalog the generalizable, conserved transcriptomic effects of toxicants across conditions. Inspection of the datasets indicated more human-derived signatures than any of the other species

(Fig. 3.3A), a higher coverage for the liver relative to other tissues (Fig. 3.3B), slightly more *in vitro* than *in vivo* study designs (Fig. 3.3C), and a higher number of cDNA microarray-derived signatures than RNASeq signatures (Fig. 3.3D). This higher representation of microarray datasets is not particularly surprising, as cDNA microarray was the dominant technology in gene expression studies prior to the adoption of RNASeq in the more recent decade¹⁴⁷.

Web server implementation

To allow easy access and use of the ToxiOmics database, we have implemented it with our existing Mergeomics^{21,22} and PharmOmics²³ tools to provide chemical DEG and pathway signature query, species- and tissue-level comparison, comparison between chemicals using gene overlap- and network-based approaches, and chemical GRN visualization (Fig. 3.4). The ToxiOmics web server contains three main functions. First, both stratified and meta DEG and pathway signatures can be queried, and species- and tissue-level comparative analyses to determine similarities and differences between chemicals can be conducted. Second, chemicals can be compared in more detail via an overlap-based and correlation-based approach. In this function, chemicals can be compared for transcriptomic effect similarity based on the degree of overlap of their up- and downregulated DEGs using the signed Jaccard score approach established in PharmOmics²³. Chemicals can also be compared using a global correlation -based approach by calculating Spearman's correlation coefficient across the changes (logFCs) of all genes for each chemical and performing hierarchical clustering analysis. The resulting clustering patterns, along with the relative similarity of each study based on the measured perturbations in gene expression, are output as heatmaps. The third function of the ToxiOmics web server utilizes a network-based approach to compare chemicals to each other and to pharmaceutical drugs from PharmOmics. This Bayesian

GRN distance measure-based approach, outlined in our methods and described in detail in PharmOmics²³, assesses the subnetworks impacted by chemical-induced transcriptomic changes, and can be used to infer concordant or discordant effects based on the directionality of within-subnetwork genes impacted by each chemical exposure. These features are all showcased in the following case study comparing legacy PFAS chemicals, which are being phased out due to their pollutant and chronic health effects^{148–152}, with emerging replacement “GenX” shorter-chain PFAS chemicals¹⁵³

PFAS as a use-case for ToxiOmics

PFAS are a class of persistent organic pollutants which have been used in a variety of industrial applications¹⁵⁴. Elevated PFAS exposure levels are associated with numerous disorders including hypercholesterolemia, hepatic steatosis, thyroid disease, ulcerative colitis, and cancers of the liver, kidney, breast, and testes¹⁵⁵. The fluorocarbon bonds of PFAS make them highly resistant to degradation and enable accumulation in the liver, kidney, bone, and blood, with a slow serum elimination half-life of 4-5 years¹⁵⁶. PFAS generally fall into two categories. “Legacy” PFAS include representative compounds perfluorooctanoate (PFOA), perfluorononanoate (PFNA), and perfluorooctane sulfonate (PFOS). This category of PFAS was widely adopted in applications including food packaging and firefighting foam before their toxicological risk profiles were properly understood, leading to a significant lag in toxicity evaluation and exposing the population to then-unknown health risks^{151,157}. After the toxic effects were identified, these compounds were phased out and replaced with the ostensibly safer “GenX” compounds such as hexafluoropropylene oxide dimer acid (HFPO-DA). While GenX compounds are more rapidly

broken down due to their shorter fluorocarbon chains, they have been found to exhibit toxic effects in their own right¹⁵⁸ and are potential precursors of legacy PFAS¹⁵⁹.

Due to the relatively understudied nature of the emerging GenX PFAS compounds, we sought to gain mechanistic insight based on the comparison of their transcriptomic effects to those of the legacy PFAS compounds. In ToxiOmics, 21 signatures derived from 8 GSEs were curated for legacy PFAS, and 5 signatures derived from 2 GSEs for GenX. Among these, mouse liver studies dominate. We began by comparing the meta DEG and pathway signature sizes for the mouse liver using the first function of ToxiOmics (Fig. 3.5). At the DEG level, GenX had the highest number of both total and unique meta DEGs, with PFOA having the next highest numbers (Fig. 3.5A). At the pathway level, however, PFOA had the highest number of both total and unique perturbed pathways (Fig. 3.5B). These comparison results suggest high degrees of perturbation exhibited by both GenX and PFOA.

To further investigate potential similarities in transcriptomic perturbations upon exposure to legacy and emerging GenX PFAS, we then conducted a correlation analysis using the second function of ToxiOmics to assess concordance of changes (logFC) in global gene expression between chemicals (Fig. 3.6) using Spearman's correlation and hierarchical clustering analysis. The clustering heatmap revealed a large central cluster of PFOA and HFPO-DA (GenX) transcriptomics signatures across doses and developmental stages (e.g. effects on the maternal vs. fetal livers after *in utero* exposure to HFPO-DA) (Fig. 3.6, boxed in blue). By comparison, other PFAS chemicals showed weaker correlations and clustering with PFOA and HFPO-DA (Fig. 3.6, boxed in green), but retained positive correlations, indicating weaker similarity in transcriptomic perturbations of the murine liver.

At the pathway level (Fig. 3.7), PPAR signaling and lipid metabolism are strongly perturbed across all the PFAS compounds, with additional alterations in insulin signaling and resistance, and FoxO signaling and cell cycle alterations. These results are consistent with the documented effects of PFAS exposure^{160,161}. PFAS exposure has also been associated with insulin resistance and diabetes onset in epidemiological studies^{162–164}. Perturbations in cell cycle regulation and FoxO signaling are associated with cancer^{165–168}, and are consistent with the status of PFAS as potential hepatocarcinogens. The enrichment for these pathways not just in the legacy PFAS, but also in the relatively understudied GenX, indicating that similar mechanisms of toxicity and disease mediation may be at play.

To evaluate network level similarity or differences across PFAS chemicals, we leveraged the integration of ToxiOmics with Mergeomics to conduct key driver analysis (KDA). We took the top 300 DEGs by FDR (all genes had FDR < 0.05) in the murine liver meta DEG signature of each compound, converted to human ortholog gene symbols, and conducted KDA as outlined in Mergeomics using the Bayesian human liver GRN^{21,22}. This analysis revealed 12 key drivers (KDs), with 9 of those KDs being detected as DEGs in multiple PFAS compounds, and 10 of them being shared KDs across multiple PFAS compounds (Fig. 3.8). These KDs reveal significant shared biological perturbations induced in both legacy PFAS and GenX exposures. Among the most significant KDs in both legacy PFAS and GenX exposures were *VNNI*, an oxidative stress sensor heavily implicated in hepatic gluconeogenesis and PPAR α -driven lipid metabolism^{169–171}, and *PEX11A*, a peroxin with significant roles in PPAR-mediated lipid metabolism and the reaction to oxidative stress^{172–174}.

The other KDs shared between the legacy PFAS and GenX are also largely involved in lipid metabolism, and perturbations in their expression as detected here are linked to the

progression of associated liver diseases. *ACOT2*, a shared KD between all four compounds, drives mitochondrial fatty acid oxidation¹⁷⁵, regulates Acyl-CoA metabolism which initiates fatty acid metabolism¹⁷⁶, and is associated with abnormal lipid metabolism and worse outcomes for cancer patients¹⁷⁷. *CD36* has been thoroughly linked to lipid metabolism and deposition^{178,179}, so much so that the degree of its involvement has drawn the attention of researchers as a potential target for therapeutic intervention to counteract Non-Alcoholic Fatty Liver Disease (NAFLD)^{180,181}. Our network analysis also implicates *CIDEA* as a discordant KD between the legacy PFAS and GenX, as this gene was shared between only the three legacy PFAS compounds (Fig. 3.8). However, *CIDEA* has a number of relevant functional associations implicated in both the legacy PFAS and GenX which indicate a concordant mechanism. *CIDEA* has been documented to have functional associations with lipid droplet formation and lipid deposition^{182,183}, PPAR α signaling in the development of hepatosteatosis¹⁸⁴, PPAR γ in lipid droplet aggregation in the liver¹⁸⁵, and triglyceride accumulation in adipocytes¹⁸⁶. Taken together, these results indicate that both legacy PFAS and GenX increase the risk of liver disease through lipid metabolism, specifically through PPAR α - and PPAR γ -mediated processes.

Finally, several of the KDs have also been implicated in liver fibrosis and cancer. The shared KDs *DECRI* and *SLC22A5* have been linked to liver fibrosis and non-alcoholic steatohepatitis (NASH)¹⁸⁷ and liver cancer-relevant intercellular transport¹⁸⁸, respectively. *TAT*, a legacy PFAS-specific KD (Fig. 3.8), has been shown to enhance hepatocarcinogenesis pathways^{189,190}. Another legacy PFAS-specific KD, *WEE1*, has been associated with liver cancer and cisplatin resistance *in vitro*^{191,192}. These findings are consistent with epidemiological studies which have associated increased PFAS levels with higher incidences of liver cancer and thus raised concern about PFAS as potential hepatocarcinogens^{193,194}.

Leveraging PFAS mechanistic insights from ToxiOmics to predict potential interactions with therapeutic drugs

Due to the ubiquity of PFAS exposure in the human population^{195–197}, we sought to determine whether PFAS exposure interacts with certain pharmaceutical drugs used to treat diseases associated with the major PFAS target pathways. Thus, we focused on Pioglitazone, a PPAR γ agonist used to treat type II diabetes^{198,199}, and sought to leverage the third function of ToxiOmics and the integration of ToxiOmics with PharmOmics²³ to examine network level concordance between PFAS and Pioglitazone. Adapting the network-based repositioning approach described in PharmOmics²³, we utilized the third function of ToxiOmics to assess the similarity between the Pioglitazone liver meta signature and the liver meta signature of the legacy PFAS compound PFOA. The resulting human liver GRN, with shared member genes enlarged, is shown in Fig. 3.9. Genes altered in opposite directions by pioglitazone and PFOA included the PFOA KDs *TAT* and *CD36* (Figs. 3.8, 3.9), *APOA5*, which has been associated with NAFLD^{200–202}, and genes with associations to abnormalities in lipid metabolism and hepatic steatosis including *CTH*^{203,204}, *ITIH3*²⁰⁵, *CDO1*²⁰⁶, and *MEI*²⁰⁷. The identification of these oppositely-altered shared network member genes in the liver GRN by PFOA and pioglitazone implicate potential antagonistic effects that may reduce the efficacy of pioglitazone treatment in individuals with high levels of PFOA exposure.

3.4 Discussion

Here, we introduce ToxiOmics, a transcriptome-based database and web tool to systematically catalog and analyze transcriptomic signatures of toxicologically relevant exposures. We curated species-, tissue-, dose-, and time-stratified DEG and pathway signatures for 222 compounds and

determined species- and tissue-level meta DEG and pathway signatures. By integrating these with advanced systems biology approaches including network analysis from Mergeomics^{21,22}, gene overlap- and network-based signature comparison approaches from PharmOmics²³, and gene signature comparison, we provide a framework for mechanistic toxicology research and translational discovery. Our proof-of-concept application of ToxiOmics to PFAS chemicals illustrates the utility of this resource in elucidating conserved and diverging mechanisms of toxicity, identifying KDs perturbed by legacy PFAS and emerging GenX exposure, and evaluating potential chemical-drug interactions.

Our first major contribution is the assembly of a large-scale, curated resource of toxicant-induced transcriptomic signatures. Drawing on CTD and GEO, we identified 222 toxicologically relevant compounds with sufficient transcriptomics data to be incorporated into ToxiOmics. These compounds spanned diverse chemical classes and covered many categories of high toxicological relevance, including persistent organic pollutants, endocrine disruptors, metals, pesticides, air pollutants, and psychoactive compounds (Fig. 3.2).

The curated database revealed several patterns reflective of the broader history of toxicogenomics research. Notably, there was a higher representation of human and liver datasets relative to other species and tissues, which is the result of the extensive use of *in vitro* human cell line studies in toxicogenomics research^{208,209,210}. The enrichment of liver datasets relative to other tissues reflects the liver's central role in xenobiotic metabolism which has made it the target of toxicologists. These species and tissue biases highlight both the strengths and limitations of currently available transcriptomics datasets, as although human and liver models provide rich, translatable data for specific mechanistic insights, coverage gaps exist for other species and tissues.

Nevertheless, the computation and inclusion of species- and tissue-specific meta signatures across doses and timepoints represents a significant advantage in understanding the broad effects of toxicological exposures. These meta signatures capture more robust transcriptional responses that are conserved across variable exposure conditions, demonstrating the core biology at play in these toxicological perturbations. We previously successfully applied this approach in PharmOmics, and meta analyses in general have long been used to enhance cross-study signal detection. Our application of this principle to toxicological exposures provides a new opportunity to derive consensus, data-driven toxicogenomic signatures and overcome experimental heterogeneity.

In this work, we used PFAS as a ToxiOmics case study to demonstrate the resource's capacity for mechanistic toxicology research. Murine hepatic meta-signatures showed that GenX produced the most DEGs and clustered tightly with PFOA across doses and developmental stages, with PFOS and PFNA showing weaker positive associations. These results are consistent with reports that PFOA and GenX elicit similar hepatic and intestinal histopathology, and liver metabolic and intestinal mucosal barrier impairments in mice²¹¹, and suggest that shorter-chain replacement PFAS are not mechanistically safer than their longer-chain legacy PFAS counterparts. Integrating ToxiOmics with Mergeomics KDA^{21,22} prioritized KDs, most of which were shared across compounds, linked to PPAR α - and PPAR γ -mediated lipid metabolism (e.g. *VNN1*, *PEX11A*, *ACOT2*, *CD36*) and processes central to PFAS-associated liver phenotypes^{212–215}. *CIDEA* appeared to be legacy PFAS-specific, however its functions in lipid droplet aggregation, PPAR γ -driven triglyceride accumulation, and steatosis align with the same lipid-centric mechanism^{182–185}. Cancer-related KDs (e.g. *WEE1*, *TAT*) further support epidemiologic associations between PFAS and hepatocellular carcinoma^{193,194,216,217}. Pathway enrichment

highlighted PPAR signaling, lipid metabolism, insulin signaling and resistance, FoxO signaling, and cell cycle regulation. These align with dysfunctional lipid metabolism²¹⁸, diabetes and insulin resistance^{219,220}, and cancer¹⁹³, which are concerning amid rising metabolic disease prevalence and the prevalence of PFAS exposure²²¹. Finally, via PharmOmics²³ integration, we demonstrated that PFOA and the PPAR γ agonist pioglitazone display oppositely directed subnetworks, highlighting how environmental exposures might antagonize or synergize with therapies and motivating targeted follow-up experiments.

As a whole, ToxiOmics serves as a powerful resource to derive toxicological insight and motivate future follow-up experiments. Our data-driven approach to derive comprehensive species-, tissue-, dose-, time-stratified and cross-dose, cross-tissue meta DEG signatures that consider global gene expression is unique over existing toxicogenomics resources, whose reliance on the literature and crowdsourcing introduces potential biases and limitations associated with selective gene reporting. The integration of ToxiOmics with our other multiomics tools enables researchers to leverage network biology to more thoroughly understand and compare the effects of toxicants to each other, and identify potential synergistic and antagonistic effects of the combination of therapeutic drug treatments with high environmental toxicological exposure burdens that may warrant future experimental investigation. The overlap-based approaches to exposure comparisons also invite users to query gene sets of interest to identify relevant associated toxicological exposures, thereby guiding future comparative mechanistic study. ToxiOmics represents a significant advance in the application of transcriptomics data-driven approaches to studying toxicology with broad applications to fully leverage this data type.

Despite the advances made by ToxiOmics, we must acknowledge some potential limitations. The reliance on publicly available GEO datasets introduces potential biases toward

certain species, tissues, and platforms, as demonstrated by the higher prevalences of human, liver, and microarray datasets. While we determined meta DEG and pathway signatures in an attempt to mitigate the issue, differences in dose, duration, and experimental conditions will naturally lead to variability in DEG signatures which remains, especially in relatively poorly represented chemicals, species, and tissues. Further, although we employ cross-species ortholog mapping to enhance cross-species comparison, interspecies differences in toxicodynamics and toxicokinetics complicate the direct extrapolation of ToxiOmics data to human health outcomes²²². Finally, ToxiOmics currently only incorporates transcriptomics data, leaving open questions about toxicological perturbations at other omics layers including the metabolome, proteome, and epigenome. In light of these challenges, future versions of ToxiOmics will expand data and chemical coverage to include more relevant RNAseq studies conducted since the explosion in popularity of the method, and further integrate with our Mergeomics and PharmOmics tools while also expanding into other omics layers to enable further understanding of the multiomic effects of toxicological exposures. Additionally, due to the reliance of studies of single-chemical exposures, ToxiOmics will also be expanded to account for exposures to toxicant mixtures to better reflect the real conditions in which human populations are exposed to toxicants.

Overall, ToxiOmics represents a novel toxicogenomics resource enabling the systematic exploration and comparison of the transcriptomic responses to toxicants. Integration with our Mergeomics and PharmOmics tools enables powerful analytical capabilities to glean insight into the GRN perturbations of toxicant exposures, compare toxicants to each other and to pharmaceutical drugs, and detect toxicants enriched for various biological pathway gene sets. By making ToxiOmics publicly available, we aim to catalyze further mechanistic toxicology research, support risk assessment, and foster translational applications to identify human health outcomes

associated with toxicological exposures. Ultimately, ToxiOmics represents a step toward more integrative, systems biology data-driven approaches to toxicogenomics research which will be fully realized through iterative expansion and integration across omics layers.

3.5 Figures

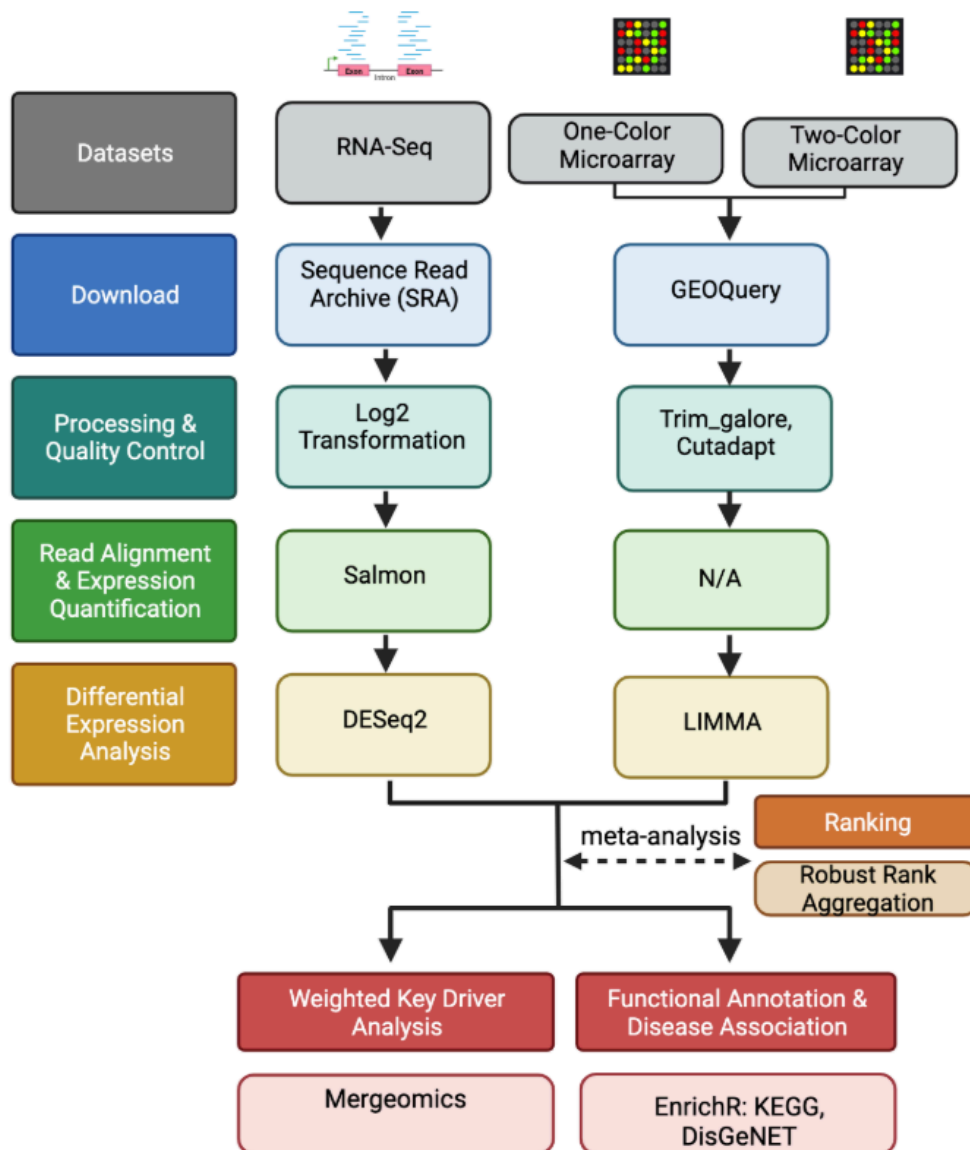


Figure 3.1. ToxiOmics analytical workflow. ToxiOmics signatures are derived from either cDNA microarray or RNAseq experiments. Microarray datasets are analyzed using our limma pipeline while RNASeq datasets are analyzed using our Salmon-DESeq2 pipeline to detect DEGs, which are used for downstream pathway enrichment and network based analysis, and included for use in the other comparative analyses enabled by the ToxiOmics web server.

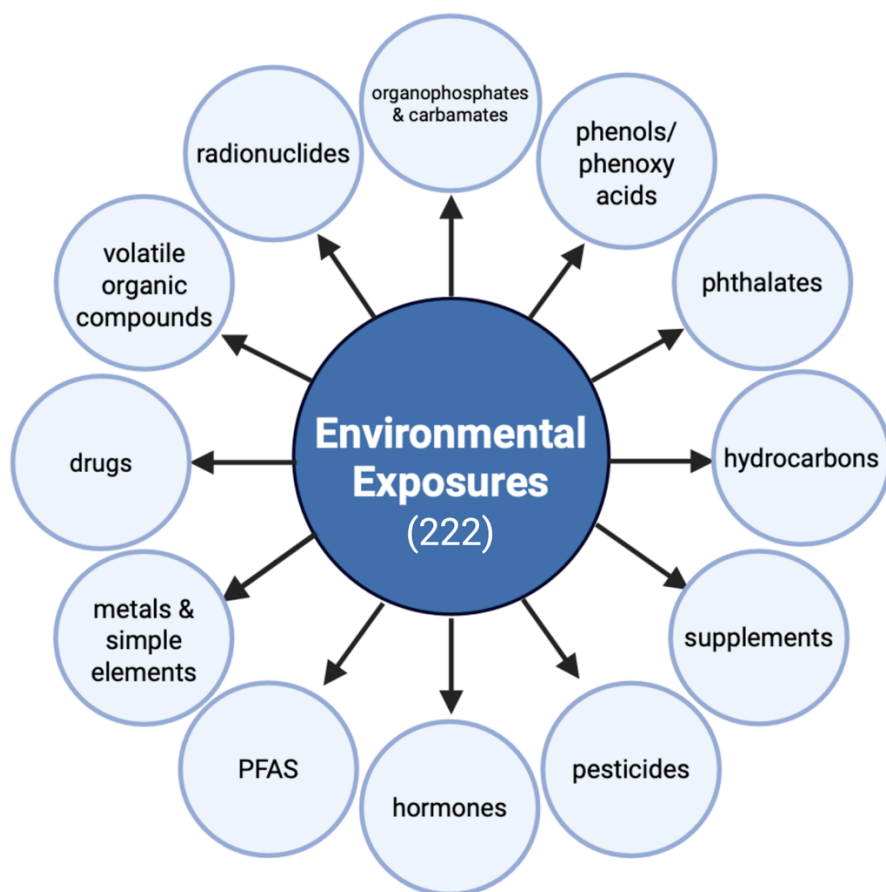


Figure 3.2. Chemical classes included in ToxiOmics. This is an overview of the important toxicologically relevant classes of chemical compounds contained in the ToxiOmics database.

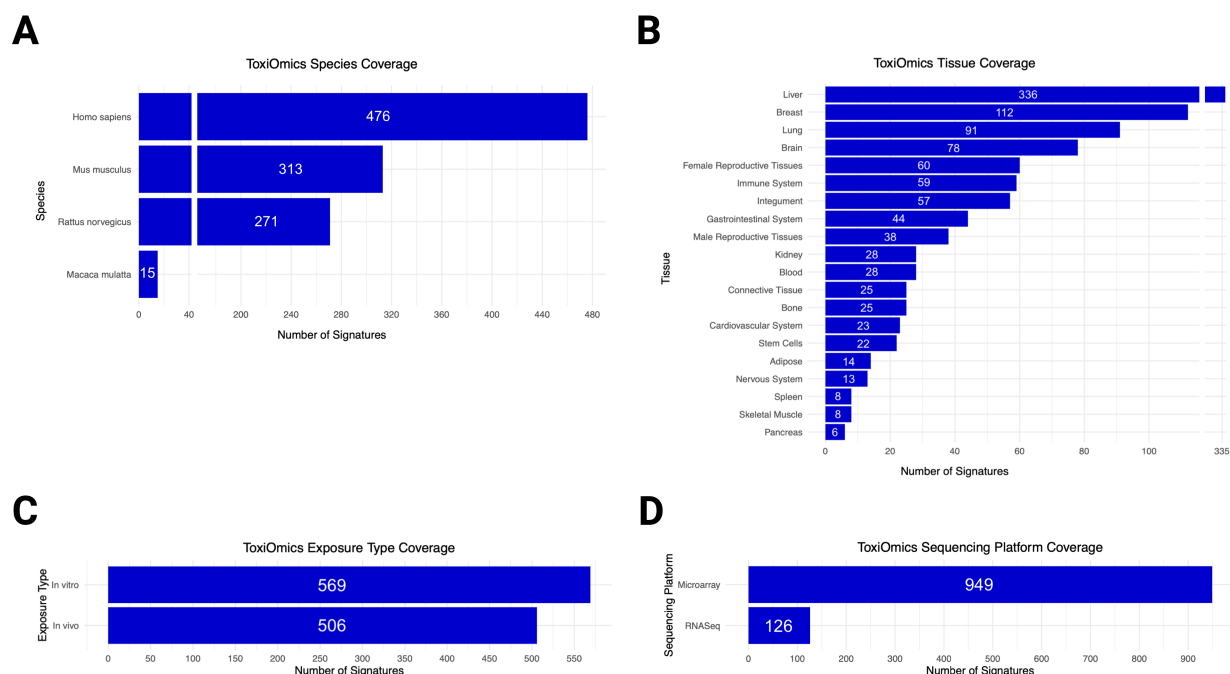


Figure 3.3. ToxiOmics signature coverage summaries. ToxiOmics contains 1,075 stratified DEG signatures derived from 304 studies posted to GEO of 222 toxicologically relevant compounds across 4 species and 29 tissues. **A)** The number of signatures associated with each species included in ToxiOmics. **B)** The number of signatures associated with each tissue type included in ToxiOmics. **C)** The number of signatures associated derived from either *in vitro* or *in vivo* exposure types in ToxiOmics. **D)** The number of signatures derived from either cDNA microarray or RNAseq experiments in ToxiOmics.

