Integrated New Approach Methods Using Zebrafish (*Danio rerio*) Larvae and Transcriptomics Produce Points of Departure that are Protective of Chronic Toxicity Effects

Jory Curry

A thesis submitted in partial fulfillment of the requirements for the
Master's degree in Biology with collaborative specialization in Bioinformatics

Department of Biology
Faculty of Science
University of Ottawa

Abstract

There is growing concern that current toxicological test methods are too slow and expensive to evaluate the safety of the thousands of chemicals in the Canadian economy. In this thesis, a novel zebrafish embryo test, with integrated behaviour, energy expenditure and gene transcription assays, was used to assess the hazard of a diverse suite of 29 chemicals. I hypothesized that points of departure (PODs) from the integrated test would be protective of the long-term toxic effects of these chemicals. I found that