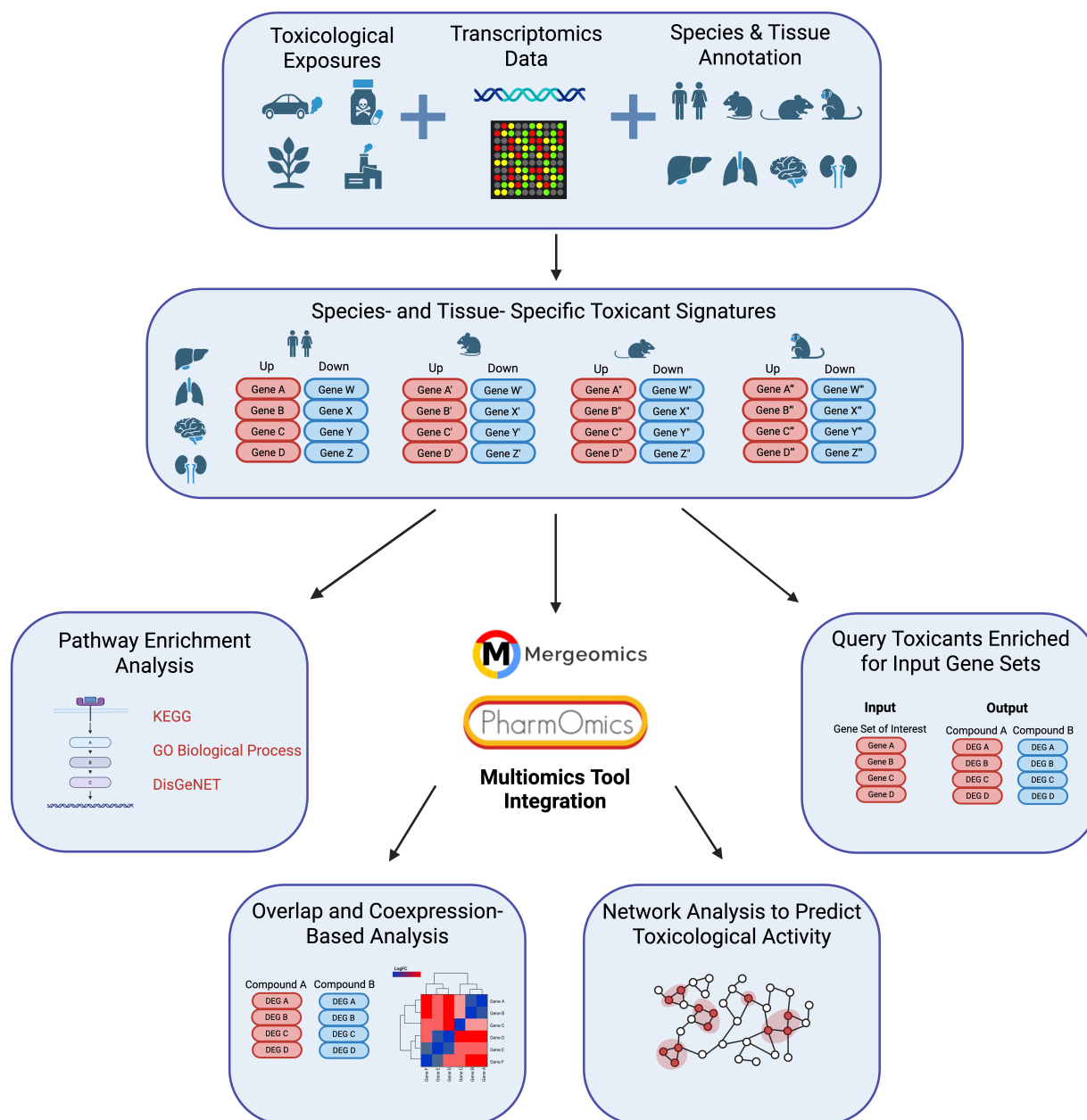


Figure 3.4. An overview of the data contained within ToxiOmics and the functional applications of the web tool. ToxiOmics derives species-, tissue-, dose-, and time-stratified DEG signatures from publicly available transcriptomics data associated with toxicologically relevant exposures. ToxiOmics also derives species- and tissue-specific meta DEG signatures to capture the robust transcriptomic effects across doses and lengths of exposure. These signatures are used to determine biologically enriched pathways and compare toxicants to each other using overlap-, coexpression-, and network-based approaches. Users can also supply gene sets of interest to determine toxicologically relevant compounds in which the gene sets are enriched. Integration with Mergeomics and PharmOmics enables the visualization of key driver networks and the prediction of potential interactions between toxicants and pharmaceutical drugs.

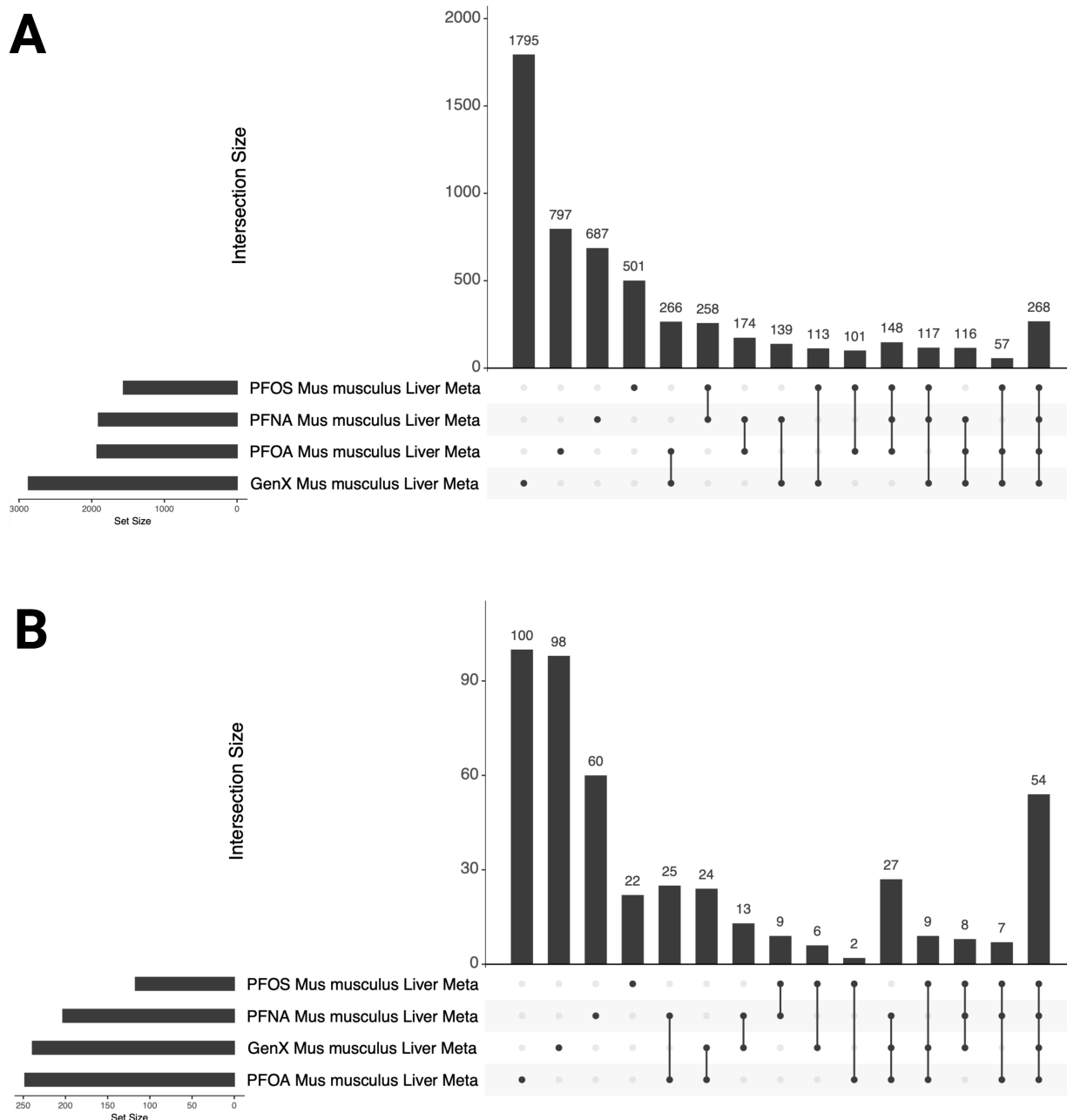


Figure 3.5. Comparison of murine liver meta DEG and pathway signatures for legacy PFAS compounds and GenX. The murine liver meta DEG and pathway signatures for the legacy PFAS compounds PFOA, PFNA, and PFOS, and the emerging replacement shorter-chain PFAS compound GenX were pulled from ToxiOmics. **A)** Comparison of the meta DEG signature sizes reveals that GenX has the most unique DEGs, followed by PFOA. GenX and PFOA share the highest number of DEGs out of any pair of compounds, and several hundred DEGs are observed in all four PFAS compounds, indicating relatively similar transcriptomic effects. **B)** Comparison of the meta pathway sizes reveals that both GenX and PFOA have high numbers of uniquely perturbed pathways in the murine liver. More than 50 pathways are perturbed in all four PFAS compounds, again highlighting relatively similar transcriptomic effects.

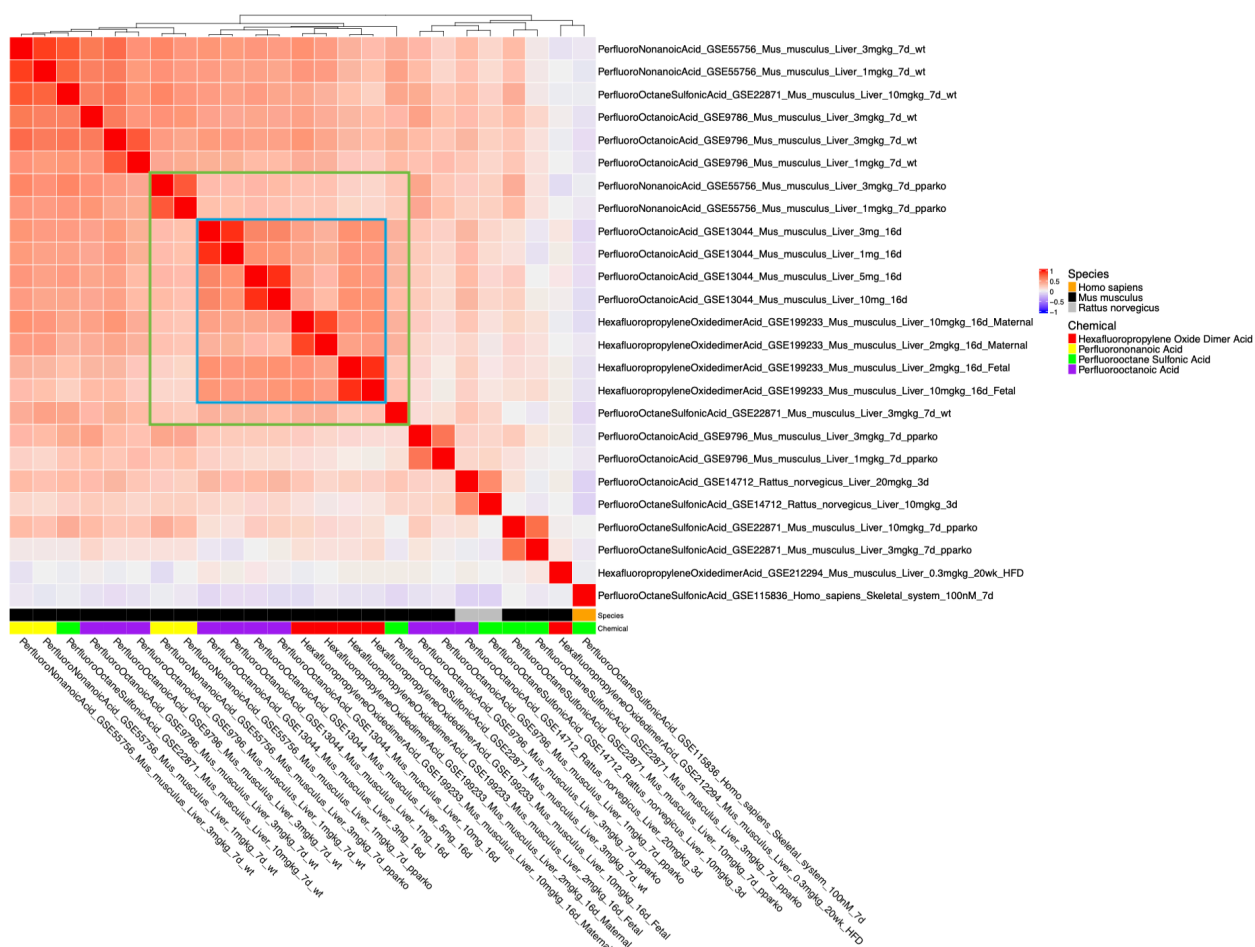


Figure 3.6. Clustering analysis reveals transcriptomic similarities between legacy PFAS compounds Perfluorooctanoic acid (PFOA), Perfluorooctane sulfonic acid (PFOS), Perfluorononanoic acid (PFNA), and Hexafluoropropylene oxide dimer acid (HFPO-DA, “GenX”) across doses, durations of exposure, and developmental stages. This heatmap shows the correlation in global gene expression patterns across stratified signatures associated with each PFAS compound cataloged in ToxiOmics. Spearman’s rank correlation coefficients were used to assess the correlation between signatures, and hierarchical clustering was performed to order the signatures based on how similar they are to each other. Redder hues indicate signatures in which the magnitude and direction of the changes in the expression of all genes are more similar, while bluer hues indicate dissimilarity. The tight central cluster, surrounded by a blue box, contains PFOA and GenX stratified signatures, indicating a high degree of similarity between these two compounds relative to the rest of the legacy PFAS. This central cluster also includes GenX signatures derived from different developmental stages of mice, indicating potentially preserved transcriptomic responses to GenX exposure across life stages. A slight extension of this cluster, indicated with the green box, includes the other legacy PFAS compounds PFOS and PFNA. This close ordering after hierarchical clustering, combined with their red hued heatmap cells, indicate transcriptomic similarity between these compounds, PFOA, and GenX, albeit to a lesser degree than PFOA and GenX have to each other.

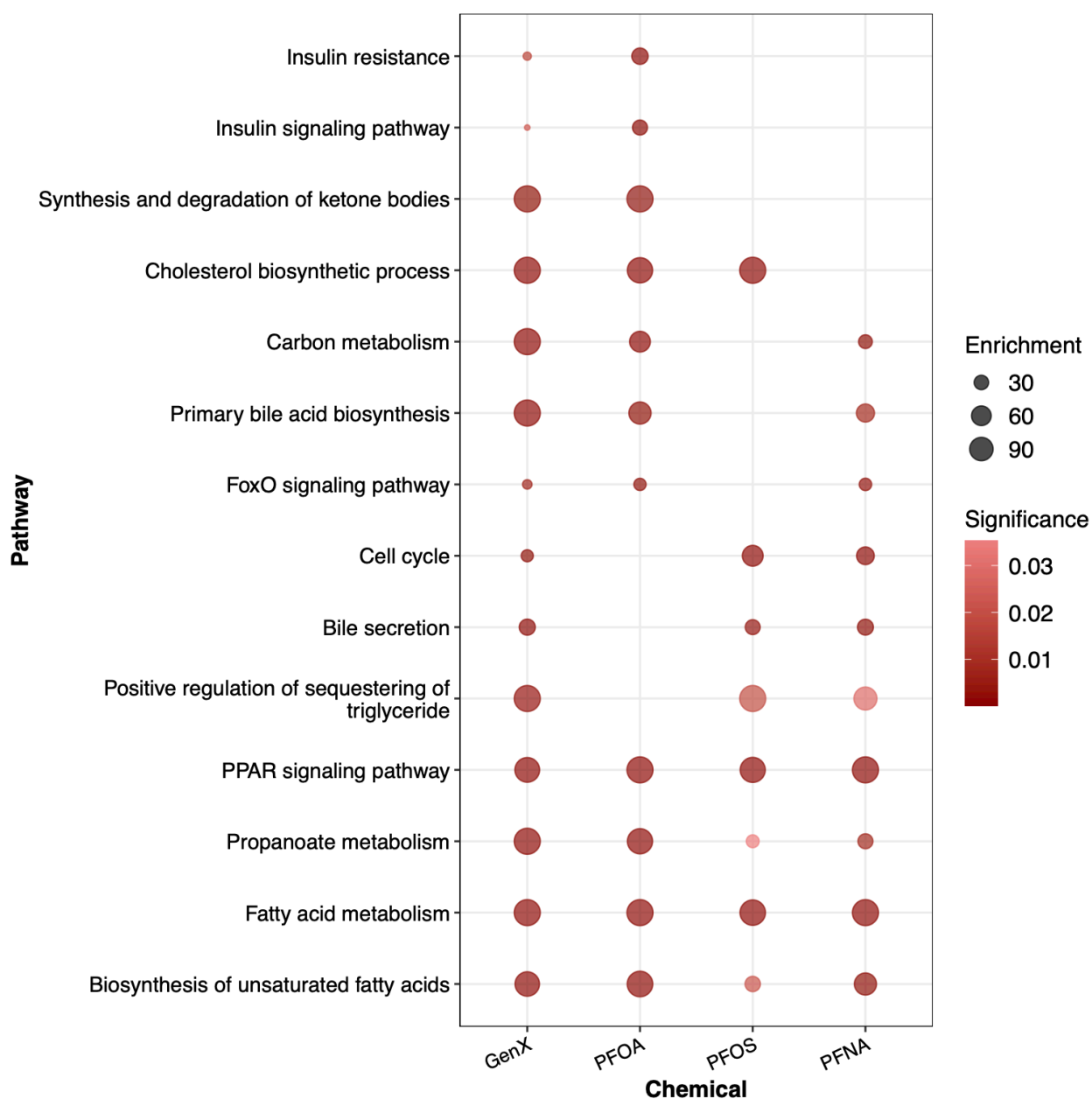


Figure 3.7. Pathway enrichment analysis reveals shared biological pathway perturbations elicited by the transcriptomic response to both legacy PFAS and GenX exposure in the murine liver. The murine liver meta DEG signatures of each PFAS compound were used for pathway enrichment analysis of KEGG and GO pathways with EnrichR. Pathways with a lower FDR are shown in darker red hues, and the size of the circle corresponds to the significance of EnrichR's Enrichment Score metric. Significant pathways shared between legacy PFAS compounds and GenX include PPAR signaling, lipid metabolism processes, pathways related to insulin resistance, and signaling and cell cycle pathways related to cancer.

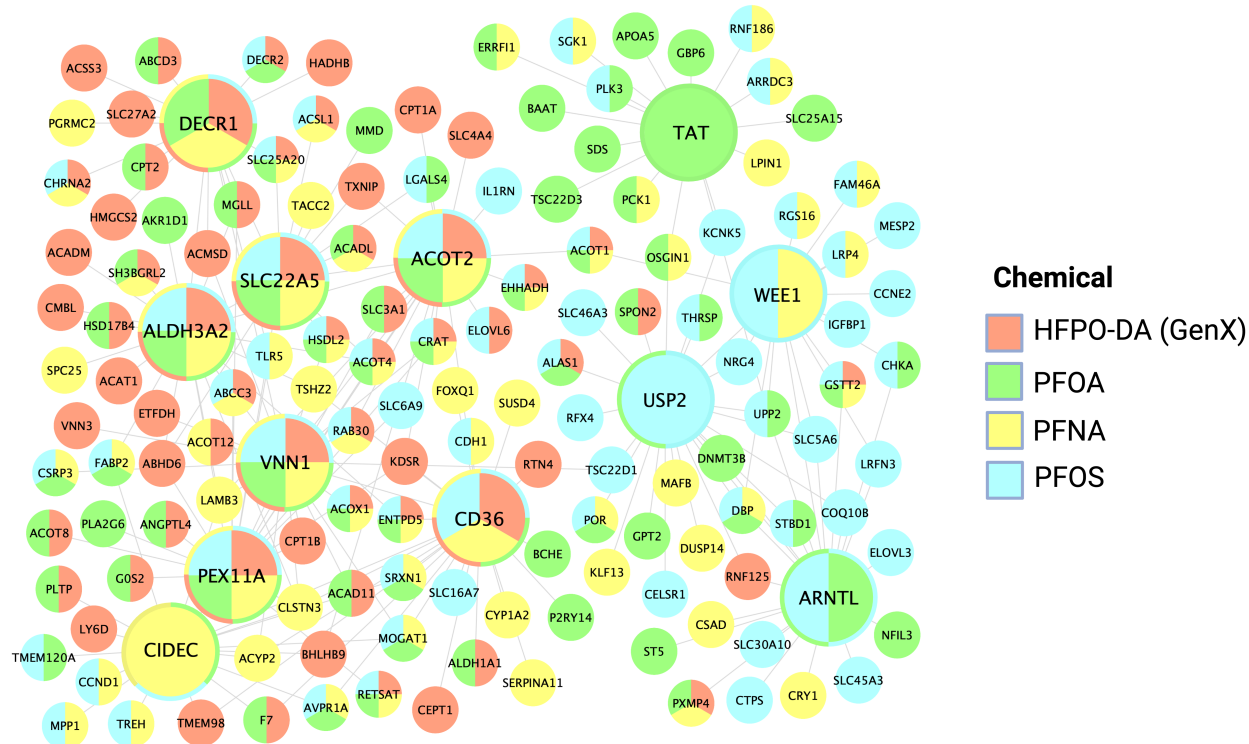


Figure 3.8. Key driver analysis reveals shared key drivers of the transcriptomic response to both legacy PFAS and GenX exposure in the murine liver. The top 300 murine liver meta DEGs by FDR were used as inputs for KDA via Mergeomics integration in ToxiOmics. Network nodes are colored according to the chemical compounds for which they were detected as murine liver meta DEGs. Larger nodes are key drivers, and their border color indicates the chemical compounds for which they are key drivers. Of the 12 detected key drivers, 7 are key drivers in the transcriptomic response to perturbation by each of the 4 PFAS compounds, suggesting a high degree of similarity in the transcriptomic response between the legacy PFAS compounds and GenX.

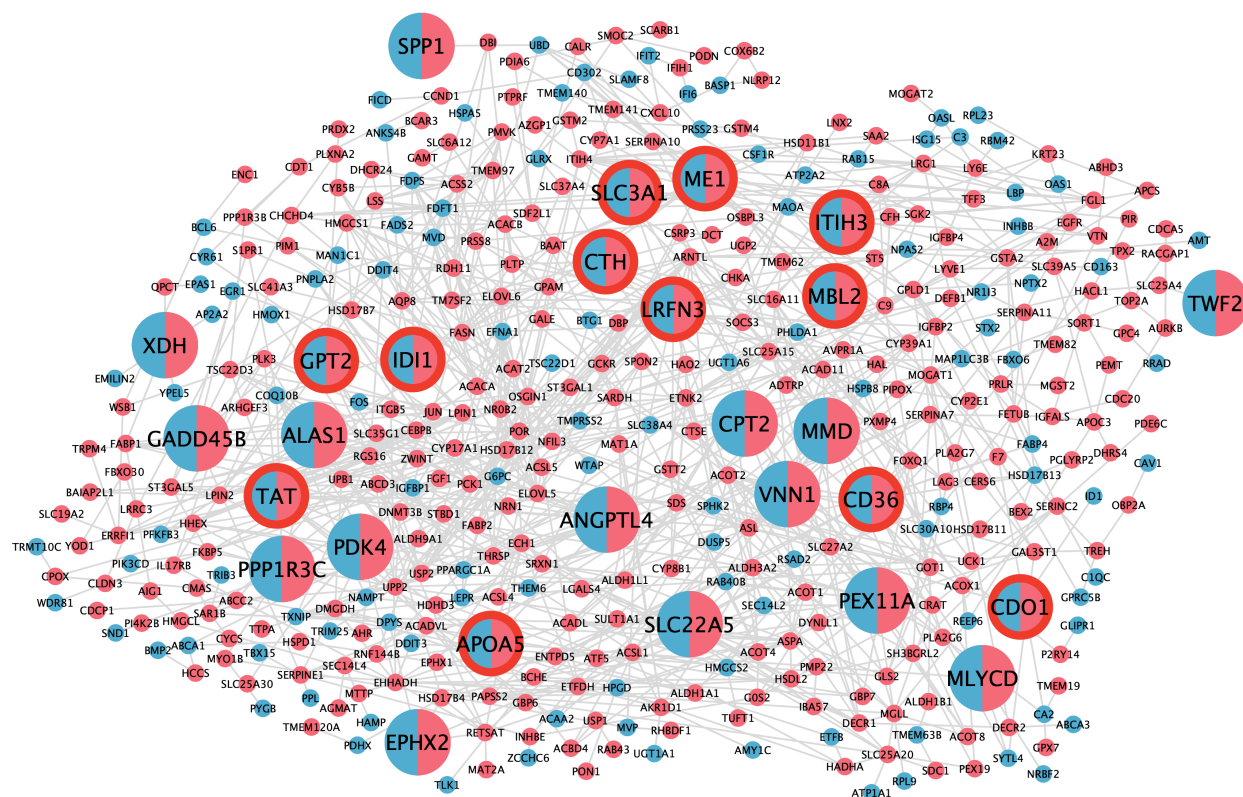


Figure 3.9. Network-based comparative analysis highlights potential antagonistic effects between PFOA exposure and pioglitazone treatment. Through integration with PharmOmics, both the murine liver meta DEG signature and the pioglitazone human meta DEG signature were mapped to the Bayesian human liver GRN. The murine liver meta DEG signature was first converted to human ortholog gene symbols prior to the analysis. Using the network distance metric outlined in PharmOmics, the signatures were compared to identify subnetworks perturbed by each signature and examine their degree of overlap. Shared network member genes are enlarged, and genes are colored according to the signature in which they are meta DEGs. The labels of genes that are expressed in the same direction are colored in grey, while those of genes expressed in opposite directions are colored in black. The labels of non-shared genes are colored in grey to emphasize the shared genes that are expressed in different directions. The difference in direction of expression in the meta signature indicates that the transcriptomic responses in these subnetworks are potentially antagonistic between PFOA and pioglitazone exposure/treatment, prioritizing these genes for further study to examine whether a true antagonistic effect is elicited.

Chapter 4. Computational Integration of *In Vivo* Single Cell and *In Vitro* Bulk Transcriptomics Across 221 Human and Mouse Datasets Differentiates Physiological versus Non-Physiological Hepatic Cell Lines for Hepatotoxicity Screening

4.1 Introduction

In recent years, the field of toxicology has pursued the implementation of novel alternative method (NAM) approaches to increase the sustainability and ethicality of biomedical research^{4,223,224}. One area of intense NAM focus has been the development of *in vitro* experimental paradigms as alternatives to traditional animal testing^{225,226}. However, the selection of suitable *in vitro* models for toxicity testing is a challenge with many important considerations.

In vitro cell culture utilizes either primary cells or immortalized cell lines. Primary cells, which come directly from donor tissues, are known for their ability to maintain *in vivo* functionality within a limited timeframe (e.g., gluconeogenesis within 48 hours for primary hepatocytes) and are a great proxy for studying *in vivo* processes *in vitro*^{227–229}. However, primary cells are short-lived, highly variable between donors, and not a renewable resource, making it difficult to replicate experiments and to test the cells under multiple or prolonged conditions. One solution to this challenge is immortalized cell lines, which have been modified to proliferate indefinitely and can be maintained for a long period of time²³⁰. However, the immortalization processes, such as genetic modification or the introduction of external agents to growth media, are known to induce functional modification that differentiates immortalized cell lines from *in vivo* tissue²³¹. Therefore, the higher replicability derived from cell lines must be balanced against the functional deficiencies induced through immortalization.

To select immortalized cell lines for *in vitro* studies, common considerations include the presence of key marker genes or genetic abnormalities of interest, the developmental stage of the

tissue from which the immortalized cell lines were established, the up- or down-regulation of specific biological processes of interest, and compatibility with experimental procedures or assays²⁴⁻²⁷. In addition to these criteria, it is also important to consider the baseline physiological and toxicological relevance of the cell lines. Incorporating cell lines which are both maximally compatible with experimental conditions and optimally reflective of *in vivo* physiology and toxicological responses will lead to more impactful and translatable *in vitro* toxicity studies.

The transcriptome is an informative gauge of cellular function. Changes in gene expression patterns have long been used to elucidate the molecular and functional effects of chemical compounds in exposure studies, as differentially expressed genes (DEGs) may be used to identify important cellular and gene regulatory processes altered by exposure to toxicants⁷. As proof-of-concept, transcriptome-based drug exposure signatures have proven to be effective and accurate to guide drug repositioning and drug toxicity studies^{23,232,233}. Resemblance of *in vitro* models to *in vivo* cells at the level of the transcriptome helps ensure both functional relevance at baseline and extrapolation of differential expression results from *in vitro* studies to the *in vivo* context. Therefore, herein we propose and develop a novel, transcriptomics-driven approach to *in vitro* model selection which seeks to identify immortalized cell lines which most closely resemble the basal transcriptome patterns of their corresponding *in vivo* cell types. Due to the liver being both the driver of chemical clearance in the body and a target of many toxic agents, we use hepatic cell types to demonstrate this approach. The utilization of this computational approach to identify an *in vitro* model for the liver and extrapolation to other tissues will guide the future development of accurate, efficient, human-relevant, and cost-effective *in vitro* models for toxicology studies and further the pursuit of NAM approaches to toxicological and general biomedical research with improved physiological relevance and translational value.

4.2 Methods

Data curation

A total of 214 transcriptomic datasets of hepatic cell lines and primary cells were queried and curated from the Gene Expression Omnibus (GEO). We found datasets involving the following immortalized cell lines: human hepatocytes (HepaRG, HepG2, Hep3B, Huh7), human hepatic stellate cells (LX2, TWNT4), human cholangiocytes (HuCCT1, EGI1, KKU213, TFK1, KKU100, NHC), mouse hepatocytes (AML12, Hepa1-6, Hepa1c1c7), and mouse Kupffer cells (KUP5). Datasets from primary human and mouse hepatocytes, hepatic stellate cells, and Kupffer cells were also considered. For each cell line and primary cell type, the total numbers and technical platforms (RNA sequencing [RNAseq] or cDNA microarray) of datasets curated are shown in Tables S.4.1 (summary information detailing cell type, species, and platform) and S.4.2 (full GEO accession numbers, platform, sample size, etc.). Datasets were checked to select the control group (e.g., exposure to an empty vehicle or inert control agent) to reflect baseline physiology. As reference *in vivo* cell type gene expression data, 7 single cell RNA sequencing (scRNAseq) datasets from human (4 studies) and mouse (3 studies) liver tissue were curated from published liver studies^{234–240} and the Single Cell Portal²⁴¹. A summary of the scRNAseq data sources is shown in Table S.4.3. In total, 221 bulk and scRNAseq studies were included in our analysis.

Data download and pre-processing

Microarray data that had undergone pre-processing and quality control were downloaded from GEO using the GEOquery package⁴⁸. We verified that microarray data were normalized and log₂-transformed prior to downstream analysis. Raw RNAseq datasets were downloaded from NCBI's Sequence Read Archive (SRA)¹³², quality checked, and processed in the Anaconda

environment¹³⁴. FASTQ files were retrieved from SRA using the parallel-fastq-dump wrapper¹³³. We performed quality control on the FASTQ files using Trim Galore (v 0.6.6) to trim the 3' end of reads with low quality base calls⁵⁰ and Cutadapt (v 2.1.0) to remove adapter sequences⁵¹, followed by the removal of short sequences (<20 bp).

Transcript mapping for RNAseq studies

Control samples from each RNAseq dataset were mapped to the corresponding species-specific reference genome using Salmon (v 1.10.0)⁵²; *Homo sapiens* GRCh38.p13 genome assembly¹³⁵ was used for human cell lines and the *Mus musculus* GRCm39 genome assembly¹³⁶ was used for mouse cell lines. Reference species genomes were built using the “index” tool within the Salmon package. Tximport (v 1.22.0) was employed to import and summarize the gene expression quantification results in R prior to downstream analysis⁵³.

scRNAseq data processing

For each reference scRNAseq dataset from human and mouse livers, raw count data from control samples (e.g. healthy adult human or wild type unexposed mouse) were normalized using the NormalizeData function in Seurat (v 5.1.0)²⁴² unless the dataset was otherwise specified to have already been normalized. Due to the sparsity of scRNAseq data, read counts of zero in these single cell expression matrices were replaced with NA before read counts were averaged on a per-gene basis. The resulting average counts per gene across all cells (i.e., pseudobulk) from each hepatic cell type within each scRNAseq dataset source were used for comparison with the microarray and bulk RNAseq studies of the cell lines.

Data harmonization across platforms

As datasets from three technical platforms (bulk RNAseq, microarray, scRNAseq) were used in our analysis, data comparability was an important issue. To alleviate the issue, we chose to use rank-based quantile normalization for each dataset to harmonize data across platforms, as gene expression ranks are less dependent on the quantification range and technical variability of platforms and rank-based quantile normalization ensures all datasets follow the same distribution. Quantile normalization across all microarray, bulk RNAseq, and scRNAseq datasets was performed using the `normalize.quantiles` function from the `preprocessCore` R package²⁴³. Accordingly, rank-based statistics was used in downstream analysis.

Assessment of global gene expression similarity across datasets for each cell type

To assess similarity and differences across datasets and cell lines within each cell type, quantile-normalized gene expression levels for each dataset (i.e., the mean expression values for individual genes across samples within each dataset) were compiled into a dataframe in R. We then used two different approaches to assess similarity/difference between datasets. First, we calculated non-parametric Spearman's rank correlation coefficient between pairs of transcriptomic datasets. These correlation coefficients were used to make heatmaps using the R package `ComplexHeatmap`¹⁴⁴, where datasets with similar transcriptomic patterns are clustered closer to each other. Second, we performed nonlinear dimension reduction using uniform manifold approximation and projection (UMAP)²⁴⁴ on the following genes: 1) the ~2500 most highly variable genes across all datasets and 2) genes classified as cell type marker genes curated from PanglaoDB²⁴⁵. Missing values were imputed using the `missForest` R package²⁴⁶ before carrying out UMAP. Based on the clustering

results from both approaches, we determined the cell lines with the highest similarity in gene expression patterns to the *in vivo* human and mouse cells from scRNAseq for each cell type.

Differential gene expression and pathway analysis to obtain cell line signatures

Differentially expressed genes (DEGs) between each cell line and their corresponding *in vivo* cell type (e.g. HepaRG vs *in vivo* human hepatocytes, LX2 vs *in vivo* human hepatic stellate cells, etc.) were determined using the Wilcoxon Rank Sum test. To derive the physiological gene signatures for each cell type, we compared the datasets sharing similar expression patterns to the *in vivo* human and mouse cells with those from other dissimilar cell lines. Genes with Benjamini-Hochberg corrected p-value < 0.05, representing a false discovery rate (FDR) of <5% were considered DEGs.

Following DEG determination, pathway enrichment analysis was performed using EnrichR⁵⁸, using pathways from the KEGG²⁴⁷, GO Biological Process¹⁴¹, and DisGeNET²⁴⁸ databases. Pathways were considered significantly enriched if they met the FDR < 0.05 threshold and had at least 5 genes overlapping with the input DEG lists. These DEGs and pathways reflect main biological functions that differ between cell lines and physiological states.

Network analysis of DEGs to identify key drivers and network visualization

To understand the regulatory relations among DEGs and identify potential regulators, DEGs from each cell line were used to perform key driver analysis (KDA) using Mergeomics²¹ and cell-type specific gene regulatory networks (GRNs). KDA detects key drivers (KDs), genes which are connected to significantly more DEGs compared to the expected number for a randomly selected network gene within a given GRN using a Chi-square-like statistic²¹. KDs are potential network

regulators of the DEGs. Cell type-specific GRNs were constructed from human and mouse liver scRNAseq data using SCING²⁴⁹. For human cell line DEGs, human cell type SCING networks were used for KDA analysis, while mouse liver GRNs and mouse cell type SCING networks were used for KDA analysis for mouse cell line DEGs. Cytoscape⁶⁰ was used to visualize the network of the top KDs and DEGs in their network neighborhoods.

Hepatotoxicity application study: assessment of tetrachlorodibenzodioxin (TCDD) transcriptomic response in hepatocytes

To evaluate whether the transcriptome-based cell line selection agrees with hepatotoxicity findings, we curated three datasets covering TCDD exposure in HepaRG, HepG2, and *in vivo* hepatocytes. The datasets were GSE122518²⁵⁰ (microarray) for HepaRG, GSE293127²⁵¹ (RNASeq) for HepG2, and a time-course scRNAseq study of TCDD exposure in mice²⁵² curated from the Single Cell Portal. For the HepaRG and HepG2 datasets, gene expression data were downloaded and DEGs were determined using limma for the HepaRG microarray and DESeq2 for the HepG2 RNAseq. The *in vivo* DEG signature of TCDD from the scRNAseq study of TCDD exposure in mice²⁵² were previously determined by the study authors using DESeq2 and were downloaded from the Single Cell Portal. Subsequent pathway analysis on all sets of DEGs was performed using EnrichR as described above. We then compared the DEGs and pathways across datasets to assess similarity and difference between cell lines and *in vivo* hepatocytes in TCDD-induced gene signatures.

4.3 Results

UMAP and correlation heatmap analyses reveal gene expression similarity/dissimilarity between *in vivo* liver cell types, cell lines, and primary cells for each hepatic cell type

We first aimed to identify optimal *in vitro* models for each hepatic cell type by examining the overall similarity/dissimilarity in basal gene expression patterns between *in vivo* liver cells, *in vitro* cell lines, and primary cells. Cell lines that closely mimic the basal gene expression patterns of their corresponding *in vivo* cell types are expected to better capture physiological states.

For hepatocytes, UMAP clustering of the 4 human and 3 murine hepatocyte cell lines, human and murine primary hepatocytes, and human and murine *in vivo* hepatocyte data derived from a total of 168 studies (see Tables S.4.1 and S.4.3 for studies for each cell line) revealed segregation of the datasets into multiple clusters (Fig. 4.1A). The UMAP plot showed five discernable clusters for the Huh7 cell line (lower left quadrant), a sparse mixture of Hep3B and HepG2 datasets (lower center of the plot), HepaRG and the physiological cells containing *in vivo* hepatocytes and primary cells (center of the plot highlighted by a dotted circle), the murine cell lines (lower right quadrant), and a distinct subcluster of HepG2 datasets (upper right quadrant). In general, different datasets from each hepatocyte cell line tended to cluster together despite differences in study source, lab and experimental settings, media, and control vector used, indicating robust cell line-specific basal gene expression patterns that are less influenced by batch effects and environmental influence. However, there were 2 cell lines, Hep3B and Huh7, that were spread more widely across the UMAP, suggesting that the gene expression patterns of these cell lines are less stable and more variable between datasets. While environmental factors related to the experimental conditions of different studies may be a contributing factor to this phenomenon, the unique disparity among datasets corresponding to these cell lines rather than blanket instability

across all cell lines examined was notable. We also observed a high degree of similarity between the HepaRG cell line and the *in vivo* human hepatocytes, which also clustered with primary human hepatocytes (Fig. 4.1A, circled in red), indicating that HepaRG has a high degree of similarity in basal transcriptome to physiological hepatocytes.

Out of this total clustering of all hepatocyte studies used in this analysis, the murine cell lines clustered together distinctly from the human cell lines in the lower right quadrant (Fig. 4.1A) but notably were more distant from their corresponding *in vivo* mouse hepatocyte reference data than the HepaRG data was from the human hepatocyte reference data. This suggests that the murine hepatocyte cell lines are more transcriptomically distinct from their corresponding *in vivo* murine hepatocytes than the HepaRG cell line is from *in vivo* human hepatocytes. Spearman correlation-based heatmap analysis also supports the UMAP results (Fig. S.4.3), as stronger correlations in gene expression patterns were observed between HepaRG and *in vivo* hepatocytes compared to other cell lines. Therefore, our clustering results using two methods support that the HepaRG cell line shares the most similar gene expression pattern with the *in vivo* and primary hepatocytes compared to other commonly used cell lines, forming a transcriptomic basis for utilizing this cell line as an optimal model for human hepatocytes *in vitro*.

The same clustering analysis was also performed for hepatic stellate cells (HSCs) and cholangiocytes, for which sufficient datasets were available for meta-analysis. UMAP clustering of 2 human HSC cell lines (LX2 and TWNT4), human and mouse primary HSCs, and *in vivo* human and mouse HSCs from a total of 27 studies (see Tables S.4.1 and S.4.3 for details) revealed a lack of distinct clusters in the plot (Fig. 4.1B), reflecting weaker transcriptomic distinction between cell lines for HSCs than there were for hepatocytes (Fig. 4.1B). As such, a clear determination of a “better” cell line could not be definitively made based solely on the clustering

of their basal transcriptome profiles. Two of the LX2 cell line datasets clustered closely with the human and mouse *in vivo* HSCs and the primary mouse HSCs (Fig. 4.1B, circled in red), but this pattern was not observed for the entire set of LX2 datasets. Additionally, TWNT4 had just 2 datasets to consider versus the 11 available for LX2, and the addition of more data, when available, could alter these interpretations.

Similar to the HSC cell lines, UMAP clustering of human cholangiocyte cell lines and human and mouse *in vivo* cholangiocytes from a total of 26 studies (see Tables S.4.1 and S.4.3) revealed a lack of distinct clusters (Fig. 4.1C). Pearson correlation heatmaps for both HSCs and cholangiocytes (Fig. S.4.4 and S.4.5) show a high degree of positive correlation across the datasets for each cell type, and the hierarchical clustering revealed a lack of distinct clusters of cell lines. Therefore, both the UMAP and Pearson correlation and hierarchical clustering analyses reveal a lack of distinction between *in vitro* cell lines and *in vivo* cells in the non-parenchymal cell types, in contrast to what was observed for the hepatocyte cell lines. While these results may suggest that the *in vitro* cell lines for HSCs and cholangiocytes are reasonably similar to *in vivo* cells, the lack of distinct patterns could also be the result of markedly less data for these non-parenchymal cell types which limited the power to derive definitive conclusions.

Experimental condition differences do not explain the distinct HepG2 subclusters

We noted that the most widely used HepG2 cell line split into two distinct subclusters based on basal gene expression values, neither of which was closely clustered with *in vivo* hepatocyte references (Fig. 4.1A). Thus, we investigated whether any known confounding factors found in the metadata of the contributing studies might explain the emergence of the two HepG2 subclusters (Fig. 4.2). We found that studies using Minimum Essential Medium (MEM) cluster together in

the top sub-cluster, however many studies using Dulbecco's Modified Essential Medium (DMEM) also cluster among them in addition to comprising the lower sub-cluster (Fig. 4.2A). The choice of control experimental condition (e.g., addition of control vehicle, scrambled siRNA, etc.) also did not lead to distinct patterns (Fig. 4.2B). While there may be an effect introduced by the profiling platform (Fig. 4.2C), there were only two studies that utilized microarrays and the RNASeq datasets also split into two subclusters; therefore, this effect is not definitive. Stock sources also did not explain this difference, as nearly all studies used HepG2 cells obtained from the American Type Culture Collection. Overall, confounding factors readily identifiable from the metadata of the contributing studies fail to explain the emergence of the distinct HepG2 subclusters derived from their basal gene expression patterns. Other unreported factors, such as passage number, could contribute to the sub-clustering. Our finding on the two distinct clusters across studies of the most widely used hepatic cell line HepG2 and the dissimilarity of HepG2 to *in vivo* hepatocytes, suggests that its function has diverged within the cell line and away from the *in vivo* hepatocyte functions.

DEG analysis reveals transcriptomic similarities and differences between cell lines and *in vivo* cell types

To further assess the similarity or difference between each cell line and its corresponding *in vivo* cell type, we utilized the non-parametric Wilcoxon Rank Sum test to determine significant DEGs between the two groups - more DEGs indicate stronger differences and vice versa. A summary of the number of DEGs detected for each cell line against its corresponding species-matched *in vivo* cell type is shown in Table 4.1. Based on DEG numbers, among human hepatocyte cell lines, Hep3B and HepRG had fewer DEGs and HepG2 and Huh7 had more DEGs compared to *in vivo*

hepatocytes, generally agreeing with the UMAP and correlation clustering results. However, despite the strong clustering results indicating similar basal gene expression patterns in HepaRG and *in vivo* hepatocyte data, HepaRG had the second fewest DEGs (861 genes) and Hep3B had the fewest DEGs (700 genes). The seeming discrepancy from the clustering results can be explained by the fact that the UMAP and clustering analyses were based on the set of the most variable genes and hepatocyte marker genes which are not necessarily DEGs. The murine cell lines' higher numbers of DEGs are consistent with the observed clustering pattern in which they appeared farther away from the murine reference data than HepaRG did from the human reference data (Fig. 4.1A).

Analysis of HSC (LX2, TWNT4) and cholangiocyte (HuCCT1, EGI1, NHC, TFK1, KKU100, KKU213) cell lines against their respective *in vivo* cell types revealed no significant DEGs. This agrees with the lack of distinct clusters in the UMAP and correlation analysis above. Alternatively, this may be due to the relative lack of statistical power (1-12 datasets each versus 26-37 datasets for the human hepatocyte cell lines) for these cell types as they are not studied to the same degree that hepatocyte cell lines are, and the higher degree of within-cell-line variance also reduced the power to detect significant DEGs.

We then sought to highlight the genes and pathways in both HepaRG and the *in vivo* hepatocyte data that most closely recapitulate physiological liver function and that differ between the physiological data and non-HepaRG cell line data. We compared human and mouse physiological (*in vivo*, primary hepatocyte, and HepaRG cell line) data with those from the sub-optimal (non-HepaRG) cell line models to identify 7,075 DEGs. DEGs with physiological and toxicological relevance include *CYP2E1*, which plays a critical role in ethanol and drug metabolism and associated liver injury²⁵³; *SAA1* and *ITIH4*, which have roles in platelet

aggregation during the immune response²⁵⁴ and extracellular matrix stabilization²⁵⁵, and whose dysregulation contributes to the onset and progression of liver fibrosis in response to liver injury; *RDH16*, which has roles in retinol metabolism and *de novo* fatty acid synthesis and whose dysregulation contributes to the onset and progression of hepatocellular carcinoma²⁵⁶. The altered expression of these genes in the non-HepaRG cell lines suggest that these cell lines deviate from the physiological metabolic and immune functions of hepatocytes.

Pathway analysis to assess functional differences in immortalized cell lines

The DEGs detected for each cell line were assessed for pathway enrichment using EnrichR to further determine the functional impact of the DEGs on the cell lines that differentiates them from their corresponding *in vivo* cell types. DEGs from most of the cell lines were highly enriched for cancer, cell cycle, and viral related pathways (Table 4.1, Fig. 4.3), which is consistent with their status as immortalized cancer cells or their derivation from patients infected with hepatitis or permissiveness to hepatitis viral infection^{257,258}. These features likely make them more appropriate for cancer and virology studies. Protein ubiquitination, a key process involved in protein degradation and numerous diseases²⁵⁹, appeared as a shared enriched pathway from the DEGs of several cell lines, further highlighting functional alterations in commonly used cell lines. For HepaRG, differences in lysosomal and autophagy processes set this cell line apart from the physiological cells, potentially limiting HepaRG as a model for studying these pathways in hepatocytes. Xenobiotic metabolism pathway was enriched in DEGs of other cell lines but not in HepaRG DEGs, which is further evidence that the core toxicological functions assessed in high-throughput hepatotoxicity screening studies are more accurately reflected in HepRG. This is

substantiated by HepaRG being well known to preserve cytochromes P450s that are key genes in xenobiotic metabolism^{260–262}.

Similar to the DEG analysis, the pathway enrichment analysis for the core hepatic physiology comparison (*in vivo* and primary hepatocytes and HepaRG vs. non-HepaRG cell lines) supports that pathways critical to normal liver physiology are enriched in physiological cells but altered in the other cell lines. These include immune responses, waste clearance, various metabolic processes including fatty acid oxidation, cholesterol metabolism, and PPAR signaling. Notably, circadian regulation of gene expression was identified as a differential pathway response, as cells in culture do not maintain synchronized circadian rhythms without external “training” stimuli, while healthy intact livers do²⁶³. This is an important consideration in toxicology as the timing of toxicological exposures determines the severity of their impact based on the specific biological pathways that are activated at a given point in the circadian cycle^{263,264}.

Gene regulatory network analysis reveals subnetworks altered in hepatocyte cell lines and key regulators

To understand the regulation of the DEGs, we conducted key driver analysis using Mergeomics to detect key regulator genes (Key Drivers or KDs) that may drive the differences between physiological hepatocytes and cell lines using the DEGs from 1) physiological hepatocytes as compared to all the non-physiological cell lines and 2) each non-physiological hepatocyte cell line compared to their *in vivo* counterpart. We then visualized the corresponding subnetworks of top KDs to understand the network interactions among DEGs. We conducted these analyses using a single-cell hepatocyte GRN to assess the cell lines in the specific context of modeling hepatocytes *in vitro*. In the comparison of the physiological hepatocytes versus the non-physiological cell lines

mapped to the single cell hepatocyte GRN, 5 KDs were detected: *EEF1A1*, *DCXR*, *APOE*, *TPT1*, and *PEBP1* (Fig. 4.4). *EEF1A1* has been linked to oncogenesis and cancer progression²⁶⁵, and changes in lipid droplet size and hepatocyte lipotoxicity^{266,267}. *DCXR*, which reduces dicarbonyl compounds and thus has high relevance to carbohydrate metabolism and detoxification²⁶⁸, has been linked to poor prognosis in human hepatocellular carcinoma patients upon altered expression²⁶⁹. *APOE* is a classic regulator of hepatocyte physiology, with roles in regulating low and very low density lipoprotein secretion by the liver^{270,271}, protection against hepatitis C virus-induced liver disease²⁷², and serving as a pathological marker of non-alcoholic fatty liver disease and hepatitis B and C by complexing with *CIq* in the classical complement cascade²⁷³. *TPT1* is an oncogene, with known roles in cancer progression and migration in various tissues including the liver^{274–276}. Taken together, along with the presence of toxicologically relevant DEGs such as *PON3*, *ALDH6A1*, and *CYB5A* in the single cell hepatocyte subnetwork, these results demonstrate alterations in core physiology related to metabolism, toxicological clearance, and cancer progression in the immortalized non-physiological hepatocyte cell lines.

In the subnetwork associated with comparison of each cell line-specific DEGs, which shared no connections with the subnetwork associated with the physiological hepatocytes, we obtained even sparser results with just one KD detected in one cell line, *CALMI* (Fig 4.4). This indicates that the cell lines do not have thoroughly connected subnetworks of DEGs, and instead the differential effects in their basal transcriptomes are more isolated. *CALMI*, or calmodulin 1, is an exosome-related gene whose aberrant expression has been linked to the progression of cancers of the liver^{277,278}. Interestingly, despite mapping all cell lines' DEGs to the single cell hepatocyte GRN, the network is dominated by HepG2 DEGs, and the *CALMI* KD is detected as a KD for HepG2 DEGs. The several *RPL* genes are ribosomal proteins with roles as cancer markers and in

the regulation of apoptosis²⁷⁹. *SLC25A3* is a mitochondrial solute carrier protein with critical roles in nutrient homeostasis and waste removal, with direct relevance to toxicological clearance^{280,281}. *SERBP1* and *XRCC5* have both been linked to cancer metastasis in the liver^{282,283}. Taken together, it is not surprising to see so many cancer marker genes highly represented in the subnetwork considering the cancerous origin of the HepG2 cell line. However, the cohesive nature of the cancer marker DEGs, combined with the inclusion of the proximal (by network distance) toxicologically-relevant *SLC25A3*, cautions the use of HepG2 as a physiologically relevant model of hepatocyte metabolism and toxicological response.

Hepatotoxicity case study supports toxicological response is better preserved in HepaRG

The above analyses support that the HepaRG cell line showed better similarity to *in vivo* hepatocytes based on basal gene expression patterns compared to other cell lines. We next sought to evaluate whether the baseline transcriptomic similarity between HepaRG and *in vivo* hepatocytes translated into HepaRG exhibiting a more similar toxicological response to *in vivo* hepatocytes, as this will inform the utility of this approach to *in vitro* cell line selection for hepatotoxicity studies. To this end, we curated 3 transcriptomic datasets from 24-hour tetrachlorodibenzodioxin (TCDD) exposure studies - one in HepaRG (GSE122518, 10 nM, microarray)²⁵⁰, one in HepG2 (GSE293127, 100 nM, bulk RNAseq)²⁵¹, and the third in a mouse model²⁵² (scRNAseq, 30 mg/kg, DEGs extracted from Single Cell Portal (SCP2482)). We expect that the DEGs from HepaRG would be more similar to those from the hepatocytes of the *in vivo* mouse scRNAseq studies compared to HepG2. We found 252 DEGs from the HepaRG study, 667 DEGs from the *in vivo* hepatocytes, and 1840 DEGs from HepG2 (Table 4.2). The differences in DEG numbers are likely at least in part due to differences in the sequencing platforms and the

dosage. With these limitations in mind, we assessed the overlap in the up- and downregulated DEGs derived from each cell line with those of the *in vivo* hepatocytes using Fisher's Exact Test.

The HepaRG DEGs in response to 24hr TCDD exposure showed stronger overlap with the *in vivo* DEGs (fold enrichment=3.73, Fisher's exact test $p=7.33\text{e-}8$), particularly the upregulated DEGs (fold enrichment=8.26, $p=4.57\text{e-}10$) than the HepG2 DEGs (fold enrichment=1.84 and $p=7.30\text{e-}8$ for overall DEGs; fold enrichment=2.78 and $p=1.78\text{e-}5$ for upregulated DEGs) (Table 4.2). By contrast, downregulated DEGs in HepG2 showed better overlap with the *in vivo* DEGs. However, while upregulated DEGs overlapping between HepaRG and *in vivo* hepatocytes included many cytochrome P450 genes directly involved in detoxification, only two of the downregulated DEGs (*SLC12A2* and *GNMT* overlapping between HepG2 and in *in vivo* hepatocytes (Fig. S.4.6) are relevant to detoxification. These results suggest that HepaRG may better recapitulate the hepatocyte toxicological response to TCDD.

Finally, we compared the biological pathways enriched by the up- and downregulated DEGs of each cell type (Fig. S.4.7). HepaRG and *in vivo* hepatocyte upregulated DEGs were enriched for pathways known to be affected by TCDD including xenobiotic metabolism by Cytochrome P450 and steroid hormone biosynthesis^{284–286}. HepG2, however, was more enriched for pathways and processes related to cellular structure and matrix organization among the upregulated DEGs. Interestingly, these pathways were downregulated in HepaRG and *in vivo* hepatocytes, further supporting the divergence of HepG2 from physiological hepatocytes. These results are also consistent with the well-characterized differences in HepG2 size and structure from *in vivo* hepatocytes²⁸⁷.

Overall, both the DEG and pathway comparisons across TCDD studies support that beyond better transcriptomic alignment at baseline, in a hepatotoxicity setting HepaRG better recapitulates the *in vivo* gene expression changes induced by the toxicant TCDD than does HepG2.

4.4 Discussion

A core challenge in toxicology research is balancing experimental feasibility with physiological relevance when selecting *in vitro* cell line models. The widely used human hepatocyte cell lines - HepG2, Hep3B, Huh7, and HepaRG – have been selected for use based on a variety of factors that have more to do with experimental feasibility and historical momentum than physiological relevance. These factors include the presence of key marker genes of interest, the developmental stage of the tissue from which cell lines were established, up- or downregulation of biological processes of interest, and compatibility with experimental procedures or assays²⁴⁻²⁷. While these factors are important to ensure the smooth execution of experiments, the value of those experiments diminishes substantially if the cell lines chosen do not accurately recapitulate *in vivo* physiology. The result of this is that many widely used cell lines fail to recapitulate fundamental aspects of hepatocyte biology, reducing the translational value of the resulting experiments.

To address this challenge, we applied a cross-platform transcriptomic comparison approach leveraging over 200 datasets from both human and murine sources. Our analysis revealed that HepaRG most closely resembles both human and mouse hepatocytes and primary hepatocytes, supporting its recognition as the most physiologically relevant immortalized hepatocyte model. HepaRG also captured TCDD-induced transcriptional responses more faithfully than the most commonly used HepG2 cell line, particularly for the upregulated genes. HepG2 exhibited greater divergence from *in vivo* transcriptional responses and performed better in capturing downregulated

rather than upregulated responses, but the toxicological relevance of the downregulated genes was relatively low. Other commonly used cell lines such as Hep3B and Huh7 demonstrated further divergence in baseline physiological transcriptomics, with enrichment of pathways more relevant to cancer and viral infection than to normal liver physiology. Together, these results underscore the need for careful selection of cell lines to balance physiological relevance and recapitulation of toxicological responses.

Our transcriptomic comparisons consistently demonstrated that the HepaRG cell line most closely resembles *in vivo* liver tissue and primary hepatocytes across both human and mouse sources. These results provide molecular support for prior reports lauding HepaRG's superior physiological fidelity over other cell lines, particularly in its ability to accurately induce cytochrome P450 enzymes, which are essential for xenobiotic metabolism^{260–262}. Thanks to its aligned basal gene expression with that of *in vivo* hepatocytes, HepaRG offers a strong platform for *in vitro* toxicity screening where translatability is a priority.

In evaluating toxicological response, HepaRG again proved advantageous. When assessing transcriptional responses to TCDD exposure, HepaRG showed a stronger concordance with *in vivo* hepatocytes than did HepG2. While HepG2 produced more total DEGs, at least partially due to the higher dose and superior sequencing platform employed in the curated studies, the HepaRG upregulated DEGs exhibited significantly better overlap with *in vivo* upregulated DEGs. This suggests that HepaRG more faithfully captures TCDD-induced transcriptional perturbations. These findings highlight HepaRG as the preferred choice for modeling toxicant-induced hepatocyte transcriptional perturbations *in vitro*.

Despite its widespread adoption as a “default” hepatocyte cell line, our analysis emphasizes several limitations of HepG2. Historically, HepG2 has been favored due to its experimental

tractability, robust growth, and compatibility with high throughput screening. However, there have been some concerns regarding the fact that they have been shown to differ in their sensitivity to genotoxic compounds due to alterations in the expression levels of cytochrome P450 enzymes from primary human hepatocytes²⁸⁸. Our analysis demonstrates that HepG2 diverges substantially from *in vivo* hepatocytes at the basal transcriptomic level. Particularly, several genes related to glucose and lipid metabolism were perturbed. While these differences have been leveraged to study hyperglycemia and glucose intolerance²⁸⁹, the stark divergence from core liver physiology emphasizes the relatively poor physiological relevance of the HepG2 cell line.

Our analysis also revealed incredible inconsistency in HepG2 basal transcriptomics across studies, as we observed distinct subclusters in our clustering analysis. We were not able to isolate consistent batch effects across studies that could explain this phenomenon. In our opinion, the most likely explanation for the subclusters is from differences in the passage number of the HepG2 cells used, as passage number is known to affect the gene expression patterns and physiological processes of *in vitro* cells^{290–292}. Specifically, with HepG2, the phenomenon of altered gene expression and toxicological clearance enzymes with increasing passage number has been reported²⁹³. Unfortunately, passage number often is not reported in the methods of *in vitro* studies, and this case highlights the need for comprehensive detailing of such information for replicability reasons. As demonstrated by our HepG2 results, neglecting to consider cell line stability under maintenance conditions may lead to significant deviations from physiological relevance.

In the toxicological response comparison, HepG2 showed partial utility in that its TCDD-induced downregulated DEGs aligned more closely with the *in vivo* response than those of HepaRG. However, it showed weaker overlap in the upregulated genes and the overall DEG sets, supporting the selection of HepaRG over HepG2. These results suggest that HepG2 may retain

partial relevance in studying gene downregulation, but overall caution its use due to its divergence between datasets and less transcriptomic similarity to human and mouse tissue and primary hepatocytes compared to HepaRG.

Compared to HepaRG and HepG2, Hep3B and Huh7 demonstrated further transcriptomic divergence from *in vivo* hepatocytes and deficits in core hepatocyte physiology. Our analysis revealed that Hepatitis B, Wnt signaling, and apoptosis pathways were enriched among Hep3B-specific DEGs that differ from *in vivo* hepatocytes. Huh7 demonstrated enrichment for Hepatitis C and lipid metabolism pathways relative to baseline *in vivo* hepatocytes. While these findings highlight deviations from core liver physiology, they are consistent with some of the applications in which these cell lines are applied. Hep3B, which lacks the p53 tumor suppressor gene and harbors integrated Hepatitis B viral DNA, is commonly used for studies of cancer-related signaling, apoptosis, and Hepatitis B viral studies^{294–298}. Huh7, known to be permissive to Hepatitis C replication, is frequently used in studies of Hepatitis C^{299–301}. Huh7 also contains abnormal intracellular lipid droplets and therefore has been used in studies of altered hepatocyte lipid composition³⁰². Taken together, these results indicate that these other hepatocyte cell lines may be more suitable as models in oncology or virology research than in translatable toxicological contexts. These results reinforce the importance of context-specific cell line selection and highlight the limitations of relying on convenience or historical precedence in hepatocyte *in vitro* model choice.

Beyond hepatocytes, our analysis also revealed insight about non-parenchymal cell lines corresponding to HSC and cholangiocyte cell types. Our clustering analysis did not reveal an optimal cell line candidate for HSCs or cholangiocytes in the same way that the hepatocyte cell lines did for hepatocytes with HepaRG. Additionally, the DEG analysis showed relatively small

perturbations, failing to generate very much meaningful insight. This is likely due to a combination of comparatively poor data availability (far fewer studies of non-parenchymal cell lines were available for curation), small sample sizes in those few studies, and high variance between them. More data is needed to apply this approach to generate meaningful insight about non-parenchymal cells.

Overall, our transcriptomics data-driven approach has enabled the large-scale cross-comparison of immortalized liver cell lines to guide *in vitro* cell line selection based on physiological resemblance to their corresponding *in vivo* cell types. Our analysis has produced results that underscore some of the biological considerations for the application of specific cell lines in specific experimental contexts, and demonstrate the advantage of using the HepaRG cell line for toxicology studies. This approach is able to consider data across sequencing platforms and experimental conditions to identify immortalized cell lines that best represent their corresponding *in vivo* cell types, and offers a facile method for experimenters to identify optimal cell lines for modeling other tissue types *in vitro* by leveraging existing data. However, we must acknowledge the inherent limitation is that the strength of this approach is directly related to the amount of data available for each cell line and *in vivo* cell type. Without sufficient data, reliable statistical analysis cannot be performed, and while trends may be able to be identified, the confidence in cell line selection decreases without as much statistical confidence. This phenomenon was exemplified by our findings with the non-parenchymal cell line analysis, which did not reveal an optimal cell line candidate for HSCs or cholangiocytes in the same way that the hepatocyte cell lines did for hepatocytes with HepaRG. To further improve *in vitro* modeling in toxicology, we advocate for the combination of this approach with 3D experimental methods such as spheroids to allow cells to better preserve their *in vivo* form and function while including cells that best resemble baseline

in vivo physiology. Additionally, the development of additional cell lines that more broadly cover genetic diversity will be critical to obtain population-wide translatable *in vitro* toxicity screening results. Overall, baseline transcriptomics comparison offers a valuable, data-driven approach to guide *in vitro* cell line selection for different purposes to obtain translatable results from *in vitro* screening studies.

4.5 Figures

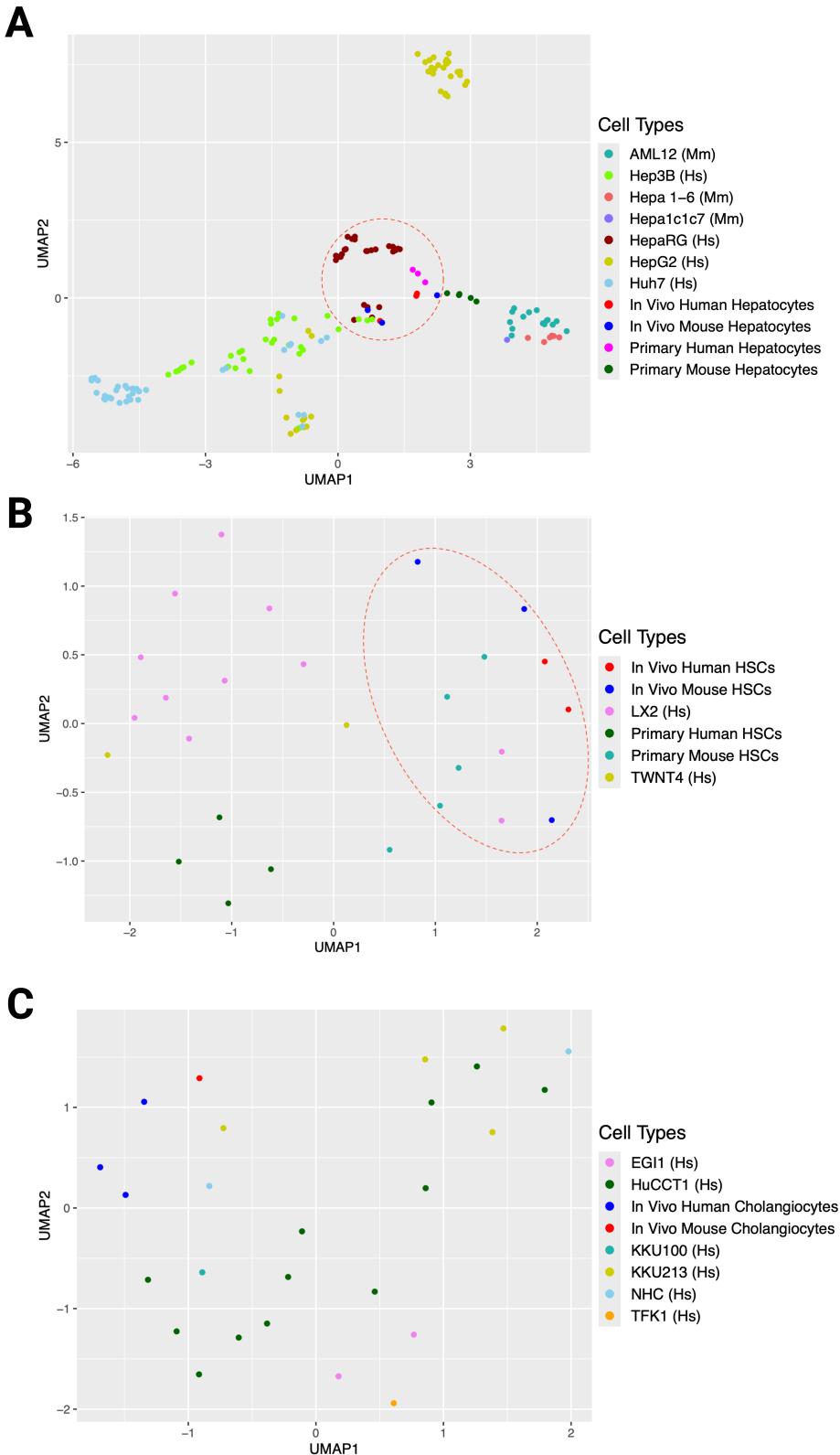


Fig. 4.1. UMAP reveals immortalized cell lines that closely resemble hepatic cell types at the level of the basal transcriptome. For hepatocyte (A), HSC (B), and cholangiocyte (C) datasets curated in this study, the top 2,500 most variable genes after quantile normalization across datasets and the curated marker genes for each cell type were taken for clustering analysis using UMAP. Missing basal expression values were imputed using MissForest in R prior to performing UMAP. Each point on each plot corresponds to a GSE Accession Number's basal gene expression values in the control samples. Points on each plot are colored according to the corresponding cell line as shown in each color key.

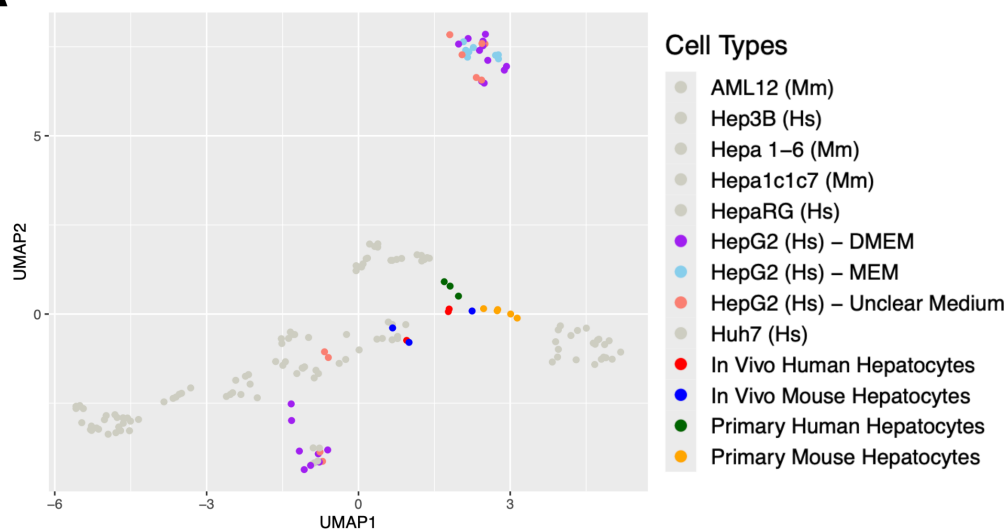
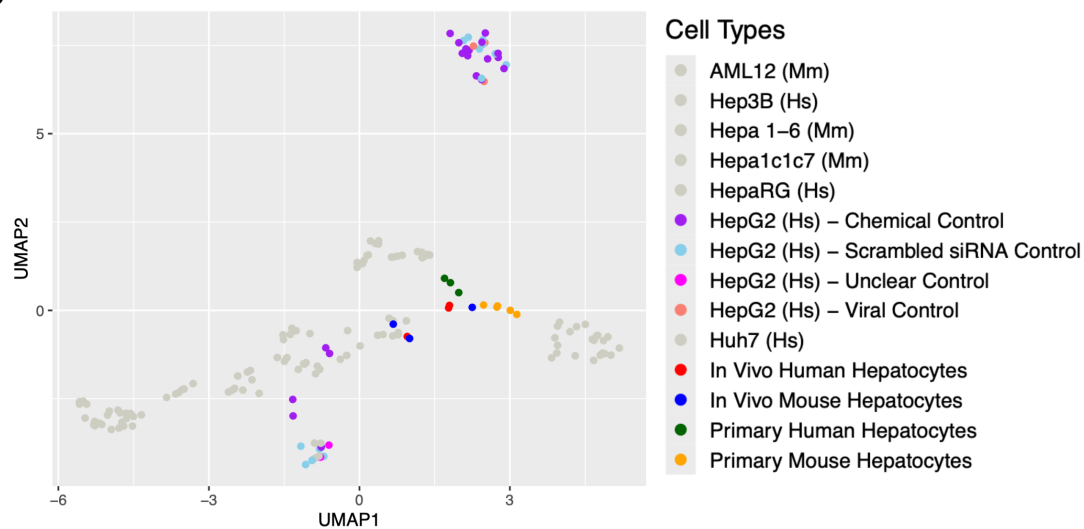
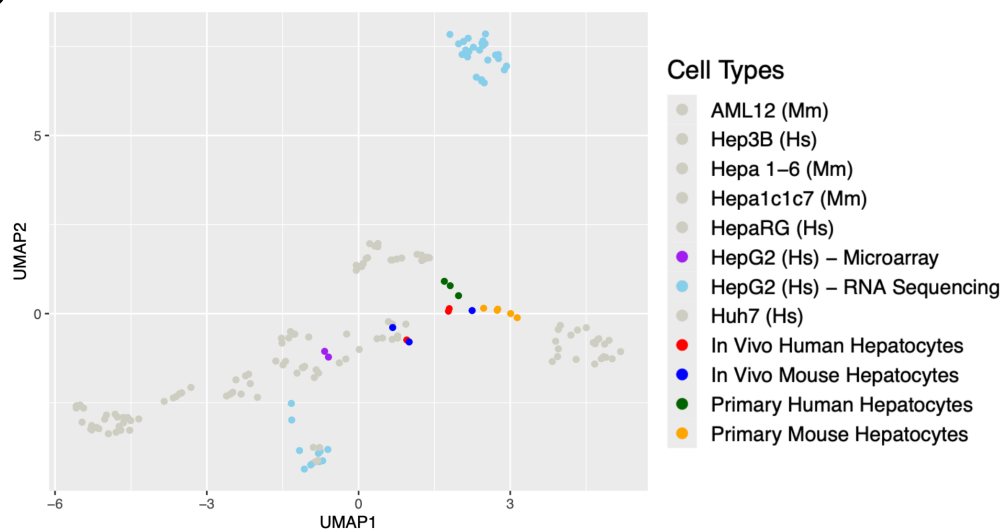
A**B****C**

Fig. 4.2. UMAP rules out several likely confounding variables as explanations for the emergence of distinct HepG2 subclusters. The hepatocyte cell line UMAP plot from Fig. 4.1A was annotated according to the medium (A), control (B), and sequencing platform (C) information found in the methods sections of each study curated for HepG2. Dots on the plots corresponding to HepG2 studies are colored according to the appropriate information. Notably, no distinct patterns emerge indicating subclusters based on any of these most-likely confounding variables.

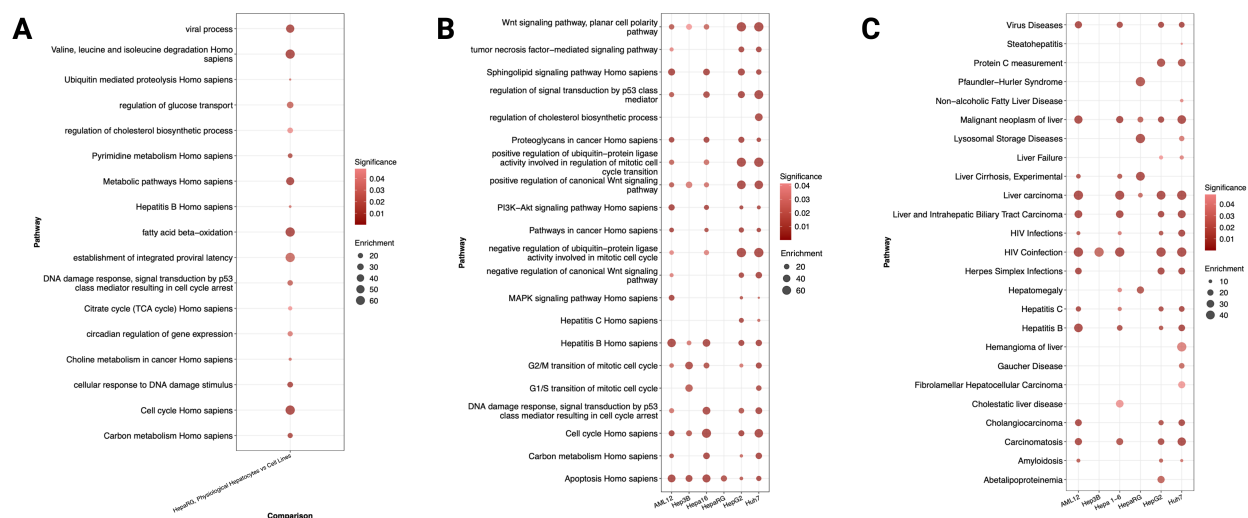


Fig. 4.3. Pathway enrichment results elucidate perturbed pathways and highlight relevant human diseases among hepatocyte cell lines. Pathway enrichment analysis was conducted with EnrichR using the set of all significant DEGs identified above. The lists of DEGs were analyzed against the GO Biological Process, KEGG, and DisGeNET databases. Pathways with a lower FDR are shown in darker red hues, and the size of the circle corresponds to the significance of EnrichR's Enrichment Score. **A)** GOBP and KEGG pathway enrichment results for the comparison of HepaRG and physiological hepatocytes against the other hepatocyte cell lines reveals key biological pathways enriched in physiological hepatocytes relative to immortalized hepatocyte cell lines, highlighting important physiological distinctions from popular *in vitro* hepatocyte models. **B)** GOBP and KEGG pathway enrichment results for each cell line relative to *in vivo* hepatocytes reveal shared and distinct pathways enriched across cell lines, highlighting differences from core hepatocyte physiology. **C)** DisGeNET disease enrichment results for each cell line relative to *in vivo* hepatocytes reveal human diseases for which the basal transcriptome of each cell line are enriched, highlighting relevant potential applications of each cell line in human disease research.

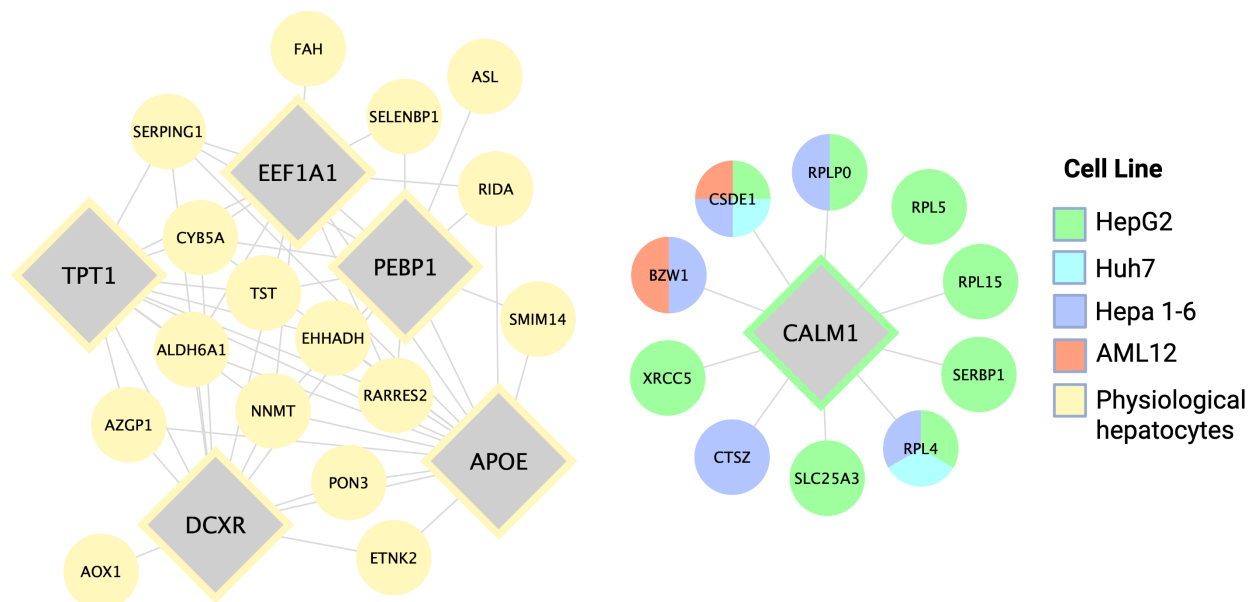


Fig. 4.4. Key driver analysis reveals key regulators driving the differences in core hepatocyte physiology from immortalized hepatocyte cell lines. The top 300 DEGs determined by the Wilcoxon Rank-Sum Test were used to conduct KDA using Mergeomics. The resulting key drivers at FDR < 0.05 were mapped onto the single cell hepatocyte GRN from SCING. Key drivers from the comparison of HepaRG and physiological hepatocyte (e.g. primary hepatocytes and *in vivo* single cell reference data from human and mouse) gene expression against the other hepatocyte cell lines, and for each cell line relative to physiological hepatocytes, were mapped onto the single cell hepatocyte GRN to highlight key regulators driving core hepatocyte physiology that are altered in the other cell lines. Network nodes are colored according to the cell line(s) in which the genes are DEGs, and the KD border color indicates the cell line for which the gene is a KD. Network nodes that are grey were not detected as DEGs in the Wilcoxon Rank-Sum Test but were still detected as KDs. Despite being mapped onto the same network, the physiologically relevant and cell line-specific perturbed hepatocyte subnetworks did not overlap.

4.6 Tables

Table 4.1. DEG number and top pathway enrichment results enable comparison between hepatocyte cell lines and *in vivo* hepatocyte basal gene expression data. Global basal gene expression data was used to perform DEG analysis using the Wilcoxon Rank-Sum Test. DEGs in each cell line relative to *in vivo* hepatocytes were determined at FDR < 0.05. Log2 fold change values were also determined using the quantile normalized expression values of each dataset, and the top 5 up- and downregulated genes by Log2FC are shown. All DEGs were used for pathway enrichment analysis using EnrichR. The top 5 enriched pathways are also shown.

Cell Line Comparison	Significant DEGs	Top Upregulated DEGs by Log2FC	Top Downregulated DEGs by Log2FC	Top 5 KEGG/GOBP Pathways
HepaRG vs <i>In Vivo</i> Human Hepatocytes	861	<i>GJA1, IFI30, TM4SF1, PMP22, SLPI</i>	<i>MED12L, SCRG1, NEK5, ADCY2, GPR65</i>	Lysosomal processes Neutrophil degranulation Regulation of macroautophagy Purine metabolism Apoptosis
Hep3B vs <i>In Vivo</i> Human Hepatocytes	700	<i>IFI30, TGFBI, ITM2C, CNN2, CAPG</i>	<i>NFE2, C4B, NR1I3, PRSS27, SEC14L2</i>	mRNA splicing Lysosomal processes G2/M transition of cell cycle Pyrimidine metabolism Drug metabolism
HepG2 vs <i>In Vivo</i> Human Hepatocytes	2,871	<i>FLNA, TM4SF1, GDF15, TGFBI, TNFRSF21</i>	<i>DNAJC28, SLC27A3, CYP2E1, ARHGAP27, PRRT2</i>	mRNA splicing Protein processing Protein ubiquitination Viral transcription Wnt signaling pathway
Huh7 vs <i>In Vivo</i> Human Hepatocytes	3,036	<i>GJA1, NID1, FLNA, ITM2C, CNN2</i>	<i>IGSF10, PCDHGA5, PCDHGA9, ENKUR, PRSS27</i>	mRNA splicing Protein ubiquitination Amino acid metabolism Endocytosis processes Membrane organization
AML12 vs <i>In Vivo</i> Mouse Hepatocytes	2,033	<i>CCL2, FLNA, ANXA3, OXCT1, TAGLN2</i>	<i>SAP25, RAB3B, BHMT2, TBX15, PKD2L2</i>	mRNA splicing Focal adhesion Protein ubiquitination Hepatitis B Actin cytoskeleton regulation
Hepa 1-6 vs <i>In Vivo</i> Mouse Hepatocytes	1,089	<i>FLNA, TAGLN2, ATP6V0C, TPM4, ANXA3</i>	<i>PHKG1, INCA1, BHMT2, NR1I3, FMO5</i>	mRNA splicing Protein processing Cell cycle Actin cytoskeleton regulation Hepatitis B
HepaRG, <i>In Vivo</i> Physiological Hepatocytes vs All Hepatocyte Cell Lines	7,075	<i>SAAI, ANXA10, CYP2E1, ITIH4, RDH16</i>	<i>BRDT, SCRNI, DPPA4, SLC7A5, PXDN</i>	SRP-dependent cotranslational protein targeting Translational initiation Viral transcription Nonsense-mediated mRNA decay Translation

Table 4.2. Fisher’s Exact Test overlap statistics potentially indicate HepaRG as better capturing the toxicological response to TCDD exposure than HepG2, but highlight the importance of experimental context in cell line selection. Three time-matched exposure studies to TCDD were curated to assess the transcriptomic responses in HepaRG (GSE122518), HepG2 (GSE293127), and *in vivo* mouse hepatocytes (SCP2482). Fisher’s Exact Test was performed between the sets of upregulated, downregulated, and total DEGs for each cell line against those of the *in vivo* hepatocytes. HepaRG showed more significant enrichment in the upregulated genes, while HepG2 showed a higher capture rate of the total *in vivo* DEGs and more significant enrichment in the downregulated genes.

Signature Comparison	Cell Line Signature Size	<i>In Vivo</i> Overlap Number	P-Value	Fold Enrichment	Odds Ratio	Percent Cell Line Signature <i>In Vivo</i>
HepaRG Up vs <i>In Vivo</i> Up	136	15	4.57e-10	8.26	9.51	11.03%
HepaRG Down vs <i>In Vivo</i> Down	116	2	0.37	1.55	1.57	1.72%
HepaRG Total vs <i>In Vivo</i> Total	252	23	7.33e-8	3.73	4.11	9.13%
HepG2 Up vs <i>In Vivo</i> Up	1,128	30	4.43e-4	1.99	2.11	2.66%
HepG2 Down vs <i>In Vivo</i> Down	712	22	1.78e-5	2.78	2.98	7.26%
HepG2 Total vs <i>In Vivo</i> Total	1,840	83	7.30e-8	1.84	2.01	12.44%

Chapter 5. Conclusions and Future Directions

This dissertation sought to contribute to the advancement of the field of toxicology by aiding in the development and adoption of NAM approaches to toxicological research. Specifically, we developed and applied big transcriptomics data-driven approaches to better understand toxicologically relevant exposures using both *in silico* and *in vitro* approaches consistent with the goals of NAM development. Across the three aims, I showcased the utility of large-scale computational analyses in generating mechanistic insights across toxicologically relevant compounds, providing resources for translational toxicology, and guiding *in vitro* experimental design. Together, these studies emphasize the value of leveraging hundreds to thousands of publicly available transcriptomics datasets to drive both scientific discovery and practical improvements in experimental design as well as to evaluate reproducibility and variability across studies.

In **Aim 1**, we conducted a comprehensive multi-species, multi-tissue transcriptomic meta analyses of the cannabinoids THC and CBD to date. This study revealed fundamental differences in their transcriptional stability across species and tissues, with CBD eliciting more consistent effects on metabolic, inflammatory, and cell cycle pathways, while THC responses were more highly variable and context-dependent. These findings provide a molecular framework for balancing the discordant therapeutic potential with the potential health risks of cannabinoids, and highlights how the integration of diverse transcriptomics datasets can overcome individual study limitations to extract robust, generalizable biological signatures.

Aim 2 expanded this principle in the development of ToxiOmics, an open access database and web tool cataloging the transcriptomic signatures for 222 toxicologically relevant compounds across species and tissues to enable the research community to understand the generalizable and

context-specific activities of toxicants. Our PFAS case study demonstrated the translational power of ToxiOmics to show that GenX and legacy PFAS compounds share PPAR-mediated lipid metabolism-centric mechanisms that carry implications for steatosis, insulin resistance, and potentially carcinogenesis. Beyond providing mechanistic insights, ToxiOmics enables cross-chemical comparisons via overlap-, coexpression-, and network-based approaches, and enables the exploration of potential toxicant-drug interactions via integration with PharmOmics. ToxiOmics represents a scalable platform for mechanistic toxicology and hypothesis generation to help focus future research through the exploration of cataloged and user-supplied data which will help community users generate data-driven hypotheses and plan informed mechanistic and validation experiments.

Finally, **Aim 3** addressed the physiological fidelity of *in vitro* models in recapitulating *in vivo* cellular functions and toxicological responses. Through the comparison of 16 hepatic cell lines with *in vivo* physiological human and mouse liver cells, we demonstrated that HepaRG most accurately recapitulates core hepatocyte physiology, while other widely used cell lines such as the most prevalent HepG2 diverge in core metabolic pathways and show high variability across studies. This analysis provides a transcriptomics-based big data-driven framework for guiding *in vitro* cell line selection, emphasizing the need to balance physiological relevance with experimental compatibility while underscoring the strong influence that *in vitro* cell line selection has over toxicological inference.

Together, these aims support the development and advancement of NAM approaches by offering both conceptual frameworks and practical resources. We have shown that publicly available transcriptomics data can be leveraged to extract robust, generalizable molecular insights

across diverse biological contexts, build scalable platforms for toxicogenomics research, and guide sensible selection of physiologically relevant *in vitro* models.

Several future directions arise from this work, particularly in the case of ToxiOmics. First, the inclusion of other omics data types such as proteomics, metabolomics, and epigenomics will enhance mechanistic resolution and provide a more complete biological picture of the effects of toxicological perturbation than transcriptomics will alone. Second, the development of a pipeline to account for mixtures of toxicological exposures will enable ToxiOmics to more accurately reflect real exposures faced by real people. Finally, in the case of the cell lines, the incorporation of this transcriptomics-guided approach with 3D cell culture techniques will enable researchers to better preserve *in vivo* physiology both at the level of individual cells and at the level of the culture as a whole, as the preservation of physical cell dynamics will allow cells to connect and communicate in a fashion that more accurately represents their *in vivo* alignment.

Overall, this dissertation establishes the value of big transcriptomics data-based, systems-level approaches for advancing mechanistic understanding, improving model selection in toxicity studies, and building translational *in silico* resources for toxicologists. Through the pairing of computational innovation with experimental relevance, this work contributes to the ongoing efforts in the field to shift toward data-driven, ethical, sustainable, and human-relevant toxicological science.

REFERENCES

1. Stokes WS. Animals and the 3Rs in toxicology research and testing: The way forward: The way forward. *Hum Exp Toxicol* [Internet]. SAGE Publications; 2015 Dec;34(12):1297–1303. Available from: <http://dx.doi.org/10.1177/0960327115598410> PMID: 26614819
2. Chapman KL, Holzgrefe H, Black LE, Brown M, Chellman G, Copeman C, Couch J, Creton S, Gehen S, Hoberman A, Kinter LB, Madden S, Mattis C, Stemple HA, Wilson S. Pharmaceutical toxicology: designing studies to reduce animal use, while maximizing human translation. *Regul Toxicol Pharmacol* [Internet]. Elsevier BV; 2013 Jun;66(1):88–103. Available from: <http://dx.doi.org/10.1016/j.yrtph.2013.03.001> PMID: 23524271
3. Jaeschke H, Ramachandran A. Are new Approach Methodologies (NAMs) the holy grail of toxicology? *Toxicol Sci* [Internet]. Oxford University Press (OUP); 2025 Aug 7;(kfaf113):kfaf113. Available from: <https://dx.doi.org/10.1093/toxsci/kfaf113> PMID: 40795217
4. Mathisen GH, Bearth A, Jones LB, Hoffmann S, Vist GE, Ames HM, Husøy T, Svendsen C, Tsaïoun K, Ashikaga T, Bloch D, Cavoski A, Chiu WA, Davies HG, Giusti A, Hartung T, Hirabayashi Y, Hogberg HT, Joglekar R, Kojima H, Krishnan K, Kwon S, Osborne OJ, Roggen E, Rooney AA, Rousselle C, Sass JB, Sepai O, Simanainen U, Thayer KA, Tong W, Wikoff D, Wright F, Whaley P. Time for CHANGE: system-level interventions for bringing forward the date of effective use of NAMs in regulatory toxicology. *Arch Toxicol* [Internet]. Springer Science and Business Media LLC; 2024 Aug;98(8):2299–2308. Available from: <http://dx.doi.org/10.1007/s00204-024-03802-6> PMCID: PMC11784933
5. Luch A, editor. Molecular, clinical and environmental toxicology: Volume 1: Molecular toxicology [Internet]. 2009th ed. Basel, Switzerland: Birkhauser Verlag AG; 2008. Available from: <http://dx.doi.org/10.1007/978-3-7643-8336-7>
6. Searfoss GH, Ryan TP, Jolly RA. The role of transcriptome analysis in pre-clinical toxicology. *Curr Mol Med* [Internet]. Bentham Science Publishers Ltd.; 2005 Feb;5(1):53–64. Available from: <http://dx.doi.org/10.2174/1566524053152825> PMID: 15720270
7. Joseph P. Transcriptomics in toxicology. *Food Chem Toxicol* [Internet]. 2017 Nov;109(Pt 1):650–662. Available from: <http://dx.doi.org/10.1016/j.fct.2017.07.031> PMCID: PMC6419952
8. Mathieson I. The omnigenic model and polygenic prediction of complex traits. *Am J Hum Genet* [Internet]. Elsevier BV; 2021 Sep 2;108(9):1558–1563. Available from: <http://dx.doi.org/10.1016/j.ajhg.2021.07.003> PMCID: PMC8456163
9. Boyle EA, Li YI, Pritchard JK. An expanded view of complex traits: From polygenic to omnigenic. *Cell* [Internet]. Elsevier; 2017 Jun 15;169(7):1177–1186. Available from: <http://dx.doi.org/10.1016/j.cell.2017.05.038> PMCID: PMC5536862
10. Rock EM, Parker LA. Constituents of cannabis Sativa. *Adv Exp Med Biol* [Internet]. Adv

- Exp Med Biol; 2021;1264:1–13. Available from: http://dx.doi.org/10.1007/978-3-030-57369-0_1 PMID: 33332000
11. Carliner H, Brown QL, Sarvet AL, Hasin DS. Cannabis use, attitudes, and legal status in the U.S.: A review. *Prev Med* [Internet]. 2017 Nov;104:13–23. Available from: <http://dx.doi.org/10.1016/j.ypmed.2017.07.008> PMCID: PMC6348863
 12. Hasin DS. US epidemiology of cannabis use and associated problems. *Neuropsychopharmacology* [Internet]. Nature Publishing Group; 2018 Jan;43(1):195–212. Available from: <http://dx.doi.org/10.1038/npp.2017.198> PMCID: PMC5719106
 13. Blake A, Wan BA, Malek L, DeAngelis C, Diaz P, Lao N, Chow E, O’Hearn S. A selective review of medical cannabis in cancer pain management. *Ann Palliat Med* [Internet]. AME Publishing Company; 2017 Dec;6(Suppl 2):S215–S222. Available from: <http://dx.doi.org/10.21037/apm.2017.08.05> PMID: 28866904
 14. Häuser W, Fitzcharles MA, Radbruch L, Petzke F. Cannabinoids in pain management and palliative medicine. *Dtsch Arztebl Int* [Internet]. 2017 Sep 22;114(38):627–634. Available from: <http://dx.doi.org/10.3238/arztebl.2017.0627> PMCID: PMC5645627
 15. Andrew Kerklaan DC. Exploring CBD for Anxiety Symptoms: A Balanced Perspective [Internet]. Dr. Kerklaan Therapeutics. 2025 [cited 2025 Aug 16]. Available from: <https://drkerklaan.com/blogs/explore/cbd-for-anxiety-symptoms>
 16. Sherman A, Chin J. Cannabis and CBD for health and wellness: An essential guide for using nature’s medicine to relieve stress, anxiety, chronic pain, inflammation, and more. Berkeley, CA: Ten Speed Press; 2019.
 17. Knapp AA, Lee DC, Borodovsky JT, Auty SG, Gabrielli J, Budney AJ. Emerging trends in cannabis administration among adolescent cannabis users. *J Adolesc Health* [Internet]. Elsevier BV; 2019 Apr;64(4):487–493. Available from: <http://dx.doi.org/10.1016/j.jadohealth.2018.07.012> PMCID: PMC6408312
 18. Sultan RS, Zhang AW, Olfson M, Kwizera MH, Levin FR. Nondisordered cannabis use among US adolescents. *JAMA Netw Open* [Internet]. American Medical Association; 2023 May 1;6(5):e2311294. Available from: <http://dx.doi.org/10.1001/jamanetworkopen.2023.11294> PMCID: PMC10157425
 19. Mattingly CJ, Colby GT, Forrest JN, Boyer JL. The Comparative Toxicogenomics Database (CTD). *Environ Health Perspect* [Internet]. 2003 May;111(6):793–795. Available from: <http://dx.doi.org/10.1289/ehp.6028> PMCID: PMC1241500
 20. Wang Z, Monteiro CD, Jagodnik KM, Fernandez NF, Gundersen GW, Rouillard AD, Jenkins SL, Feldmann AS, Hu KS, McDermott MG, Duan Q, Clark NR, Jones MR, Kou Y, Goff T, Woodland H, Amaral FMR, Szeto GL, Fuchs O, Schüssler-Fiorenza Rose SM, Sharma S, Schwartz U, Bausela XB, Szymkiewicz M, Maroulis V, Salykin A, Barra CM, Kruth CD, Bongio NJ, Mathur V, Todoric RD, Rubin UE, Malatras A, Fulp CT, Galindo JA, Motiejunaite R, Jüschke C, Dishuck PC, Lahl K, Jafari M, Aibar S, Zaravinos A, Steenhuizen

- LH, Allison LR, Gamallo P, de Andres Segura F, Dae Devlin T, Pérez-García V, Ma'ayan A. Extraction and analysis of signatures from the Gene Expression Omnibus by the crowd. *Nat Commun* [Internet]. Springer Science and Business Media LLC; 2016 Sep 26;7(1):12846. Available from: <http://dx.doi.org/10.1038/ncomms12846> PMID: PMC5052684
21. Shu L, Zhao Y, Kurt Z, Byars SG, Tukiainen T, Kettunen J, Orozco LD, Pellegrini M, Lusi AJ, Ripatti S, Zhang B, Inouye M, Mäkinen VP, Yang X. Mergeomics: multidimensional data integration to identify pathogenic perturbations to biological systems. *BMC Genomics* [Internet]. Springer Science and Business Media LLC; 2016 Nov 4;17(1):874. Available from: <http://dx.doi.org/10.1186/s12864-016-3198-9> PMID: PMC5097440
 22. Ding J, Blencowe M, Nghiem T, Ha SM, Chen YW, Li G, Yang X. Mergeomics 2.0: a web server for multi-omics data integration to elucidate disease networks and predict therapeutics. *Nucleic Acids Res* [Internet]. Oxford University Press (OUP); 2021 Jul 2;49(W1):W375–W387. Available from: <https://dx.doi.org/10.1093/nar/gkab405> PMID: PMC8262738
 23. Chen YW, Diamante G, Ding J, Nghiem TX, Yang J, Ha SM, Cohn P, Arneson D, Blencowe M, Garcia J, Zaghari N, Patel P, Yang X. PharmOmics: A species- and tissue-specific drug signature database and gene-network-based drug repositioning tool. *iScience* [Internet]. Elsevier BV; 2022 Apr 15;25(4):104052. Available from: <http://dx.doi.org/10.1016/j.isci.2022.104052> PMID: PMC8957031
 24. Skopek R, Palusińska M, Kaczor-Keller K, Pingwara R, Papierniak-Wyglądała A, Schenk T, Lewicki S, Zelent A, Szymański Ł. Choosing the Right Cell Line for Acute Myeloid Leukemia (AML) Research. *Int J Mol Sci* [Internet]. 2023 Mar 11;24(6). Available from: <http://dx.doi.org/10.3390/ijms24065377> PMID: PMC10049680
 25. Lendahl U, McKay RD. The use of cell lines in neurobiology. *Trends Neurosci* [Internet]. 1990 Apr;13(4):132–137. Available from: [http://dx.doi.org/10.1016/0166-2236\(90\)90004-t](http://dx.doi.org/10.1016/0166-2236(90)90004-t) PMID: 1692169
 26. Brodaczewska KK, Szczylik C, Fiedorowicz M, Porta C, Czarnecka AM. Choosing the right cell line for renal cell cancer research. *Mol Cancer* [Internet]. 2016 Dec 19;15(1):83. Available from: <http://dx.doi.org/10.1186/s12943-016-0565-8> PMID: PMC5168717
 27. Hare D, Collins S, Cuddington B, Mossman K. The Importance of Physiologically Relevant Cell Lines for Studying Virus-Host Interactions. *Viruses* [Internet]. 2016 Nov 1;8(11). Available from: <http://dx.doi.org/10.3390/v8110297> PMID: PMC5127011
 28. Smart R, Pacula RL. Early evidence of the impact of cannabis legalization on cannabis use, cannabis use disorder, and the use of other substances: Findings from state policy evaluations. *Am J Drug Alcohol Abuse* [Internet]. Informa UK Limited; 2019 Oct 11;45(6):644–663. Available from: <http://dx.doi.org/10.1080/00952990.2019.1669626> PMID: PMC6934162
 29. Miech RA, Johnston LD, O'Malley PM, Bachman JG, Schulenberg JE, Patrick ME. Monitoring the Future national survey results on drug use, 1975-2017: Volume I, secondary school students [Internet]. The University of Michigan Institute for Social Research; 2018. Available from: <http://dx.doi.org/10.3998/2027.42/146530>

30. Mattingly DT, Richardson MK, Hart JL. Prevalence of and trends in current cannabis use among US youth and adults, 2013-2022. *Drug Alcohol Depend Rep* [Internet]. Elsevier BV; 2024 Sep;12(100253):100253. Available from: <http://dx.doi.org/10.1016/j.dadr.2024.100253> PMID: PMC11262189
31. Maroon J, Bost J. Review of the neurological benefits of phytocannabinoids. *Surg Neurol Int* [Internet]. Scientific Scholar; 2018;9(1):91. Available from: http://dx.doi.org/10.4103/sni.sni_45_18
32. Lazarjani P, Torres M, Hooker S, Fowlie T, Young C, Seyfoddin O. Methods for quantification of cannabinoids: a narrative review. *J Cannabis Res*. 2020;2:1–10.
33. Pertwee RG. The pharmacology of cannabinoid receptors and their ligands: an overview. *Int J Obes (Lond)* [Internet]. Springer Science and Business Media LLC; 2006 Apr;30 Suppl 1(S1):S13–8. Available from: <http://dx.doi.org/10.1038/sj.ijo.0803272> PMID: 16570099
34. Pagano C, Navarra G, Coppola L, Avilia G, Bifulco M, Laezza C. Cannabinoids: Therapeutic use in clinical practice. *Int J Mol Sci* [Internet]. MDPI AG; 2022 Mar 19;23(6):3344. Available from: <http://dx.doi.org/10.3390/ijms23063344> PMID: PMC8952215
35. Batalla A, Bos J, Postma A, Bossong MG. The impact of cannabidiol on human brain function: A systematic review. *Front Pharmacol* [Internet]. Frontiers Media SA; 2020;11:618184. Available from: <http://dx.doi.org/10.3389/fphar.2020.618184> PMID: PMC7858248
36. Malabadi RB, Kolkar K, Chalannavar R. Δ^9 -Tetrahydrocannabinol (THC): The major psychoactive component is of botanical origin. *Int J Innov Sci Res Rev*. 2023;5(3):4177–4184.
37. Machado Bergamaschi M, Queiroz HC, Zuardi W, Crippa A. Safety and side effects of cannabidiol, a Cannabis sativa constituent. *Curr Drug Saf*. 2011;6(4):237–249.
38. Sachs J, McGlade E, Yurgelun-Todd D. Safety and toxicology of cannabinoids. *Neurotherapeutics* [Internet]. Springer Science and Business Media LLC; 2015 Oct;12(4):735–746. Available from: <http://dx.doi.org/10.1007/s13311-015-0380-8> PMID: PMC4604177
39. Zuo Y, Iemolo A, Montilla-Perez P, Li HR, Yang X, Telese F. Chronic adolescent exposure to cannabis in mice leads to sex-biased changes in gene expression networks across brain regions. *Neuropsychopharmacology* [Internet]. Springer Science and Business Media LLC; 2022 Nov;47(12):2071–2080. Available from: <http://dx.doi.org/10.1038/s41386-022-01413-2> PMID: PMC9556757
40. Puighermanal E, Luna-Sánchez M, Gella A, van der Walt G, Urpi A, Royo M, Tena-Morraja P, Appiah I, de Donato MH, Menardy F, Bianchi P, Esteve-Codina A, Rodríguez-Pascual L, Vergara C, Gómez-Pallarès M, Marsicano G, Bellocchio L, Martinell M, Sanz E, Jurado S, Soriano FX, Pizcueta P, Quintana A. Cannabidiol ameliorates mitochondrial disease via PPAR γ activation in preclinical models. *Nat Commun* [Internet]. Springer Science and

Business Media LLC; 2024 Sep 4;15(1):7730. Available from: <http://dx.doi.org/10.1038/s41467-024-51884-8> PMID: PMC11375224

41. Shapira G, Israel-Elgali I, Grad M, Avnat E, Rachmany L, Sarne Y, Shomron N. Hippocampal differential expression underlying the neuroprotective effect of delta-9-tetrahydrocannabinol microdose on old mice. *Front Neurosci* [Internet]. 2023 Jul 18;17:1182932. Available from: <http://dx.doi.org/10.3389/fnins.2023.1182932> PMID: PMC10393280
42. Bilkei-Gorzo A, Albayram O, Draffehn A, Michel K, Piyanova A, Oppenheimer H. A chronic low dose of Δ 9-tetrahydrocannabinol (THC) restores cognitive function in old mice. *Nat Med*. 2017;23(6).
43. Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* [Internet]. Oxford University Press (OUP); 2002 Jan 1;30(1):207–210. Available from: <http://dx.doi.org/10.1093/nar/30.1.207> PMID: PMC99122
44. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M. NCBI GEO: archive for functional genomics data sets-update. *Nucleic Acids Res*. 2012;41(D1):D991–995.
45. Zamora Z, Wang S, Chen YW, Diamante G, Yang X. Systematic transcriptome-wide meta-analysis across endocrine disrupting chemicals reveals shared and unique liver pathways, gene networks, and disease associations. *Environ Int* [Internet]. Elsevier BV; 2024 Jan;183(108339):108339. Available from: <http://dx.doi.org/10.1016/j.envint.2023.108339> PMID: PMC11216742
46. Diamante G, Cely I, Zamora Z, Ding J, Blencowe M, Lang J, Bline A, Singh M, Lusi AJ, Yang X. Systems toxicogenomics of prenatal low-dose BPA exposure on liver metabolic pathways, gut microbiota, and metabolic health in mice. *Environ Int* [Internet]. Elsevier BV; 2021 Jan;146(106260):106260. Available from: <http://dx.doi.org/10.1016/j.envint.2020.106260> PMID: PMC7775895
47. Chen Y, Shu L, Qiu Z, Lee DY, Settle SJ, Que Hee S, Telesca D, Yang X, Allard P. Exposure to the BPA-substitute Bisphenol S causes unique alterations of germline function. *PLoS Genet* [Internet]. Public Library of Science (PLoS); 2016 Jul;12(7):e1006223. Available from: <http://dx.doi.org/10.1371/journal.pgen.1006223> PMID: PMC4966967
48. Davis S, Meltzer PS. GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. *Bioinformatics* [Internet]. Oxford University Press (OUP); 2007 Jul 15;23(14):1846–1847. Available from: <http://dx.doi.org/10.1093/bioinformatics/btm254> PMID: 17496320
49. Leinonen R, Sugawara H, Shumway M. International Nucleotide Sequence Database Collaboration. The sequence read archive. *Nucleic Acids Res*. 2010;39(suppl_1):D19–21.
50. Krueger F. Trim Galore!: A wrapper around Cutadapt and FastQC to consistently apply

- adapter and quality trimming to FastQ files, with extra functionality for RRBS data. Babraham Institute [Internet]. 2015 [cited 2024 Jan 29]; Available from: <https://cir.nii.ac.jp/crid/1370294643762929691>
51. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J [Internet]. EMBnet Stichting; 2011 May 2;17(1):10. Available from: <http://dx.doi.org/10.14806/ej.17.1.200>
 52. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. Nat Methods [Internet]. Springer Science and Business Media LLC; 2017 Apr;14(4):417–419. Available from: <http://dx.doi.org/10.1038/nmeth.4197> PMID: PMC5600148
 53. Sonesson C, Love MI, Robinson MD. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. F1000Res [Internet]. 2015 Dec 30;4:1521. Available from: <http://dx.doi.org/10.12688/f1000research.7563.2> PMID: PMC4712774
 54. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res [Internet]. Oxford University Press (OUP); 2015 Apr 20;43(7):e47. Available from: <http://dx.doi.org/10.1093/nar/gkv007> PMID: PMC4402510
 55. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2 [Internet]. bioRxiv. bioRxiv; 2014. Available from: <http://dx.doi.org/10.1101/002832>
 56. Stekhoven DJ, Bühlmann P. MissForest--non-parametric missing value imputation for mixed-type data. Bioinformatics [Internet]. Oxford University Press (OUP); 2012 Jan 1;28(1):112–118. Available from: <https://dx.doi.org/10.1093/bioinformatics/btr597> PMID: 22039212
 57. Kolde R, Laur S, Adler P, Vilo J. Robust rank aggregation for gene list integration and meta-analysis. Bioinformatics [Internet]. Oxford University Press (OUP); 2012 Feb 15;28(4):573–580. Available from: <https://dx.doi.org/10.1093/bioinformatics/btr709> PMID: PMC3278763
 58. Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles GV, Clark NR, Ma'ayan A. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics [Internet]. Springer Science and Business Media LLC; 2013 Apr 15;14(1):128. Available from: <http://dx.doi.org/10.1186/1471-2105-14-128> PMID: PMC3637064
 59. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet [Internet]. Springer Science and Business Media LLC; 2000 May;25(1):25–29. Available from: <http://dx.doi.org/10.1038/75556> PMID: PMC3037419

60. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* [Internet]. Cold Spring Harbor Laboratory; 2003 Nov;13(11):2498–2504. Available from: <http://dx.doi.org/10.1101/gr.1239303> PMID: PMC403769
61. MacArthur J, Bowler E, Cerezo M, Gil L, Hall P, Hastings E, Junkins H, McMahon A, Milano A, Morales J, Pendlington ZM, Welter D, Burdett T, Hindorff L, Flicek P, Cunningham F, Parkinson H. The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res* [Internet]. 2017 Jan 4;45(D1):D896–D901. Available from: <http://dx.doi.org/10.1093/nar/gkw1133> PMID: PMC5210590
62. Hevner RF, Shi L, Justice N, Hsueh Y, Sheng M, Smiga S, Bulfone A, Goffinet AM, Campagnoni AT, Rubenstein JL. *Tbr1* regulates differentiation of the preplate and layer 6. *Neuron* [Internet]. Elsevier BV; 2001 Feb;29(2):353–366. Available from: [http://dx.doi.org/10.1016/s0896-6273\(01\)00211-2](http://dx.doi.org/10.1016/s0896-6273(01)00211-2) PMID: 11239428
63. Zhou J, Wade SD, Graykowski D, Xiao MF, Zhao B, Giannini LAA, Hanson JE, van Swieten JC, Sheng M, Worley PF, Dejanovic B. The neuronal pentraxin Nptx2 regulates complement activity and restrains microglia-mediated synapse loss in neurodegeneration. *Sci Transl Med* [Internet]. 2023 Mar 29;15(689):eadf0141. Available from: <http://dx.doi.org/10.1126/scitranslmed.adf0141> PMID: PMC10467038
64. Pulkkinen V, Ezer S, Sundman L, Hagström J, Remes S, Söderhäll C, Greco D, Haglund C, Kere J, Arola J. Neuropeptide S receptor 1 (NPSR1) activates cancer-related pathways and is widely expressed in neuroendocrine tumors. *Virchows Arch* [Internet]. Springer Science and Business Media LLC; 2014 Aug;465(2):173–183. Available from: <http://dx.doi.org/10.1007/s00428-014-1602-x> PMID: PMC4116602
65. Xie W, Cappiello M, Meng M, Rosenthal R, Zhang W. ADRA2B deletion variant and enhanced cognitive processing of emotional information: A meta-analytical review. *Neurosci Biobehav Rev* [Internet]. Elsevier BV; 2018 Sep;92:402–416. Available from: <http://dx.doi.org/10.1016/j.neubiorev.2018.05.010>
66. Hohoff C, Kroll T, Zhao B, Kerkenberg N, Lang I, Schwarte K, Elmenhorst D, Elmenhorst EM, Aeschbach D, Zhang W, Baune BT, Neumaier B, Bauer A, Deckert J. ADORA2A variation and adenosine A1 receptor availability in the human brain with a focus on anxiety-related brain regions: modulation by ADORA1 variation. *Transl Psychiatry* [Internet]. Springer Science and Business Media LLC; 2020 Nov 24;10(1):406. Available from: <http://dx.doi.org/10.1038/s41398-020-01085-w> PMID: PMC7686488
67. Shimizu T, Nakazawa M, Kani S, Bae YK, Shimizu T, Kageyama R, Hibi M. Zinc finger genes *Fezf1* and *Fezf2* control neuronal differentiation by repressing *Hes5* expression in the forebrain. *Development* [Internet]. The Company of Biologists; 2010 Jun;137(11):1875–1885. Available from: <http://dx.doi.org/10.1242/dev.047167> PMID: 20431123
68. Heide M, Zhang Y, Zhou X, Zhao T, Miquelajáuregui A, Varela-Echavarría A, Alvarez-Bolado G. *Lhx5* controls mamillary differentiation in the developing hypothalamus of the

- mouse. *Front Neuroanat* [Internet]. Frontiers Media SA; 2015 Aug 14;9:113. Available from: <http://dx.doi.org/10.3389/fnana.2015.00113> PMID: PMC4536661
69. Zahid Khan M, He L. The emerging role of GPR50 receptor in brain. *Biomed Pharmacother*. 2016;78:121–128.
 70. Szymanowicz O, Drużdż A, Słowikowski B, Pawlak S, Potocka E, Goutor U, Konieczny M, Ciastoń M, Lewandowska A, Jagodziński PP, Kozubski W, Dorszewska J. A review of the CACNA gene family: Its role in neurological disorders. *Diseases* [Internet]. MDPI AG; 2024 May 5;12(5):90. Available from: <http://dx.doi.org/10.3390/diseases12050090> PMID: PMC11119137
 71. Wu HJ, Bondada S. Positive and negative roles of CD72 in B cell function. *Immunol Res*. 2002;25:155–166.
 72. Scherer SW, Soder S, Duvoisin RM, Huizenga JJ, Tsui LC. The human metabotropic glutamate receptor 8 (GRM8) gene: a disproportionately large gene located at 7q31.3-q32.1. *Genomics*. 1997;44(2):232–236.
 73. Andrews GK. Cellular zinc sensors: MTF-1 regulation of gene expression. *Biometals* [Internet]. 2001 Sep;14(3-4):223–237. Available from: <http://dx.doi.org/10.1023/a:1012932712483> PMID: 11831458
 74. Wang P, Yang L, Guo Y, Qi S, Liang J, Tian G, Tian Z. SENP3 mediates the activation of the Wnt/ β -catenin signaling pathway to accelerate the growth and metastasis of oesophageal squamous cell carcinoma in mice. *Funct Integr Genomics* [Internet]. Springer Science and Business Media LLC; 2024 Feb 22;24(2):40. Available from: <http://dx.doi.org/10.1007/s10142-024-01321-2> PMID: 38383667
 75. Ryu MS, Lichten LA, Liuzzi JP, Cousins RJ. Zinc transporters ZnT1 (Slc30a1), Zip8 (Slc39a8), and Zip10 (Slc39a10) in mouse red blood cells are differentially regulated during erythroid development and by dietary zinc deficiency. *J Nutr* [Internet]. Elsevier BV; 2008 Nov;138(11):2076–2083. Available from: <http://dx.doi.org/10.3945/jn.108.093575> PMID: PMC2635520
 76. Casares L, García V, Garrido-Rodríguez M, Millán E, Collado JA, García-Martín A, Peñarando J, Calzado MA, de la Vega L, Muñoz E. Cannabidiol induces antioxidant pathways in keratinocytes by targeting BACH1. *Redox Biol* [Internet]. Elsevier BV; 2020 Jan;28(101321):101321. Available from: <http://dx.doi.org/10.1016/j.redox.2019.101321> PMID: PMC6742916
 77. Ren X, Feng C, Wang Y, Chen P, Wang S, Wang J, Cao H, Li Y, Ji M, Hou P. SLC39A10 promotes malignant phenotypes of gastric cancer cells by activating the CK2-mediated MAPK/ERK and PI3K/AKT pathways. *Exp Mol Med* [Internet]. Springer Science and Business Media LLC; 2023 Aug;55(8):1757–1769. Available from: <http://dx.doi.org/10.1038/s12276-023-01062-5> PMID: PMC10474099
 78. Schulkens IA, Castricum KCM, Weijers EM, Koolwijk P, Griffioen AW, Thijssen VL.

- Expression, regulation and function of human metallothioneins in endothelial cells. *J Vasc Res* [Internet]. S. Karger AG; 2014 Aug 9;51(3):231–238. Available from: <http://dx.doi.org/10.1159/000365550> PMID: 25116857
79. Chen YF, Wei YY, Yang CC, Liu CJ, Yeh LY, Chou CH, Chang KW, Lin SC. miR-125b suppresses oral oncogenicity by targeting the anti-oxidative gene PRXL2A. *Redox Biol* [Internet]. Elsevier BV; 2019 Apr;22(101140):101140. Available from: <http://dx.doi.org/10.1016/j.redox.2019.101140> PMID: PMC6383183
 80. Tan Y, Xia H, Song Q. Research mapping of cannabinoids and endocannabinoid system in cancer over the past three decades: insights from bibliometric analysis. *Front Pharmacol* [Internet]. Frontiers Media SA; 2025 Apr 2;16:1540619. Available from: <http://dx.doi.org/10.3389/fphar.2025.1540619> PMID: PMC12000044
 81. Wu J. Cannabis, cannabinoid receptors, and endocannabinoid system: yesterday, today, and tomorrow. *Acta Pharmacol Sin* [Internet]. Springer Science and Business Media LLC; 2019 Mar;40(3):297–299. Available from: <http://dx.doi.org/10.1038/s41401-019-0210-3> PMID: PMC6460362
 82. Piñero J, Bravo À, Queralt-Rosinach N, Gutiérrez-Sacristán A, Deu-Pons J, Centeno E, García-García J, Sanz F, Furlong LI. DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants. *Nucleic Acids Res* [Internet]. Oxford Academic; 2017 Jan 4;45(D1):D833–D839. Available from: <https://dx.doi.org/10.1093/nar/gkw943> PMID: PMC5210640
 83. Piñero J, Ramírez-Angueta JM, Saüch-Pitarch J, Ronzano F, Centeno E, Sanz F, Furlong LI. The DisGeNET knowledge platform for disease genomics: 2019 update. *Nucleic Acids Res* [Internet]. Oxford University Press (OUP); 2020 Jan 8;48(D1):D845–D855. Available from: <http://dx.doi.org/10.1093/nar/gkz1021> PMID: PMC7145631
 84. Lowe H, Toyang N, Steele B, Bryant J, Ngwa W. The Endocannabinoid System: A potential target for the treatment of various diseases. *Int J Mol Sci* [Internet]. MDPI AG; 2021 Aug 31;22(17):9472. Available from: <http://dx.doi.org/10.3390/ijms22179472> PMID: PMC8430969
 85. Massoumi R. CYLD: a deubiquitination enzyme with multiple roles in cancer. *Future Oncol* [Internet]. Informa UK Limited; 2011 Feb;7(2):285–297. Available from: <http://dx.doi.org/10.2217/fon.10.187> PMID: 21345146
 86. Negroni A, Pierdomenico M, Cucchiara S, Stronati L. NOD2 and inflammation: current insights. *J Inflamm Res* [Internet]. Dove Medical Press Ltd.; 2018 Feb;11:49–60. Available from: <http://dx.doi.org/10.2147/jir.s137606>
 87. Elding H, Lau W, Swallow DM, Maniatis N. Dissecting the genetics of complex inheritance: linkage disequilibrium mapping provides insight into Crohn disease. *Am J Hum Genet* [Internet]. Elsevier BV; 2011 Dec 9;89(6):798–805. Available from: <http://dx.doi.org/10.1016/j.ajhg.2011.11.006> PMID: PMC3234369

88. Del Bosque-Plata L, Hernández-Cortés EP, Gragnoli C. The broad pathogenetic role of TCF7L2 in human diseases beyond type 2 diabetes. *J Cell Physiol* [Internet]. Wiley; 2022 Jan;237(1):301–312. Available from: <http://dx.doi.org/10.1002/jcp.30581> PMID: PMC9292842
89. Weedon MN. The importance of TCF7L2. *Diabet Med* [Internet]. Wiley; 2007 Oct;24(10):1062–1066. Available from: <https://doi.org/10.1111/j.1464-5491.2007.02258.x> PMID: 17888129
90. Zivotić I, Djurić T, Stanković A, Milasinovic D, Stankovic G, Dekleva M. CDKN2B gene expression is affected by 9p21. 3 rs10757278 in CAD patients, six months after the MI. *Clin Biochem*. 2019;73:70–76.
91. Martin BR. Cellular effects of cannabinoids. *Pharmacol Rev* [Internet]. 1986 Mar;38(1):45–74. Available from: <https://www.ncbi.nlm.nih.gov/pubmed/2872689> PMID: 2872689
92. Al-Khazaleh AK, Zhou X, Bhuyan DJ, Münch GW, Al-Dalabeeh EA, Jaye K, Chang D. The neurotherapeutic arsenal in cannabis sativa: Insights into anti-neuroinflammatory and neuroprotective activity and potential entourage effects. *Molecules* [Internet]. 2024 Jan 15;29(2). Available from: <http://dx.doi.org/10.3390/molecules29020410> PMID: PMC10821245
93. Atalay S, Jarocka-Karpowicz I, Skrzydlewska E. Antioxidative and anti-inflammatory properties of cannabidiol. *Antioxidants (Basel)* [Internet]. MDPI AG; 2019 Dec 25;9(1):21. Available from: <http://dx.doi.org/10.3390/antiox9010021> PMID: PMC7023045
94. Ekiner SA, Gęgotek A, Skrzydlewska E. The molecular activity of cannabidiol in the regulation of Nrf2 system interacting with NF-κB pathway under oxidative stress. *Redox Biol*. 2022;57.
95. Baswan SM, Klosner AE, Glynn K, Rajgopal A, Malik K, Yim S, Stern N. Therapeutic potential of cannabidiol (CBD) for skin health and disorders. *Clin Cosmet Investig Dermatol* [Internet]. Informa UK Limited; 2020 Dec 8;13:927–942. Available from: <http://dx.doi.org/10.2147/CCID.S286411> PMID: PMC7736837
96. Juknat A, Rimmerman N, Levy R, Vogel Z, Kozela E. Cannabidiol affects the expression of genes involved in zinc homeostasis in BV-2 microglial cells. *Neurochem Int* [Internet]. Elsevier BV; 2012 Nov;61(6):923–930. Available from: <http://dx.doi.org/10.1016/j.neuint.2011.12.002> PMID: 22178458
97. Kozela E, Juknat A, Gao F, Kaushansky N, Coppola G, Vogel Z. Pathways and gene networks mediating the regulatory effects of cannabidiol, a nonpsychoactive cannabinoid, in autoimmune T cells. *J Neuroinflammation* [Internet]. Springer Science and Business Media LLC; 2016 Dec;13(1). Available from: <http://dx.doi.org/10.1186/s12974-016-0603-x>
98. Pagano C, Navarra G, Coppola L, Bifulco M, Laezza C. Molecular mechanism of cannabinoids in cancer progression. *Int J Mol Sci* [Internet]. MDPI AG; 2021 Apr 1;22(7):3680. Available from: <http://dx.doi.org/10.3390/ijms22073680> PMID:

99. Nagarajan R, Svaren J, Le N, Araki T, Watson M, Milbrandt J. EGR2 mutations in inherited neuropathies dominant-negatively inhibit myelin gene expression. *Neuron* [Internet]. Elsevier BV; 2001 May;30(2):355–368. Available from: [http://dx.doi.org/10.1016/s0896-6273\(01\)00282-3](http://dx.doi.org/10.1016/s0896-6273(01)00282-3) PMID: 11394999
100. Pavinato L, Stanic J, Barzasi M, Gurgone A, Chiantia G, Cipriani V, Eberini I, Palazzolo L, Di Luca M, Costa A, Marcantoni A, Biamino E, Spada M, Hiatt SM, Kelley WV, Vestito L, Sisodiya SM, Genomics England Research Consortium, Efthymiou S, Chand P, Kaiyrzhanov R, Bruxelles A, Cardaropoli S, Tartaglia M, De Rubeis S, Buxbaum JD, Smedley D, Ferrero GB, Giustetto M, Gardoni F, Brusco A. Missense variants in RPH3A cause defects in excitatory synaptic function and are associated with a clinically variable neurodevelopmental disorder. *Genet Med* [Internet]. Elsevier BV; 2023 Nov;25(11):100922. Available from: <http://dx.doi.org/10.1016/j.gim.2023.100922> PMID: 37403762
101. Wei AD, Wakenight P, Zwingman TA, Bard AM, Sahai N, Willemsen MH, Schelhaas HJ, Stegmann APA, Verhoeven JS, de Man SA, Wessels MW, Kleefstra T, Shinde DN, Helbig KL, Basinger A, Wagner VF, Rodriguez-Buritica D, Bryant E, Millichap JJ, Millen KJ, Dobyns WB, Ramirez JM, Kalume FK. Human KCNQ5 de novo mutations underlie epilepsy and intellectual disability. *J Neurophysiol* [Internet]. American Physiological Society; 2022 Jul 1;128(1):40–61. Available from: <http://dx.doi.org/10.1152/jn.00509.2021> PMCID: PMC9236882
102. Paulraj P, Bosworth M, Longhurst M, Hornbuckle C, Gotway G, Lamb AN, Andersen EF. A novel homozygous deletion within the FRY gene associated with nonsyndromic developmental delay. *Cytogenet Genome Res* [Internet]. S. Karger AG; 2019 Sep 6;159(1):19–25. Available from: <http://dx.doi.org/10.1159/000502598> PMID: 31487712
103. Lim ASP, Chang AM, Shulman JM, Raj T, Chibnik LB, Cain SW, Rothamel K, Benoist C, Myers AJ, Czeisler CA, Buchman AS, Bennett DA, Duffy JF, Saper CB, De Jager PL. A common polymorphism near PER1 and the timing of human behavioral rhythms. *Ann Neurol* [Internet]. Wiley; 2012 Sep;72(3):324–334. Available from: <http://dx.doi.org/10.1002/ana.23636> PMCID: PMC3464954
104. Chayasirisobhon S. Mechanisms of action and pharmacokinetics of cannabis. *Perm J* [Internet]. The Permanente Press; 2020 Dec;25(5):1–3. Available from: <http://dx.doi.org/10.7812/TPP/19.200> PMCID: PMC8803256
105. Huestis MA. Human cannabinoid pharmacokinetics. *Chem Biodivers* [Internet]. Chem Biodivers; 2007 Aug;4(8):1770–1804. Available from: <http://dx.doi.org/10.1002/cbdv.200790152> PMCID: PMC2689518
106. Reimer RJ. SLC17: a functionally diverse family of organic anion transporters. *Mol Aspects Med* [Internet]. Elsevier BV; 2013 Apr;34(2-3):350–359. Available from: <http://dx.doi.org/10.1016/j.mam.2012.05.004> PMCID: PMC3927456
107. Nedelescu H, Wagner GE, De Ness GL, Carroll A, Kerr TM, Wang J, Zhang S, Chang S,

- Than AH, Emerson NE, Suto N, Weiss F. Cannabidiol produces distinct U-shaped dose-response effects on cocaine-induced conditioned place preference and associated recruitment of prelimbic neurons in male rats. *Biol Psychiatry Glob Open Sci* [Internet]. Elsevier BV; 2022 Jan;2(1):70–78. Available from: <http://dx.doi.org/10.1016/j.bpsgos.2021.06.014> PMID: PMC8896771
108. Trainito A, Muscarà C, Gugliandolo A, Chiricosta L, Salamone S, Pollastro F, Mazzon E, D'Angiolini S. Cannabinol (CBN) influences the ion channels and synaptic-related genes in NSC-34 cell line: A transcriptomic study. *Cells* [Internet]. 2024 Sep 19;13(18). Available from: <http://dx.doi.org/10.3390/cells13181573> PMID: PMC11430194
 109. Fang CY, Lai TC, Hsiao M, Chang YC. The diverse roles of TAO kinases in health and diseases. *Int J Mol Sci* [Internet]. MDPI AG; 2020 Oct 10;21(20):7463. Available from: <http://dx.doi.org/10.3390/ijms21207463> PMID: PMC7589832
 110. Simon L, Song K, Vande Stouwe C, Hollenbach A, Amedee A, Mohan M, Winsauer P, Molina P. $\Delta 9$ -tetrahydrocannabinol ($\Delta 9$ -THC) promotes neuroimmune-modulatory MicroRNA profile in striatum of simian immunodeficiency virus (SIV)-infected macaques. *J Neuroimmune Pharmacol* [Internet]. Springer Science and Business Media LLC; 2016 Mar;11(1):192–213. Available from: <http://dx.doi.org/10.1007/s11481-015-9645-6> PMID: PMC4773048
 111. Zou S, Kumar U. Cannabinoid receptors and the endocannabinoid system: Signaling and function in the central nervous system. *Int J Mol Sci* [Internet]. MDPI AG; 2018 Mar 13;19(3):833. Available from: <http://dx.doi.org/10.3390/ijms19030833>
 112. Shahbazi F, Grandi V, Banerjee A, Trant JF. Cannabinoids and Cannabinoid Receptors: The Story so Far. *iScience* [Internet]. Elsevier BV; 2020 Jul 24;23(7):101301. Available from: <http://dx.doi.org/10.1016/j.isci.2020.101301> PMID: PMC7339067
 113. Schönke M, Martinez-Tellez B, Rensen PC. Role of the endocannabinoid system in the regulation of the skeletal muscle response to exercise. *Curr Opin Pharmacol* [Internet]. Elsevier BV; 2020 Jun;52:52–60. Available from: <http://dx.doi.org/10.1016/j.coph.2020.05.003> PMID: 32619926
 114. Wiciński M, Fajkiel-Madajczyk A, Kurant Z, Gryczka K, Kurant D, Szambelan M, Malinowski B, Falkowski M, Zabrzyński J, Słupski M. The use of cannabidiol in metabolic syndrome-an opportunity to improve the patient's health or much ado about nothing? *J Clin Med* [Internet]. 2023 Jul 11;12(14). Available from: <http://dx.doi.org/10.3390/jcm12144620> PMID: PMC10380672
 115. Reis MG, Ferreira AJF, Sohoulí MH, Taimeirão DR, Vieira RAL, Guimarães NS. Effect of cannabis and subproducts on anthropometric measures: a systematic review and meta-analysis. *Int J Obes (Lond)* [Internet]. Springer Science and Business Media LLC; 2024 Jan;48(1):44–54. Available from: <http://dx.doi.org/10.1038/s41366-023-01399-x> PMID: 37935909
 116. Li X, Diviant JP, Stith SS, Brockelman F, Keeling K, Hall B, Vigil JM. The effectiveness

- of cannabis flower for immediate relief from symptoms of depression. *Yale J Biol Med* [Internet]. 2020 Jun;93(2):251–264. Available from: <https://www.ncbi.nlm.nih.gov/pubmed/32607086> PMID: PMC7309674
117. Hindley G, Beck K, Borgan F, Ginestet CE, McCutcheon R, Kleinloog D, Ganesh S, Radhakrishnan R, D'Souza DC, Howes OD. Psychiatric symptoms caused by cannabis constituents: a systematic review and meta-analysis. *Lancet Psychiatry* [Internet]. Elsevier BV; 2020 Apr;7(4):344–353. Available from: [http://dx.doi.org/10.1016/S2215-0366\(20\)30074-2](http://dx.doi.org/10.1016/S2215-0366(20)30074-2) PMID: PMC7738353
 118. Lucatch AM, Kloiber SM, Meyer JH, Rizvi SJ, George TP. Effects of extended cannabis abstinence in major depressive disorder. *Can J Addict* [Internet]. Ovid Technologies (Wolters Kluwer Health); 2020 Sep;11(3):33–41. Available from: <http://dx.doi.org/10.1097/cxa.0000000000000090>
 119. Clendennen SL, Smith J, Sumbe A, Chen B, Wilkinson AV, Harrell MB. Symptoms of depression and anxiety and subsequent use of nicotine and THC in electronic cigarettes. *Subst Use Misuse* [Internet]. 2023 Feb 22;58(5):591–600. Available from: <http://dx.doi.org/10.1080/10826084.2023.2177110> PMID: PMC10155290
 120. Jadoon KA, Ratcliffe SH, Barrett DA, Thomas EL, Stott C, Bell JD. Efficacy and Safety of Cannabidiol and Tetrahydrocannabivarin on Glycemic and Lipid Parameters in Patients With Type 2 Diabetes: A Randomized, Double-Blind, Placebo-Controlled, Parallel Group Pilot Study. *Diabetes Care*. 2016;39:1777–1786.
 121. McGuire P, Robson P, Cubala WJ, Vasile D, Morrison PD, Barron R, Taylor A, Wright S. Cannabidiol (CBD) as an adjunctive therapy in schizophrenia: A multicenter randomized controlled trial. *Am J Psychiatry* [Internet]. American Psychiatric Association Publishing; 2018 Mar 1;175(3):225–231. Available from: <http://dx.doi.org/10.1176/appi.ajp.2017.17030325> PMID: 29241357
 122. Leweke FM, Piomelli D, Pahlisch F, Muhl D, Gerth CW, Hoyer C, Klosterkötter J, Hellmich M, Koethe D. Cannabidiol enhances anandamide signaling and alleviates psychotic symptoms of schizophrenia. *Transl Psychiatry* [Internet]. Springer Science and Business Media LLC; 2012 Mar 20;2(3):e94. Available from: <http://dx.doi.org/10.1038/tp.2012.15> PMID: PMC3316151
 123. Casey W, Jacobs A, Maull E, Matheson J, Clarke C, Lowit A. A new path forward: The Interagency Coordinating Committee on the validation of Alternative Methods (ICCVAM) and National Toxicology Program's Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM). *J Am Assoc Lab Anim Sci* [Internet]. 2015 Mar;54(2):170–173. Available from: <https://www.ncbi.nlm.nih.gov/pubmed/25836963> PMID: PMC4382621
 124. Corvi R, Ahr HJ, Albertini S, Blakey DH, Clerici L, Coecke S, Douglas GR, Gribaldo L, Groten JP, Haase B, Hamernik K, Hartung T, Inoue T, Indans I, Maurici D, Orphanides G, Rembges D, Sansone SA, Snape JR, Toda E, Tong W, van Delft JH, Weis B, Schechtman LM. Meeting report: Validation of toxicogenomics-based test systems: ECVAM-

- ICCVAM/NICEATM considerations for regulatory use. *Environ Health Perspect* [Internet]. 2006 Mar;114(3):420–429. Available from: <http://dx.doi.org/10.1289/ehp.8247> PMID: PMC1392237
125. Heijne WHM, Kienhuis AS, van Ommen B, Stierum RH, Groten JP. Systems toxicology: applications of toxicogenomics, transcriptomics, proteomics and metabolomics in toxicology. *Expert Rev Proteomics* [Internet]. Informa UK Limited; 2005 Oct;2(5):767–780. Available from: <http://dx.doi.org/10.1586/14789450.2.5.767> PMID: 16209655
 126. Joseph P, Umbright C, Sellamuthu R. Blood transcriptomics: applications in toxicology. *J Appl Toxicol* [Internet]. 2013 Nov;33(11):1193–1202. Available from: <http://dx.doi.org/10.1002/jat.2861> PMID: PMC4550215
 127. Cui Y, Paules RS. Use of transcriptomics in understanding mechanisms of drug-induced toxicity. *Pharmacogenomics* [Internet]. 2010 Apr;11(4):573–585. Available from: <http://dx.doi.org/10.2217/pgs.10.37> PMID: PMC2884389
 128. Kinaret PAS, Serra A, Federico A, Kohonen P, Nymark P, Liampa I, Ha MK, Choi JS, Jagiello K, Sanabria N, Melagraki G, Cattelani L, Fratello M, Sarimveis H, Afantitis A, Yoon TH, Gulumian M, Grafström R, Puzyn T, Greco D. Transcriptomics in toxicogenomics, part I: Experimental Design, Technologies, Publicly Available Data, and Regulatory Aspects. *Nanomaterials (Basel)* [Internet]. MDPI AG; 2020 Apr 15;10(4):750. Available from: <http://dx.doi.org/10.3390/nano10040750> PMID: PMC7221878
 129. Davis AP, Murphy CG, Saraceni-Richards CA, Rosenstein MC, Wiegers TC, Mattingly CJ. Comparative Toxicogenomics Database: a knowledgebase and discovery tool for chemical-gene-disease networks. *Nucleic Acids Res* [Internet]. Oxford University Press (OUP); 2009 Jan;37(Database issue):D786–92. Available from: <http://dx.doi.org/10.1093/nar/gkn580> PMID: PMC2686584
 130. Clough E, Barrett T. The Gene Expression Omnibus database. *Methods Mol Biol* [Internet]. 2016;1418:93–110. Available from: http://dx.doi.org/10.1007/978-1-4939-3578-9_5 PMID: PMC4944384
 131. Gremse M, Chang A, Schomburg I, Grote A, Scheer M, Ebeling C, Schomburg D. The BRENDA Tissue Ontology (BTO): the first all-integrating ontology of all organisms for enzyme sources. *Nucleic Acids Res* [Internet]. Oxford University Press (OUP); 2011 Jan;39(Database issue):D507–13. Available from: <https://dx.doi.org/10.1093/nar/gkq968> PMID: PMC3013802
 132. Leinonen R, Sugawara H, Shumway M, International Nucleotide Sequence Database Collaboration. The sequence read archive. *Nucleic Acids Res* [Internet]. Oxford Academic; 2011 Jan;39(Database issue):D19–21. Available from: <https://dx.doi.org/10.1093/nar/gkq1019> PMID: PMC3013647
 133. Valieris R. Parallel-fastq-dump. GitHub. 2020;
 134. Inc. A. Anaconda software distribution. Computer software Anaconda Documentation.

2020;

135. Homo sapiens genome assembly GRCh38 [Internet]. NCBI. [cited 2025 Aug 14]. Available from: https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_000001405.26/.
136. Mus musculus genome assembly GRCm39 [Internet]. NCBI. [cited 2025 Aug 14]. Available from: https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_000001635.27/.
137. Rattus norvegicus genome assembly Rnor_6.0 [Internet]. NCBI. [cited 2025 Aug 14]. Available from: https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_000001895.5/
138. Macaca mulatta genome assembly Mmul_10 [Internet]. NCBI. [cited 2025 Aug 14]. Available from: https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_003339765.1/
139. Schurch NJ, Schofield P, Gierliński M, Cole C, Sherstnev A, Singh V, Wrobel N, Gharbi K, Simpson GG, Owen-Hughes T, Blaxter M, Barton GJ. How many biological replicates are needed in an RNA-seq experiment and which differential expression tool should you use? RNA [Internet]. 2016 Jun;22(6):839–851. Available from: <http://dx.doi.org/10.1261/rna.053959.115> PMID: PMC4878611
140. Kanehisa M. The KEGG Database. “In Silico” Simulation of Biological Processes [Internet]. Chichester, UK: John Wiley & Sons, Ltd; 2003. p. 91–103. Available from: <http://dx.doi.org/10.1002/0470857897.ch8>
141. Harris MA, Clark J, Ireland A, Lomax J, Ashburner M, Foulger R, Eilbeck K, Lewis S, Marshall B, Mungall C, Richter J, Rubin GM, Blake JA, Bult C, Dolan M, Drabkin H, Eppig JT, Hill DP, Ni L, Ringwald M, Balakrishnan R, Cherry JM, Christie KR, Costanzo MC, Dwight SS, Engel S, Fisk DG, Hirschman JE, Hong EL, Nash RS, Sethuraman A, Theesfeld CL, Botstein D, Dolinski K, Feierbach B, Berardini T, Mundodi S, Rhee SY, Apweiler R, Barrell D, Camon E, Dimmer E, Lee V, Chisholm R, Gaudet P, Kibbe W, Kishore R, Schwarz EM, Sternberg P, Gwinn M, Hannick L, Wortman J, Berriman M, Wood V, de la Cruz N, Tonellato P, Jaiswal P, Seigfried T, White R, Gene Ontology Consortium. The Gene Ontology (GO) database and informatics resource. Nucleic Acids Res [Internet]. Oxford University Press (OUP); 2004 Jan 1;32(Database issue):D258–61. Available from: <http://dx.doi.org/10.1093/nar/gkh036> PMID: PMC308770
142. Tweedie S, Braschi B, Gray K, Jones TEM, Seal RL, Yates B, Bruford EA. Genenames.Org: The HGNC and VGNC resources in 2021. Nucleic Acids Res [Internet]. Oxford University Press (OUP); 2021 Jan 8;49(D1):D939–D946. Available from: <https://dx.doi.org/10.1093/nar/gkaa980> PMID: PMC7779007
143. Wright MW, Eyre TA, Lush MJ, Povey S, Bruford EA. HCOP: the HGNC comparison of orthology predictions search tool. Mamm Genome [Internet]. Springer Science and Business Media LLC; 2005 Nov;16(11):827–828. Available from: <http://dx.doi.org/10.1007/s00335-005-0103-2> PMID: 16284797
144. Gu Z. Complex heatmap visualization. Imeta [Internet]. Wiley; 2022 Sep;1(3):e43. Available from: <http://dx.doi.org/10.1002/imt2.43> PMID: PMC10989952

145. Kurt Z, Cheng J, Barrere-Cain R, McQuillen CN, Saleem Z, Hsu N, Jiang N, Pan C, Franzén O, Koplev S, Wang S, Björkegren J, Lusi AJ, Blencowe M, Yang X. Shared and distinct pathways and networks genetically linked to coronary artery disease between human and mouse. *Elife* [Internet]. Elife; 2023 Dec 7;12. Available from: <http://dx.doi.org/10.7554/eLife.88266> PMID: PMC10703441
146. Cheng F, Desai RJ, Handy DE, Wang R, Schneeweiss S, Barabási AL, Loscalzo J. Network-based approach to prediction and population-based validation of in silico drug repurposing. *Nat Commun* [Internet]. Nat Commun; 2018 Jul 12;9(1):2691. Available from: <http://dx.doi.org/10.1038/s41467-018-05116-5> PMID: PMC6043492
147. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* [Internet]. Springer Science and Business Media LLC; 2008 Jul;5(7):621–628. Available from: <http://dx.doi.org/10.1038/nmeth.1226> PMID: 18516045
148. Janousek RM, Mayer J, Knepper TP. Is the phase-out of long-chain PFASs measurable as fingerprint in a defined area? Comparison of global PFAS concentrations and a monitoring study performed in Hesse, Germany from 2014 to 2018. *Trends Analyt Chem* [Internet]. Elsevier BV; 2019 Nov;120(115393):115393. Available from: <http://dx.doi.org/10.1016/j.trac.2019.01.017>
149. Sicairos Fernandez G. Proactive approaches towards the phase-out of the PFAS family: The PFAS Movement status and a multi-national automotive enterprise towards hazardous chemicals phase-out. IIIIEE Master Thesis [Internet]. 2024 [cited 2025 Aug 15]; Available from: <https://lup.lub.lu.se/student-papers/search/publication/9171566>
150. Bocard Colome A. Navigating the PFAS Phase-Out in the Outdoor Apparel Industry [Internet]. 2025 [cited 2025 Aug 15]. Available from: <https://studenttheses.uu.nl/handle/20.500.12932/49528>
151. Sunderland EM, Hu XC, Dassuncao C, Tokranov AK, Wagner CC, Allen JG. A review of the pathways of human exposure to poly- and perfluoroalkyl substances (PFASs) and present understanding of health effects. *J Expo Sci Environ Epidemiol* [Internet]. Springer Science and Business Media LLC; 2019 Mar;29(2):131–147. Available from: <http://dx.doi.org/10.1038/s41370-018-0094-1> PMID: PMC6380916
152. National Academies of Sciences, Engineering, Medicine, Health, Medicine Division, Division on Earth, Life Studies, Board on Population Health, Public Health Practice, Board on Environmental Studies, Toxicology, Committee on the Guidance on PFAS Testing, Health Outcomes. Potential Health Effects of PFAS [Internet]. National Academies Press (US); 2022 [cited 2025 Aug 15]. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK584690/>
153. Guo W, Hao W, Xiao W. Emerging perfluorinated chemical GenX: Environmental and biological fates and risks. *Environ Health (Wash)* [Internet]. 2025 Apr 18;3(4):338–351. Available from: <http://dx.doi.org/10.1021/envhealth.4c00164> PMID: PMC12012656
154. Glüge J, Scheringer M, Cousins IT, DeWitt JC, Goldenman G, Herzke D, Lohmann R, Ng

- CA, Trier X, Wang Z. An overview of the uses of per- and polyfluoroalkyl substances (PFAS). *Environ Sci Process Impacts* [Internet]. Royal Society of Chemistry (RSC); 2020 Dec 1;22(12):2345–2373. Available from: <http://dx.doi.org/10.1039/D0EM00291G> PMID: PMC7784712
155. Fenton SE, Ducatman A, Boobis A, DeWitt JC, Lau C, Ng C, Smith JS, Roberts SM. Per- and Polyfluoroalkyl Substance Toxicity and Human Health Review: Current State of Knowledge and Strategies for Informing Future Research. *Environ Toxicol Chem* [Internet]. 2021 Mar;40(3):606–630. Available from: <http://dx.doi.org/10.1002/etc.4890> PMID: PMC7906952
 156. Olsen GW, Burris JM, Ehresman DJ, Froehlich JW, Seacat AM, Butenhoff JL, Zobel LR. Half-life of serum elimination of perfluorooctanesulfonate, perfluorohexanesulfonate, and perfluorooctanoate in retired fluorocarbon production workers. *Environ Health Perspect* [Internet]. 2007 Sep;115(9):1298–1305. Available from: <http://dx.doi.org/10.1289/ehp.10009> PMID: PMC1964923
 157. Buttle E, Sharp EL, Fisher K. Managing ubiquitous “forever chemicals”: More-than-human possibilities for the problem of PFAS. *N Z Geog. Wiley*; 2023;79(2):97–106.
 158. Stanifer JW, Stapleton HM, Souma T, Wittmer A, Zhao X, Boulware LE. Perfluorinated chemicals as emerging environmental threats to kidney health: A scoping review: A scoping review. *Clin J Am Soc Nephrol* [Internet]. Ovid Technologies (Wolters Kluwer Health); 2018 Oct 8;13(10):1479–1492. Available from: https://journals.lww.com/cjasn/fulltext/2018/10000/perfluorinated_chemicals_as_emerging_environmental.8.aspx PMID: PMC6218824
 159. Wang Z, Cousins IT, Scheringer M, Hungerbühler K. Fluorinated alternatives to long-chain perfluoroalkyl carboxylic acids (PFCAs), perfluoroalkane sulfonic acids (PFASs) and their potential precursors. *Environ Int* [Internet]. 2013 Oct;60:242–248. Available from: <https://doi.org/10.1016/j.envint.2013.08.021> PMID: 24660230
 160. Evans N, Conley JM, Cardon M, Hartig P, Medlock-Kakaley E, Gray LE Jr. In vitro activity of a panel of per- and polyfluoroalkyl substances (PFAS), fatty acids, and pharmaceuticals in peroxisome proliferator-activated receptor (PPAR) alpha, PPAR gamma, and estrogen receptor assays. *Toxicol Appl Pharmacol* [Internet]. Elsevier BV; 2022 Aug 15;449(116136):116136. Available from: <http://dx.doi.org/10.1016/j.taap.2022.116136> PMID: PMC9341220
 161. Yang W, Ling X, He S, Cui H, Yang Z, An H, Wang L, Zou P, Chen Q, Liu J, Ao L, Cao J. PPAR α /ACOX1 as a novel target for hepatic lipid metabolism disorders induced by per- and polyfluoroalkyl substances: An integrated approach. *Environ Int* [Internet]. Elsevier BV; 2023 Aug;178(108138):108138. Available from: <http://dx.doi.org/10.1016/j.envint.2023.108138> PMID: 37572494
 162. Margolis R, Sant KE. Associations between exposures to perfluoroalkyl substances and diabetes, hyperglycemia, or insulin resistance: A scoping review. *J Xenobiot* [Internet]. MDPI AG; 2021 Sep 14;11(3):115–129. Available from: <http://dx.doi.org/10.3390/jox11030008>

163. Jia P, Yu X, Jin Y, Wang X, Yang A, Zhang L, Jing X, Kang W, Zhao G, Gao B. Relationship between per-fluoroalkyl and polyfluoroalkyl substance exposure and insulin resistance in nondiabetic adults: Evidence from NHANES 2003-2018. *Ecotoxicol Environ Saf* [Internet]. Elsevier BV; 2024 Nov 15;287(117260):117260. Available from: <http://dx.doi.org/10.1016/j.ecoenv.2024.117260> PMID: 39504878
164. Dagar M, Kumari P, Mirza AMW, Singh S, Ain NU, Munir Z, Javed T, Virk MFI, Javed S, Qizilbash FH, Kc A, Ekhtator C, Bellegarde SB. The hidden threat: Endocrine disruptors and their impact on insulin resistance. *Cureus* [Internet]. Springer Science and Business Media LLC; 2023 Oct;15(10):e47282. Available from: https://assets.cureus.com/uploads/review_article/pdf/190793/20231117-10825-r8a398.pdf PMCID: PMC10656111
165. Zhang Y, Gan B, Liu D, Paik JH. FoxO family members in cancer. *Cancer Biol Ther* [Internet]. Informa UK Limited; 2011 Aug 15;12(4):253–259. Available from: <http://dx.doi.org/10.4161/cbt.12.4.15954> PMID: 21613825
166. Farhan M, Wang H, Gaur U, Little PJ, Xu J, Zheng W. FOXO signaling pathways as therapeutic targets in cancer. *Int J Biol Sci* [Internet]. 2017 Jul 6;13(7):815–827. Available from: <http://dx.doi.org/10.7150/ijbs.20052> PMCID: PMC5555100
167. Zhang X, Tang N, Hadden TJ, Rishi AK. Akt, FoxO and regulation of apoptosis. *Biochim Biophys Acta* [Internet]. Elsevier BV; 2011 Nov;1813(11):1978–1986. Available from: <http://dx.doi.org/10.1016/j.bbamcr.2011.03.010> PMID: 21440011
168. Coomans de Brachène A, Demoulin JB. FOXO transcription factors in cancer development and therapy. *Cell Mol Life Sci* [Internet]. Springer Science and Business Media LLC; 2016 Mar;73(6):1159–1172. Available from: <http://dx.doi.org/10.1007/s00018-015-2112-y> PMCID: PMC11108379
169. Chen S, Zhang W, Tang C, Tang X, Liu L, Liu C. Vanin-1 is a key activator for hepatic gluconeogenesis. *Diabetes* [Internet]. American Diabetes Association; 2014 Jun;63(6):2073–2085. Available from: <https://dx.doi.org/10.2337/db13-0788> PMID: 24550194
170. Yu H, Cui Y, Guo F, Zhu Y, Zhang X, Shang D, Dong D, Xiang H. Vanin1 (VNN1) in chronic diseases: Future directions for targeted therapy. *Eur J Pharmacol* [Internet]. 2024 Jan 5;962:176220. Available from: <http://dx.doi.org/10.1016/j.ejphar.2023.176220> PMID: 38042463
171. Rommelaere S, Millet V, Gensollen T, Bourges C, Eeckhoutte J, Hennuyer N, Baugé E, Chasson L, Cacciatore I, Staels B, Pitari G, Galland F, Naquet P. PPAR α regulates the production of serum Vanin-1 by liver. *FEBS Lett* [Internet]. Wiley; 2013 Nov 15;587(22):3742–3748. Available from: <http://dx.doi.org/10.1016/j.febslet.2013.09.046> PMID: 24140347
172. Chen C, Wang H, Chen B, Chen D, Lu C, Li H, Qian Y, Tan Y, Weng H, Cai L. Pex11a

- deficiency causes dyslipidaemia and obesity in mice. *J Cell Mol Med* [Internet]. 2019 Mar;23(3):2020–2031. Available from: <http://dx.doi.org/10.1111/jcmm.14108> PMID: PMC6378206
173. Koch J, Pranjic K, Huber A, Ellinger A, Hartig A, Kragler F, Brocard C. PEX11 family members are membrane elongation factors that coordinate peroxisome proliferation and maintenance. *J Cell Sci* [Internet]. The Company of Biologists; 2010 Oct 1;123(Pt 19):3389–3400. Available from: <https://dx.doi.org/10.1242/jcs.064907> PMID: 20826455
 174. Rodríguez-Serrano M, Romero-Puertas MC, Sanz-Fernández M, Hu J, Sandalio LM. Peroxisomes extend peroxules in a fast response to stress via a reactive oxygen species-mediated induction of the peroxin PEX11a. *Plant Physiol* [Internet]. Oxford Academic; 2016 Jul;171(3):1665–1674. Available from: <https://dx.doi.org/10.1104/pp.16.00648> PMID: PMC4936588
 175. Moffat C, Bhatia L, Nguyen T, Lynch P, Wang M, Wang D, Ilkayeva OR, Han X, Hirschey MD, Claypool SM, Seifert EL. Acyl-CoA thioesterase-2 facilitates mitochondrial fatty acid oxidation in the liver. *J Lipid Res* [Internet]. Elsevier BV; 2014 Dec;55(12):2458–2470. Available from: <http://dx.doi.org/10.1194/jlr.M046961> PMID: PMC4242439
 176. Ellis JM, Bowman CE, Wolfgang MJ. Metabolic and tissue-specific regulation of acyl-CoA metabolism. *PLoS One* [Internet]. Public Library of Science (PLOS); 2015 Mar 11;10(3):e0116587. Available from: <http://dx.doi.org/10.1371/journal.pone.0116587> PMID: PMC4356623
 177. Yin X, Lyu C, Li Z, Wang Q, Ding Y, Wang Y, Qiu Y, Cui S, Guo D, Xu R. High Expression of ACOT2 Predicts Worse Overall Survival and Abnormal Lipid Metabolism: A Potential Target for Acute Myeloid Leukemia. Chen GS, editor. *Journal of Healthcare Engineering* [Internet]. 2022 Sep 23;2022:1–15. Available from: <http://dx.doi.org/10.1155/2022/2669114>
 178. Liu Y, Yin W. CD36 in liver diseases. *Hepatol Commun* [Internet]. Ovid Technologies (Wolters Kluwer Health); 2025 Jan 1;9(1):e0623. Available from: <http://dx.doi.org/10.1097/HC9.0000000000000623> PMID: PMC11717518
 179. Wilson CG, Tran JL, Erion DM, Vera NB, Febbraio M, Weiss EJ. Hepatocyte-specific disruption of CD36 attenuates fatty liver and improves insulin sensitivity in HFD-fed mice. *Endocrinology* [Internet]. The Endocrine Society; 2016 Feb;157(2):570–585. Available from: <https://dx.doi.org/10.1210/en.2015-1866> PMID: PMC4733118
 180. Sheedfar F, Sung MM, Aparicio-Vergara M, Kloosterhuis NJ, Miquilena-Colina ME, Vargas-Castrillón J, Febbraio M, Jacobs RL, de Bruin A, Vinciguerra M, García-Monzón C, Hofker MH, Dyck JR, Koonen DPY. Increased hepatic CD36 expression with age is associated with enhanced susceptibility to nonalcoholic fatty liver disease. *Aging (Albany NY)* [Internet]. Impact Journals, LLC; 2014 Apr;6(4):281–295. Available from: <http://dx.doi.org/10.18632/aging.100652> PMID: PMC4032795
 181. Rada P, González-Rodríguez Á, García-Monzón C, Valverde ÁM. Understanding

- lipotoxicity in NAFLD pathogenesis: is CD36 a key driver? *Cell Death Dis* [Internet]. Springer Science and Business Media LLC; 2020 Sep 25;11(9):802. Available from: <http://dx.doi.org/10.1038/s41419-020-03003-w> PMID: PMC7519685
182. Li Y, Kang H, Chu Y, Jin Y, Zhang L, Yang R, Zhang Z, Zhao S, Zhou L. Cidec differentially regulates lipid deposition and secretion through two tissue-specific isoforms. *Gene* [Internet]. 2018 Jan 30;641:265–271. Available from: <http://dx.doi.org/10.1016/j.gene.2017.10.069> PMID: 29080839
 183. Fu Q, Wang P, Li W, Cai Z, Zhao S, Ling W, Li M, Tang X, Song Z. Partial inhibition of adipose CIDEC improves insulin sensitivity and increases energy expenditure in high-fat diet-fed mice via activating ATGL-PPAR α pathway. *Biochim Biophys Acta Mol Cell Biol Lipids* [Internet]. Elsevier BV; 2025 Aug;1870(6):159659. Available from: <http://dx.doi.org/10.1016/j.bbalip.2025.159659> PMID: 40633832
 184. Langhi C, Baldán Á. CIDEC/FSP27 is regulated by peroxisome proliferator-activated receptor alpha and plays a critical role in fasting- and diet-induced hepatosteatosis. *Hepatology* [Internet]. Ovid Technologies (Wolters Kluwer Health); 2015 Apr;61(4):1227–1238. Available from: <http://dx.doi.org/10.1002/hep.27607> PMID: PMC4376564
 185. Li H, Sun J, Li B, Jiang A, Tao J, Ning C, Li R, Liu H. AMPK-PPAR γ -Cidec axis drives the fasting-induced lipid droplet aggregation in the liver of obese mice. *Front Nutr* [Internet]. 2022 Jul 4;9:917801. Available from: <http://dx.doi.org/10.3389/fnut.2022.917801> PMID: PMC9289538
 186. Kim YJ, Cho SY, Yun CH, Moon YS, Lee TR, Kim SH. Transcriptional activation of Cidec by PPAR γ 2 in adipocyte. *Biochem Biophys Res Commun* [Internet]. Elsevier BV; 2008 Dec 5;377(1):297–302. Available from: <http://dx.doi.org/10.1016/j.bbrc.2008.09.129> PMID: 18845124
 187. Wang X, Zhang L, Chang Y, Guo Y, Yang G, Xie W, Zhu M, Teng J, Shen J, Jia W, Chen S, Chen T, Deng Y. An integrated multi-omics Mendelian Randomization identifies predictive transcription gene signature of liver fibrosis [Internet]. Research Square. 2024 [cited 2025 Aug 16]. Available from: <https://www.researchsquare.com/article/rs-4395597/v1>
 188. Juraszek B, Nałęcz KA. SLC22A5 (OCTN2) carnitine transporter-indispensable for cell metabolism, a Jekyll and Hyde of human cancer. *Molecules* [Internet]. MDPI AG; 2019 Dec 19;25(1):14. Available from: <http://dx.doi.org/10.3390/molecules25010014> PMID: PMC6982704
 189. Vogel J, Hinrichs SH, Napolitano LA, Ngo L, Jay G. Liver cancer in transgenic mice carrying the human immunodeficiency virus tat gene. *Cancer Res* [Internet]. American Association for Cancer Research; 1991 Dec 15;51(24):6686–6690. Available from: <https://dx.doi.org/> PMID: 1742742
 190. Altavilla G, Caputo A, Lanfredi M, Piola C, Barbanti-Brodano G, Corallini A. Enhancement of Chemical Hepatocarcinogenesis by the HIV-1 tat Gene. *Am J Pathol* [Internet]. Elsevier BV; 2000 Oct;157(4):1081–1089. Available from:

[http://dx.doi.org/10.1016/S0002-9440\(10\)64622-6](http://dx.doi.org/10.1016/S0002-9440(10)64622-6)

191. Li D, Tang X, Li M, Zheng Y. Long noncoding RNA DLX6-AS1 promotes liver cancer by increasing the expression of WEE1 via targeting miR-424-5p. *J Cell Biochem* [Internet]. Wiley; 2019 Aug;120(8):12290–12299. Available from: <http://dx.doi.org/10.1002/jcb.28493> PMID: PMC6712946
192. Zhao W, Liu S, Dou Q, Li C, Du J, Ren W. The role and mechanism of WEE1 on the cisplatin resistance reversal of the HepG2/DDP human hepatic cancer cell line. *Oncol Lett* [Internet]. Spandidos Publications; 2015 Nov;10(5):3081–3086. Available from: <http://dx.doi.org/10.3892/ol.2015.3647> PMID: PMC4665834
193. Liu Y, Lin N, Dai C, Xu J, Zhang Y, Xu M, Wang F, Li Y, Chen D. Occurrence and distribution of per- and polyfluoroalkyl substances (PFASs) in human livers with liver cancer. *Environ Res* [Internet]. Elsevier BV; 2021 Nov;202(111775):111775. Available from: <http://dx.doi.org/10.1016/j.envres.2021.111775> PMID: 34333008
194. Dai C, Peng L, Li Y, Li Z, Chen D, Wang F, Lin N. Distribution of per- and polyfluoroalkyl substances in blood, serum, and urine of patients with liver cancer and associations with liver function biomarkers. *J Environ Sci (China)* [Internet]. 2024 May;139:418–427. Available from: <http://dx.doi.org/10.1016/j.jes.2023.05.026> PMID: 38105065
195. Abate HM, Bujnowski D, Jowell A, Moylan CA, Hoyo C, Wegermann K. Ubiquitous environmental exposures and risk of hepatocellular carcinoma: A narrative review. *Dig Dis Sci* [Internet]. 2025 Aug 10; Available from: <http://dx.doi.org/10.1007/s10620-025-09212-7> PMID: 40783888
196. Sonne C, Gustavson K, Bossi R, Søndergaard J, Desforges JP, Bonefeld-Jørgensen EC, Dietz R. Ubiquitous global use of persistent PFAS threatens Arctic Indigenous peoples for decades to come. *Cell Rep Sustain* [Internet]. Elsevier BV; 2025 Mar;2(3):100341. Available from: <http://dx.doi.org/10.1016/j.crsus.2025.100341>
197. De Silva AO, Armitage JM, Bruton TA, Dassuncao C, Heiger-Bernays W, Hu XC, Kärrman A, Kelly B, Ng C, Robuck A, Sun M, Webster TF, Sunderland EM. PFAS exposure pathways for humans and wildlife: A synthesis of current knowledge and key gaps in understanding. *Environ Toxicol Chem* [Internet]. Oxford University Press (OUP); 2021 Mar;40(3):631–657. Available from: <https://dx.doi.org/10.1002/etc.4935> PMID: PMC7906948
198. Oliveira ACP, Bertollo CM, Rocha LTS, Nascimento EB Jr, Costa KA, Coelho MM. Antinociceptive and antiedematogenic activities of fenofibrate, an agonist of PPAR alpha, and pioglitazone, an agonist of PPAR gamma. *Eur J Pharmacol* [Internet]. Elsevier BV; 2007 Apr 30;561(1-3):194–201. Available from: <http://dx.doi.org/10.1016/j.ejphar.2006.12.026> PMID: 17343847
199. Gillies PS, Dunn CJ. Pioglitazone. *Drugs* [Internet]. Springer Nature; 2000 Aug;60(2):333–43; discussion 344–5. Available from: <http://dx.doi.org/10.2165/00003495-200060020-00009> PMID: 10983737

200. Ress C, Moschen AR, Sausgruber N, Tschoner A, Graziadei I, Weiss H, Schgoer W, Ebenbichler CF, Konrad RJ, Patsch JR, Tilg H, Kaser S. The role of apolipoprotein A5 in non-alcoholic fatty liver disease. *Gut* [Internet]. BMJ; 2011 Jul;60(7):985–991. Available from: <http://dx.doi.org/10.1136/gut.2010.222224> PMID: 21339203
201. Feng Q, Baker SS, Liu W, Arbizu RA, Aljomah G, Khatib M, Nugent CA, Baker RD, Forte TM, Hu Y, Zhu L. Increased apolipoprotein A5 expression in human and rat non-alcoholic fatty livers. *Pathology* [Internet]. Elsevier BV; 2015 Jun;47(4):341–348. Available from: <http://dx.doi.org/10.1097/PAT.0000000000000251> PMCID: PMC4427056
202. Xu Y, Lu LL, Liu SS, Du SX, Zhu HL, Dong QJ, Xin YN. Apolipoprotein A5 gene polymorphisms are associated with non-alcoholic fatty liver disease. *Hepatobiliary Pancreat Dis Int* [Internet]. 2018 Jun;17(3):214–219. Available from: <http://dx.doi.org/10.1016/j.hbpd.2018.04.004> PMID: 29735301
203. Chen HT, Huang C, Chen JW, Yang SQ, Cheng JM, Li YQ, Chen HQ, Zhou YJ. Hepatocyte FoxO1 depletion exacerbates hepatic inflammation in MASH by targeting cystathionine γ -lyase. *Sci Rep* [Internet]. Springer Science and Business Media LLC; 2025 Jul 22;15(1):26631. Available from: <http://dx.doi.org/10.1038/s41598-025-10192-x> PMCID: PMC12283972
204. Katsouda A, Markou M, Valakos D, Theodorou I, Papapetropoulos A. Cth/Mpst double ablation results in early onset fatty liver disease in lean mice. *Redox Biol* [Internet]. Elsevier BV; 2025 Jun;83(103641):103641. Available from: <http://dx.doi.org/10.1016/j.redox.2025.103641> PMCID: PMC12023873
205. Talari NK, Mattam U, Kaminska D, Sotomayor-Rodriguez I, Rahman AP, Péterfy M, Pajukanta P, Pihlajamäki J, Chella Krishnan K. Hepatokine ITIH3 protects against hepatic steatosis by downregulating mitochondrial bioenergetics and de novo lipogenesis. *iScience* [Internet]. Elsevier BV; 2024 May 17;27(5):109709. Available from: <http://dx.doi.org/10.1016/j.isci.2024.109709> PMCID: PMC11059128
206. Chen M, Zhu JY, Mu WJ, Luo HY, Li Y, Li S, Yan LJ, Li RY, Guo L. Cdo1-Camkk2-AMPK axis confers the protective effects of exercise against NAFLD in mice. *Nat Commun* [Internet]. Nature Publishing Group; 2023 Dec 18;14(1):8391. Available from: <http://dx.doi.org/10.1038/s41467-023-44242-7> PMCID: PMC10728194
207. Simmen FA, Pabona JMP, Al-Dwairi A, Alhallak I, Montales MTE, Simmen RCM. Malic Enzyme 1 (ME1) promotes adiposity and hepatic steatosis and induces circulating insulin and Leptin in obese female mice. *Int J Mol Sci* [Internet]. Multidisciplinary Digital Publishing Institute; 2023 Apr 1;24(7):6613. Available from: <http://dx.doi.org/10.3390/ijms24076613> PMCID: PMC10095602
208. Devlin RB, Frampton ML, Ghio AJ. In vitro studies: what is their role in toxicology? *Exp Toxicol Pathol* [Internet]. 2005 Jul;57 Suppl 1:183–188. Available from: <http://dx.doi.org/10.1016/j.etp.2005.05.018> PMID: 16092726
209. Guillouzo A. Liver cell models in in vitro toxicology. *Environ Health Perspect* [Internet].

- 1998 Apr;106 Suppl 2:511–532. Available from: <http://dx.doi.org/10.1289/ehp.98106511> PMID: PMC1533385
210. Liebsch M, Spielmann H. Currently available in vitro methods used in the regulatory toxicology. *Toxicol Lett* [Internet]. Elsevier BV; 2002 Feb 28;127(1-3):127–134. Available from: [http://dx.doi.org/10.1016/S0378-4274\(01\)00492-1](http://dx.doi.org/10.1016/S0378-4274(01)00492-1) PMID: 12052650
 211. Zhang QY, Xu LL, Zhong MT, Chen YK, Lai MQ, Wang Q, Xie XL. Gestational GenX and PFOA exposures induce hepatotoxicity, metabolic pathway, and microbiome shifts in weanling mice. *Sci Total Environ* [Internet]. Elsevier BV; 2024 Jan 10;907(168059):168059. Available from: <http://dx.doi.org/10.1016/j.scitotenv.2023.168059> PMID: 37884144
 212. Alijagic A, Sinisalu L, Duberg D, Kotlyar O, Scherbak N, Engwall M, Orešič M, Hyötyläinen T. Metabolic and phenotypic changes induced by PFAS exposure in two human hepatocyte cell models. *Environ Int* [Internet]. Elsevier BV; 2024 Aug;190(108820):108820. Available from: <http://dx.doi.org/10.1016/j.envint.2024.108820> PMID: 38906088
 213. Jin R, McConnell R, Catherine C, Xu S, Walker DI, Stratakis N, Jones DP, Miller GW, Peng C, Conti DV, Vos MB, Chatzi L. Perfluoroalkyl substances and severity of nonalcoholic fatty liver in Children: An untargeted metabolomics approach. *Environ Int* [Internet]. Elsevier BV; 2020 Jan;134(105220):105220. Available from: <http://dx.doi.org/10.1016/j.envint.2019.105220> PMID: PMC6944061
 214. Sen P, Qadri S, Luukkonen PK, Ragnarsdottir O, McGlinchey A, Jäntti S, Juuti A, Arola J, Schlezinger JJ, Webster TF, Orešič M, Yki-Järvinen H, Hyötyläinen T. Exposure to environmental contaminants is associated with altered hepatic lipid metabolism in non-alcoholic fatty liver disease. *J Hepatol* [Internet]. Elsevier BV; 2022 Feb;76(2):283–293. Available from: <http://dx.doi.org/10.1016/j.jhep.2021.09.039> PMID: 34627976
 215. Hari A, AbdulHameed MDM, Balik-Meisner MR, Mav D, Phadke DP, Scholl EH, Shah RR, Casey W, Auerbach SS, Wallqvist A, Pannala VR. Exposure to PFAS chemicals induces sex-dependent alterations in key rate-limiting steps of lipid metabolism in liver steatosis. *Front Toxicol* [Internet]. Frontiers Media SA; 2024 Jun 5;6:1390196. Available from: <http://dx.doi.org/10.3389/ftox.2024.1390196> PMID: PMC11188372
 216. Seyyedsalehi MS, Boffetta P. Per- and poly-fluoroalkyl substances (PFAS) exposure and risk of kidney, liver, and testicular cancers: A systematic review and meta-analysis. *Med Lav* [Internet]. Mattioli1885; 2023 Oct 24;114(5):e2023040. Available from: <http://dx.doi.org/10.23749/mdl.v114i5.15065> PMID: PMC10627102
 217. McGlynn KA, Watling CZ, Hernandez BY, Groopman JD. Environmental risk factors for liver cancer. *Hepatology* [Internet]. Ovid Technologies (Wolters Kluwer Health); 2025 Jul 7;10.1097/HEP.0000000000001463. Available from: <http://dx.doi.org/10.1097/HEP.0000000000001463> PMID: 40622863
 218. Chen Z, Yang T, Walker DI, Thomas DC, Qiu C, Chatzi L, Alderete TL, Kim JS, Conti DV, Breton CV, Liang D, Hauser ER, Jones DP, Gilliland FD. Dysregulated lipid and fatty acid metabolism link perfluoroalkyl substances exposure and impaired glucose metabolism

- in young adults. *Environ Int* [Internet]. Elsevier BV; 2020 Dec;145(106091):106091. Available from: <http://dx.doi.org/10.1016/j.envint.2020.106091> PMID: PMC8009052
219. Birru RL, Liang HW, Farooq F, Bedi M, Feghali M, Haggerty CL, Mendez DD, Catov JM, Ng CA, Adibi JJ. A pathway level analysis of PFAS exposure and risk of gestational diabetes mellitus. *Environ Health* [Internet]. Springer Science and Business Media LLC; 2021 May 22;20(1):63. Available from: <http://dx.doi.org/10.1186/s12940-021-00740-z> PMID: PMC8141246
 220. Gui SY, Qiao JC, Xu KX, Li ZL, Chen YN, Wu KJ, Jiang ZX, Hu CY. Association between per- and polyfluoroalkyl substances exposure and risk of diabetes: a systematic review and meta-analysis. *J Expo Sci Environ Epidemiol* [Internet]. Springer Science and Business Media LLC; 2023 Jan;33(1):40–55. Available from: <http://dx.doi.org/10.1038/s41370-022-00464-3> PMID: 35970987
 221. Zhang H, Zhou XD, Shapiro MD, Lip GYH, Tilg H, Valenti L, Somers VK, Byrne CD, Targher G, Yang W, Viveiros O, Opio CK, Mantzoros CS, Ryan JD, Kok KYY, Jumaev NA, Perera N, Robertson AG, Abu-Abeid A, Misra A, Wong YJ, Ruiz-Úcar E, Ospanov O, Kızılkaya MC, Luo F, Méndez-Sánchez N, Zuluaga M, Lonardo A, Al Momani H, Toro-Huamanchumo CJ, Adams L, Al-Busafi SA, Sharara AI, Chan WK, Abbas SI, Sookoian S, Treeprasertsuk S, Ocama P, Alswat K, Kong APS, Ataya K, Lim-Loo MC, Oviedo RJ, Szepietowski O, Fouad Y, Zhang H, Abdelbaki TN, Katsouras CS, Prasad A, Thaher O, Ali A, Molina GA, Sung KC, Chen QF, Lesmana CRA, Zheng MH. Global burden of metabolic diseases, 1990–2021. *Metabolism* [Internet]. Elsevier BV; 2024 Nov;160(155999):155999. Available from: <http://dx.doi.org/10.1016/j.metabol.2024.155999> PMID: 39151887
 222. Travis CC. Interspecies Extrapolation of Toxicological Data. *Health Risk Assessment Dermal and Inhalation Exposure and Absorption of Toxicants* [Internet]. 1st Edition. CRC Press; 2017 [cited 2025 Aug 16]. p. 387–410. Available from: <http://dx.doi.org/10.1201/9780203711989-27>
 223. Schmeisser S, Miccoli A, von Bergen M, Berggren E, Braeuning A, Busch W, Desaintes C, Gourmelon A, Grafström R, Harrill J, Hartung T, Herzler M, Kass GEN, Kleinstreuer N, Leist M, Luijten M, Marx-Stoelting P, Poetz O, van Ravenzwaay B, Roggeband R, Rogiers V, Roth A, Sanders P, Thomas RS, Marie Vinggaard A, Vinken M, van de Water B, Luch A, Tralau T. New approach methodologies in human regulatory toxicology - Not if, but how and when! *Environ Int* [Internet]. Elsevier BV; 2023 Aug;178(108082):108082. Available from: <http://dx.doi.org/10.1016/j.envint.2023.108082> PMID: PMC10858683
 224. Avila AM, Bebenek I, Bonzo JA, Bourcier T, Davis Bruno KL, Carlson DB, Dubinion J, Elayan I, Harrouk W, Lee SL, Mendrick DL, Merrill JC, Peretz J, Place E, Saulnier M, Wange RL, Yao J, Zhao D, Brown PC. An FDA/CDER perspective on nonclinical testing strategies: Classical toxicology approaches and new approach methodologies (NAMs). *Regul Toxicol Pharmacol* [Internet]. 2020 Jul;114:104662. Available from: <http://dx.doi.org/10.1016/j.yrtph.2020.104662> PMID: 32325112
 225. Anadón A, Martínez MA, Castellano V, Martínez-Larrañaga MR. The role of in vitro

- methods as alternatives to animals in toxicity testing. *Expert Opin Drug Metab Toxicol* [Internet]. Informa UK Limited; 2014 Jan;10(1):67–79. Available from: <http://dx.doi.org/10.1517/17425255.2014.854329> PMID: 24160258
226. Bakand S, Winder C, Khalil C, Hayes A. Toxicity assessment of industrial chemicals and airborne contaminants: transition from in vivo to in vitro test methods: a review. *Inhal Toxicol* [Internet]. 2005 Dec 1;17(13):775–787. Available from: <http://dx.doi.org/10.1080/08958370500225240> PMID: 16195213
 227. Bögler O, Huang H, Cavenee WK. Loss of Wild-Type p53 Bestows a Growth Advantage on Primary Cortical Astrocytes and Facilitates Their in Vitro Transformation¹. *Cancer Res American Association for Cancer Research*. 1995;55(13):2746–2751.
 228. Wang K, Shindoh H, Inoue T, Horii I. Advantages of in vitro cytotoxicity testing by using primary rat hepatocytes in comparison with established cell lines. *J Toxicol Sci* [Internet]. 2002 Aug;27(3):229–237. Available from: <http://dx.doi.org/10.2131/jts.27.229> PMID: 12238146
 229. Butterworth BE, Smith-Oliver T, Earle L, Loury DJ, White RD, Doolittle DJ, Working PK, Cattley RC, Jirtle R, Michalopoulos G, Strom S. Use of Primary Cultures of Human Hepatocytes in Toxicology Studies¹. *Cancer Res American Association for Cancer Research*. 1989;49(5):1075–1084.
 230. Carter M, Shieh JC. *Cell Culture Techniques. Guide to Research Techniques in Neuroscience*. Elsevier; 2010. p. 281–296.
 231. Allen DD, Caviedes R, Cárdenas AM, Shimahara T, Segura-Aguilar J, Caviedes PA. Cell lines as in vitro models for drug screening and toxicity studies. *Drug Dev Ind Pharm* [Internet]. 2005 Sep;31(8):757–768. Available from: <http://dx.doi.org/10.1080/03639040500216246> PMID: 16221610
 232. Gao Y, Kim S, Lee YI, Lee J. Cellular stress-modulating drugs can potentially be identified by in silico screening with connectivity map (CMap). *Int J Mol Sci* [Internet]. MDPI AG; 2019 Nov 9;20(22):5601. Available from: <http://dx.doi.org/10.3390/ijms20225601> PMCID: PMC6888006
 233. Subramanian A, Narayan R, Corsello SM, Peck DD, Natoli TE, Lu X, Gould J, Davis JF, Tubelli AA, Asiedu JK, Lahr DL, Hirschman JE, Liu Z, Donahue M, Julian B, Khan M, Wadden D, Smith I, Lam D, Liberzon A, Toder C, Bagul M, Orzechowski M, Enache OM, Piccioni F, Berger AH, Shamji A, Brooks AN, Vrcic A, Flynn C, Rosains J, Takeda D, Davison D, Lamb J, Ardlie K, Hogstrom L, Gray NS, Clemons PA, Silver S, Wu X, Zhao WN, Read-Button W, Wu X, Haggarty SJ, Ronco LV, Boehm JS, Schreiber SL, Doench JG, Bittker JA, Root DE, Wong B, Golub TR. A next generation Connectivity Map: L1000 platform and the first 1,000,000 profiles [Internet]. *bioRxiv. bioRxiv*; 2017. Available from: <http://dx.doi.org/10.1101/136168>
 234. Andrews TS, Atif J, Liu JC, Perciani CT, Ma XZ, Thoeni C, Slyper M, Eraslan G, Segerstolpe A, Manuel J, Chung S, Winter E, Cirlan I, Khuu N, Fischer S, Rozenblatt-Rosen

- O, Regev A, McGilvray ID, Bader GD, MacParland SA. Single-cell, single-nucleus, and spatial RNA sequencing of the human liver identifies cholangiocyte and mesenchymal heterogeneity. *Hepatology* [Internet]. Ovid Technologies (Wolters Kluwer Health); 2022 Apr;6(4):821–840. Available from: <http://dx.doi.org/10.1002/hep4.1854> PMID: PMC8948611
235. Wang Z, Shao X, Wang K, Lu X, Zhuang L, Yang X, Zhang P, Yang P, Zheng S, Xu X, Fan X. Single-cell analysis reveals a pathogenic cellular module associated with early allograft dysfunction after liver transplantation [Internet]. *bioRxiv*. 2022. Available from: <http://dx.doi.org/10.1101/2022.02.09.479667>
 236. Chen YW, Ahn IS, Wang SSM, Majid S, Diamante G, Cely I, Zhang G, Cabanayan A, Komzyuk S, Bonnett J, Arneson D, Yang X. Multitissue single-cell analysis reveals differential cellular and molecular sensitivity between fructose and high-fat high-sucrose diets. *Cell Rep* [Internet]. Elsevier BV; 2025 May 27;44(5):115690. Available from: <http://dx.doi.org/10.1016/j.celrep.2025.115690> PMID: PMC12182283
 237. Qi SY, Kim F, Adewale Y, Zhou C, Aldler M, Ni Y. Single-Cell RNA Transcriptome Landscape of Hepatocytes and Non-Parenchymal Cells in Healthy and NAFLD Mouse Liver. *IScience*. 2021;24(11).
 238. Regev A, Teichmann SA, Lander ES, Amit I, Benoist C, Birney E, Bodenmiller B, Campbell P, Carninci P, Clatworthy M, Clevers H, Deplancke B, Dunham I, Eberwine J, Eils R, Enard W, Farmer A, Fugger L, Göttgens B, Hacohen N, Haniffa M, Hemberg M, Kim S, Klennerman P, Kriegstein A, Lein E, Linnarsson S, Lundberg E, Lundberg J, Majumder P, Marioni JC, Merad M, Mhlanga M, Nawijn M, Netea M, Nolan G, Pe'er D, Phillipakis A, Ponting CP, Quake S, Reik W, Rozenblatt-Rosen O, Sanes J, Satija R, Schumacher TN, Shalek A, Shapiro E, Sharma P, Shin JW, Stegle O, Stratton M, Stubbington MJT, Theis FJ, Uhlen M, van Oudenaarden A, Wagner A, Watt F, Weissman J, Wold B, Xavier R, Yosef N, Human Cell Atlas Meeting Participants. The Human Cell Atlas. *Elife* [Internet]. 2017 Dec 5;6. Available from: <http://dx.doi.org/10.7554/eLife.27041> PMID: PMC5762154
 239. Han X, Zhou Z, Fei L, Sun H, Wang R, Chen Y, Chen H, Wang J, Tang H, Ge W, Zhou Y, Ye F, Jiang M, Wu J, Xiao Y, Jia X, Zhang T, Ma X, Zhang Q, Bai X, Lai S, Yu C, Zhu L, Lin R, Gao Y, Wang M, Wu Y, Zhang J, Zhan R, Zhu S, Hu H, Wang C, Chen M, Huang H, Liang T, Chen J, Wang W, Zhang D, Guo G. Construction of a human cell landscape at single-cell level. *Nature* [Internet]. Springer Science and Business Media LLC; 2020 May;581(7808):303–309. Available from: <http://dx.doi.org/10.1038/s41586-020-2157-4> PMID: 32214235
 240. Tabula Muris Consortium, Overall coordination, Logistical coordination, Organ collection and processing, Library preparation and sequencing, Computational data analysis, Cell type annotation, Writing group, Supplemental text writing group, Principal investigators. Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. *Nature* [Internet]. Springer Science and Business Media LLC; 2018 Oct;562(7727):367–372. Available from: <http://dx.doi.org/10.1038/s41586-018-0590-4> PMID: PMC6642641

241. Tarhan L, Bistline J, Chang J, Galloway B, Hanna E, Weitz E. Single Cell Portal: an interactive home for single-cell genomics data [Internet]. bioRxiv.org. 2023. Available from: <http://dx.doi.org/10.1101/2023.07.13.548886> PMID: PMC10370058
242. Hao Y, Stuart T, Kowalski MH, Choudhary S, Hoffman P, Hartman A, Srivastava A, Molla G, Madad S, Fernandez-Granda C, Satija R. Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nat Biotechnol* [Internet]. 2024 Feb;42(2):293–304. Available from: <http://dx.doi.org/10.1038/s41587-023-01767-y> PMID: PMC10928517
243. Bolstad BM. preprocessCore: A collection of pre-processing functions. R package version. 2013;
244. Konopka T. Uniform Manifold Approximation and Projection [R package umap version 0.2.10.0] [Internet]. Comprehensive R Archive Network (CRAN). 2023 [cited 2025 Aug 17]. Available from: <https://cran.r-project.org/web/packages/umap/index.html>
245. Franzén O, Gan LM, Björkegren JLM. PanglaoDB: a web server for exploration of mouse and human single-cell RNA sequencing data. *Database (Oxford)* [Internet]. Oxford University Press (OUP); 2019 Jan 1;2019. Available from: <http://dx.doi.org/10.1093/database/baz046> PMID: PMC6450036
246. missForest: Nonparametric Missing Value Imputation using Random Forest [Internet]. Comprehensive R Archive Network (CRAN). [cited 2025 Aug 17]. Available from: <https://cran.r-project.org/web/packages/missForest/index.html>
247. Kanehisa M. *Silico” Simulation of Biological Processes*. Chichester, UK; Ltd: John Wiley & Sons; 2003. p. 91–103.
248. Piñero J, Queralt-Rosinach N, Bravo À, Deu-Pons J, Bauer-Mehren A, Baron M, Sanz F, Furlong LI. DisGeNET: a discovery platform for the dynamical exploration of human diseases and their genes. *Database* [Internet]. 2015 Apr 15;2015:bav028. Available from: <http://dx.doi.org/10.1093/database/bav028> PMID: PMC4397996
249. Littman R, Cheng M, Wang N, Peng C, Yang X. SCING: Inference of robust, interpretable gene regulatory networks from single cell and spatial transcriptomics. *iScience* [Internet]. Elsevier BV; 2023 Jul 21;26(7):107124. Available from: <http://dx.doi.org/10.1016/j.isci.2023.107124> PMID: PMC10331489
250. Svobodová J, Procházková J, Kabátková M, Krkoška M, Šmerdová L, Líbalová H, Topinka J, Kléma J, Kozubík A, Machala M, Vondráček J. 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD) disrupts control of cell proliferation and apoptosis in a human model of adult liver progenitors. *Toxicol Sci* [Internet]. Oxford University Press (OUP); 2019 Dec 1;172(2):368–384. Available from: <http://dx.doi.org/10.1093/toxsci/kfz202> PMID: 31536130
251. GEO Accession viewer [Internet]. [cited 2025 Jun 22]. Available from: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi>
252. Cholino GN, Nault R, Zacharewski T. Cell-specific AHR-driven differential gene

- expression in the mouse liver cell following acute TCDD exposure. *BMC Genomics* [Internet]. Springer Science and Business Media LLC; 2024 Aug 28;25(1):809. Available from: <http://dx.doi.org/10.1186/s12864-024-10730-3> PMID: PMC11351262
253. Harjumäki R, Pridgeon CS, Ingelman-Sundberg M. CYP2E1 in alcoholic and non-alcoholic liver injury. Roles of ROS, reactive intermediates and lipid overload. *Int J Mol Sci* [Internet]. MDPI AG; 2021 Jul 30;22(15):8221. Available from: <http://dx.doi.org/10.3390/ijms22158221> PMID: PMC8348366
 254. Li D, Xie P, Zhao S, Zhao J, Yao Y, Zhao Y, Ren G, Liu X. Hepatocytes derived increased SAA1 promotes intrahepatic platelet aggregation and aggravates liver inflammation in NAFLD. *Biochem Biophys Res Commun* [Internet]. Elsevier BV; 2021 May 28;555:54–60. Available from: <http://dx.doi.org/10.1016/j.bbrc.2021.02.124> PMID: 33813276
 255. Bhanumathy CD, Tang Y, Monga SPS, Katuri V, Cox JA, Mishra B, Mishra L. Itih-4, a serine protease inhibitor regulated in interleukin-6-dependent liver formation: role in liver development and regeneration. *Dev Dyn* [Internet]. Wiley; 2002 Jan;223(1):59–69. Available from: <http://dx.doi.org/10.1002/dvdy.1235> PMID: 11803570
 256. Zhu YH, Li JB, Wu RY, Yu Y, Li X, Li ZL, Zhang HL, Feng GK, Deng R, Zhu XF. Clinical significance and function of RDH16 as a tumor-suppressing gene in hepatocellular carcinoma. *Hepatol Res* [Internet]. Wiley; 2020 Jan;50(1):110–120. Available from: <http://dx.doi.org/10.1111/hepr.13432> PMID: 31661588
 257. Qiu GH, Xie X, Xu F, Shi X, Wang Y, Deng L. Distinctive pharmacological differences between liver cancer cell lines HepG2 and Hep3B. *Cytotechnology* [Internet]. Springer Science and Business Media LLC; 2015 Jan;67(1):1–12. Available from: <http://dx.doi.org/10.1007/s10616-014-9761-9> PMID: PMC4294832
 258. Sainz B Jr, Barretto N, Uprichard SL. Hepatitis C virus infection in phenotypically distinct Huh7 cell lines. *PLoS One* [Internet]. Public Library of Science (PLOS); 2009 Aug 10;4(8):e6561. Available from: <http://dx.doi.org/10.1371/journal.pone.0006561> PMID: PMC2720605
 259. Damgaard RB. The ubiquitin system: from cell signalling to disease biology and new therapeutic opportunities. *Cell Death Differ* [Internet]. Springer Science and Business Media LLC; 2021 Feb;28(2):423–426. Available from: <http://dx.doi.org/10.1038/s41418-020-00703-w> PMID: PMC7862391
 260. Guillouzo A, Corlu A, Aninat C, Glaize D, Morel F, Guguen-Guillouzo C. The human hepatoma HepaRG cells: a highly differentiated model for studies of liver metabolism and toxicity of xenobiotics. *Chem Biol Interact* [Internet]. Elsevier BV; 2007 May 20;168(1):66–73. Available from: <http://dx.doi.org/10.1016/j.cbi.2006.12.003> PMID: 17241619
 261. Leite SB, Wilk-Zasadna I, Zaldivar JM, Airola E, Reis-Fernandes MA, Mennecozzi M, Guguen-Guillouzo C, Chesne C, Guillou C, Alves PM, Coecke S. Three-dimensional HepaRG model as an attractive tool for toxicity testing. *Toxicol Sci* [Internet]. Oxford University Press (OUP); 2012 Nov;130(1):106–116. Available from:

<https://dx.doi.org/10.1093/toxsci/kfs232> PMID: 22843569

262. Andersson TB, Kanebratt KP, Kenna JG. The HepaRG cell line: a unique in vitro tool for understanding drug metabolism and toxicology in human. *Expert Opin Drug Metab Toxicol* [Internet]. Informa Healthcare; 2012 Jul;8(7):909–920. Available from: <http://dx.doi.org/10.1517/17425255.2012.685159> PMID: 22568886
263. Ndikung J, Storm D, Violet N, Kramer A, Schönfelder G, Ertych N, Oelgeschläger M. Restoring circadian synchrony in vitro facilitates physiological responses to environmental chemicals. *Environ Int* [Internet]. Elsevier BV; 2020 Jan;134(105265):105265. Available from: <http://dx.doi.org/10.1016/j.envint.2019.105265> PMID: 31734582
264. Güdül Ü, Kordbacheh H, Sanajou S, Şahin G. Chronotoxicity studies in pharmaceutical Science: A comprehensive review. *EMU Journal of Pharmaceutical Sciences* [Internet]. EMU Journal of Pharmaceutical Sciences; 2024 Nov 28;7(2):68–79. Available from: <https://doi.org/10.54994/emujpharmsci.1536853>
265. Abbas W, Kumar A, Herbein G. The eEF1A proteins: At the Crossroads of oncogenesis, apoptosis, and viral infections. *Front Oncol* [Internet]. Frontiers Media SA; 2015 Apr 7;5:75. Available from: <http://dx.doi.org/10.3389/fonc.2015.00075> PMCID: PMC4387925
266. Wilson RB, Chen YJ, Zhang R, Maini S, Andrews TS, Wang R, Borradaile NM. Elongation factor 1A1 inhibition elicits changes in lipid droplet size, the bulk transcriptome, and cell type-associated gene expression in MASLD mouse liver. *Am J Physiol Gastrointest Liver Physiol* [Internet]. American Physiological Society; 2024 Oct 1;327(4):G608–G622. Available from: <http://dx.doi.org/10.1152/ajpgi.00276.2023> PMCID: PMC11482270
267. Stoianov AM, Robson DL, Hetherington AM, Sawyez CG, Borradaile NM. Elongation factor 1A-1 is a mediator of hepatocyte lipotoxicity partly through its canonical function in protein synthesis. *PLoS One* [Internet]. Public Library of Science (PLOS); 2015 Jun 23;10(6):e0131269. Available from: <http://dx.doi.org/10.1371/journal.pone.0131269> PMCID: PMC4478042
268. Lee SK, Son LT, Choi HJ, Ahnn J. Dicarboxyl/l-xylulose reductase (DCXR): The multifunctional pentosuria enzyme. *Int J Biochem Cell Biol* [Internet]. Elsevier BV; 2013 Nov;45(11):2563–2567. Available from: <http://dx.doi.org/10.1016/j.biocel.2013.08.010> PMID: 23988570
269. Hang X, Wu Z, Chu K, Yu G, Peng H, Xin H, Miao X, Wang J, Xu W. Low expression of DCXR protein indicates a poor prognosis for hepatocellular carcinoma patients. *Tumour Biol* [Internet]. Springer Nature; 2016 Nov;37(11):15079–15085. Available from: <http://dx.doi.org/10.1007/s13277-016-5302-9> PMID: 27658779
270. Mensenkamp AR, Jong MC, van Goor H, van Luyn MJ, Bloks V, Havinga R, Voshol PJ, Hofker MH, van Dijk KW, Havekes LM, Kuipers F. Apolipoprotein E participates in the regulation of very low density lipoprotein-triglyceride secretion by the liver. *J Biol Chem* [Internet]. Elsevier BV; 1999 Dec 10;274(50):35711–35718. Available from: <http://dx.doi.org/10.1074/jbc.274.50.35711> PMID: 10585451

271. Mensenkamp AR, Havekes LM, Romijn JA, Kuipers F. Hepatic steatosis and very low density lipoprotein secretion: the involvement of apolipoprotein E. *J Hepatol* [Internet]. Elsevier BV; 2001 Dec;35(6):816–822. Available from: [http://dx.doi.org/10.1016/S0168-8278\(01\)00249-5](http://dx.doi.org/10.1016/S0168-8278(01)00249-5) PMID: 11738112
272. Wozniak MA, Itzhaki RF, Faragher EB, James MW, Ryder SD, Irving WL, Trent HCV Study Group. Apolipoprotein E-epsilon 4 protects against severe liver disease caused by hepatitis C virus: Apolipoprotein E-epsilon 4 protects against severe liver disease caused by hepatitis C virus. *Hepatology* [Internet]. Ovid Technologies (Wolters Kluwer Health); 2002 Aug;36(2):456–463. Available from: <http://dx.doi.org/10.1053/jhep.2002.34745> PMID: 12143056
273. Habenicht LKL, Wang Z, Zhang X, Li Y, Mogler C, Huspenina JS, Schmid RM, Weber C, Mohanta SK, Ma Z, Yin C. The C1q-ApoE complex: A new hallmark pathology of viral hepatitis and nonalcoholic fatty liver disease. *Front Immunol* [Internet]. Frontiers Media SA; 2022 Oct 6;13:970938. Available from: <http://dx.doi.org/10.3389/fimmu.2022.970938> PMCID: PMC9592549
274. Li H, Jin J, Xian J, Wang W. lncRNA TPT1-AS1 knockdown inhibits liver cancer cell proliferation, migration and invasion. *Mol Med Rep* [Internet]. Spandidos Publications; 2021 Nov;24(5). Available from: <http://dx.doi.org/10.3892/mmr.2021.12422> PMCID: PMC8441979
275. Bae SY, Byun S, Bae SH, Min DS, Woo HA, Lee K. TPT1 (tumor protein, translationally-controlled 1) negatively regulates autophagy through the BECN1 interactome and an MTORC1-mediated pathway. *Autophagy* [Internet]. 2017 May 4;13(5):820–833. Available from: <http://dx.doi.org/10.1080/15548627.2017.1287650> PMCID: PMC5446065
276. Wei W, Huang X, Shen X, Lian J, Chen Y, Wang W, Huang J, Zhang B. Overexpression of lncRNA TPT1-AS1 suppresses hepatocellular carcinoma cell proliferation by downregulating CDK2. *Crit Rev Eukaryot Gene Expr* [Internet]. Begell House; 2022;32(1):1–9. Available from: <http://dx.doi.org/10.1615/CritRevEukaryotGeneExpr.2021039224> PMID: 35377976
277. Yao M, Fu L, Liu X, Zheng D. In-silico multi-omics analysis of the functional significance of calmodulin 1 in multiple cancers. *Front Genet* [Internet]. Frontiers Media SA; 2021;12:793508. Available from: <http://dx.doi.org/10.3389/fgene.2021.793508> PMCID: PMC8790318
278. Zuo B, Kuai J, Long J, Bian J, Yang X, Yang X, Xun Z, Li Y, Sun H, Sang X, Zhao H. Differentially expressed liver exosome-related genes as candidate prognostic biomarkers for hepatocellular carcinoma. *Ann Transl Med* [Internet]. AME Publishing Company; 2022 Aug;10(15):817. Available from: <http://dx.doi.org/10.21037/atm-21-4400> PMCID: PMC9403937
279. Song MJ, Jung CK, Park CH, Hur W, Choi JE, Bae SH, Choi JY, Choi SW, Han NI, Yoon SK. RPL36 as a prognostic marker in hepatocellular carcinoma: RPL36 expression in HCC. *Pathol Int* [Internet]. Wiley; 2011 Nov;61(11):638–644. Available from:

<http://dx.doi.org/10.1111/j.1440-1827.2011.02716.x> PMID: 22029674

280. Nakanishi T, Kawasaki Y, Nakamura Y, Kimura Y, Kawamura K, Shumba MN, Shimokawa N. An implication of the mitochondrial carrier SLC25A3 as an oxidative stress modulator in NAFLD. *Exp Cell Res* [Internet]. Elsevier BV; 2023 Oct 1;431(1):113740. Available from: <http://dx.doi.org/10.1016/j.yexcr.2023.113740> PMID: 37557977
281. Palmieri F. Mitochondrial transporters of the SLC25 family and associated diseases: a review. *J Inherit Metab Dis* [Internet]. Wiley; 2014 Jul;37(4):565–575. Available from: <http://dx.doi.org/10.1007/s10545-014-9708-5> PMID: 24797559
282. Liu Y, Yu X, Wang Y, Wu J, Feng B, Li M. The role of differentially expressed genes and immune cell infiltration in the progression of nonalcoholic steatohepatitis (NASH) to hepatocellular carcinoma (HCC): a new exploration based on bioinformatics analysis. *Nucleosides Nucleotides Nucleic Acids* [Internet]. Informa UK Limited; 2024 Feb 6;43(12):1415–1430. Available from: <http://dx.doi.org/10.1080/15257770.2024.2310044> PMID: 38319987
283. Liu T, Long Q, Li L, Gan H, Hu X, Long H, Yang L, Pang P, Wang S, Deng W. The NRF2-dependent transcriptional axis, XRCC5/hTERT drives tumor progression and 5-Fu insensitivity in hepatocellular carcinoma. *Mol Ther Oncolytics* [Internet]. Elsevier BV; 2022 Mar 17;24:249–261. Available from: <http://dx.doi.org/10.1016/j.omto.2021.12.012> PMCID: PMC8762376
284. Sanderson JT. The steroid hormone biosynthesis pathway as a target for endocrine-disrupting chemicals. *Toxicol Sci* [Internet]. Oxford University Press (OUP); 2006 Nov;94(1):3–21. Available from: <https://dx.doi.org/10.1093/toxsci/kfl051> PMID: 16807284
285. Santostefano MJ, Wang X, Richardson VM, Ross DG, DeVito MJ, Birnbaum LS. A pharmacodynamic analysis of TCDD-induced cytochrome P450 gene expression in multiple tissues: dose- and time-dependent effects. *Toxicol Appl Pharmacol* [Internet]. Elsevier BV; 1998 Aug;151(2):294–310. Available from: <http://dx.doi.org/10.1006/taap.1998.8466> PMID: 9707506
286. Devito MJ, Maier WE, Diliberto JJ, Birnbaum LS. Comparative Ability of Various PCBs, PCDFs, and TCDD to Induce Cytochrome P450 1A1 and 1A2 Activity Following 4 Weeks of Treatment. *Toxicol Sci* [Internet]. Oxford University Press (OUP); 1993;20(1):125–130. Available from: <https://dx.doi.org/10.1093/toxsci/20.1.125>
287. Arzumanyan VA, Kiseleva OI, Poverennaya EV. The curious case of the HepG2 cell line: 40 years of expertise. *Int J Mol Sci* [Internet]. MDPI AG; 2021 Dec 4;22(23):13135. Available from: <http://dx.doi.org/10.3390/ijms222313135> PMCID: PMC8658661
288. Wilkening S, Stahl F, Bader A. Comparison of primary human hepatocytes and hepatoma cell line Hepg2 with regard to their biotransformation properties. *Drug Metab Dispos* [Internet]. Elsevier BV; 2003 Aug;31(8):1035–1042. Available from: <http://dx.doi.org/10.1124/dmd.31.8.1035> PMID: 12867492

289. Chandrasekaran K, Swaminathan K, Chatterjee S, Dey A. Apoptosis in HepG2 cells exposed to high glucose. *Toxicol In Vitro* [Internet]. Elsevier BV; 2010 Mar;24(2):387–396. Available from: <http://dx.doi.org/10.1016/j.tiv.2009.10.020> PMID: 19892008
290. Chang-Liu CM, Woloschak GE. Effect of passage number on cellular response to DNA-damaging agents: cell survival and gene expression. *Cancer Lett* [Internet]. Elsevier BV; 1997 Feb 26;113(1-2):77–86. Available from: [http://dx.doi.org/10.1016/S0304-3835\(97\)04599-0](http://dx.doi.org/10.1016/S0304-3835(97)04599-0) PMID: 9065805
291. Papavlassopoulos H, Mishra YK, Kaps S, Paulowicz I, Abdelaziz R, Elbahri M, Maser E, Adelung R, Röhl C. Toxicity of functional nano-micro zinc oxide tetrapods: impact of cell culture conditions, cellular age and material properties. *PLoS One* [Internet]. Public Library of Science (PLoS); 2014 Jan 13;9(1):e84983. Available from: <http://dx.doi.org/10.1371/journal.pone.0084983> PMCID: PMC3890288
292. Moreno-Torres M, García-Llorens G, Moro E, Méndez R, Quintás G, Castell JV. Factors that influence the quality of metabolomics data in in vitro cell toxicity studies: a systematic survey. *Sci Rep* [Internet]. Springer Science and Business Media LLC; 2021 Nov 11;11(1):22119. Available from: <http://dx.doi.org/10.1038/s41598-021-01652-1> PMCID: PMC8586040
293. Wilkening S, Bader A. Influence of culture time on the expression of drug-metabolizing enzymes in primary human hepatocytes and hepatoma cell line HepG2. *J Biochem Mol Toxicol* [Internet]. Wiley; 2003;17(4):207–213. Available from: <http://dx.doi.org/10.1002/jbt.10085> PMID: 12898644
294. Sakai M, Okabe M, Tachibana H, Yamada K. Apoptosis induction by gamma-tocotrienol in human hepatoma Hep3B cells. *J Nutr Biochem* [Internet]. Elsevier BV; 2006 Oct;17(10):672–676. Available from: <http://dx.doi.org/10.1016/j.jnutbio.2005.11.001> PMID: 16517139
295. Hung WC, Chang HC, Chuang LY. Activation of caspase-3-like proteases in apoptosis induced by sphingosine and other long-chain bases in Hep3B hepatoma cells. *Biochem J* [Internet]. Portland Press Ltd.; 1999 Feb 15;338(1):161–166. Available from: <https://dx.doi.org/10.1042/bj3380161>
296. Lamboley C, Bringuier AF, Feldmann G. Induction of apoptosis in normal cultured rat hepatocytes and in Hep3B, a human hepatoma cell line. *Cell Biol Toxicol* [Internet]. 2000;16(3):185–200. Available from: <http://dx.doi.org/10.1023/a:1007611022477> PMID: 11032362
297. Lin SJ, Shu PY, Chang C, Ng AK, Hu CP. IL-4 suppresses the expression and the replication of hepatitis B virus in the hepatocellular carcinoma cell line Hep3B. *J Immunol* [Internet]. The American Association of Immunologists; 2003 Nov 1;171(9):4708–4716. Available from: <https://dx.doi.org/10.4049/jimmunol.171.9.4708> PMID: 14568946
298. Vivekanandan P, Thomas D, Torbenson M. Hepatitis B viral DNA is methylated in liver tissues. *J Viral Hepat* [Internet]. Wiley; 2008 Feb;15(2):103–107. Available from:

<http://dx.doi.org/10.1111/j.1365-2893.2007.00905.x> PMID: 18184192

299. Sainz B Jr, TenCate V, Uprichard SL. Three-dimensional Huh7 cell culture system for the study of Hepatitis C virus infection. *Virology* [Internet]. Springer Science and Business Media LLC; 2009 Jul 15;6(1):103. Available from: <http://dx.doi.org/10.1186/1743-422X-6-103> PMID: 19596477
300. Eyre NS, Phillips RJ, Bowden S, Yip E, Dewar B, Locarnini SA, Beard MR. Hepatitis B virus and hepatitis C virus interaction in Huh-7 cells. *J Hepatol* [Internet]. Elsevier BV; 2009 Sep;51(3):446–457. Available from: <http://dx.doi.org/10.1016/j.jhep.2009.04.025> PMID: 19596477
301. Li K, Prow T, Lemon SM, Beard MR. Cellular response to conditional expression of hepatitis C virus core protein in Huh7 cultured human hepatoma cells: Cellular response to conditional expression of hepatitis C virus core protein in Huh7 cultured human hepatoma cells. *Hepatology* [Internet]. Ovid Technologies (Wolters Kluwer Health); 2002 May;35(5):1237–1246. Available from: <http://dx.doi.org/10.1053/jhep.2002.32968> PMID: 11981774
302. Fujimoto Y, Itabe H, Sakai J, Makita M, Noda J, Mori M, Higashi Y, Kojima S, Takano T. Identification of major proteins in the lipid droplet-enriched fraction isolated from the human hepatocyte cell line HuH7. *Biochim Biophys Acta* [Internet]. 2004 Feb 2;1644(1):47–59. Available from: <http://dx.doi.org/10.1016/j.bbamcr.2003.10.018> PMID: 14741744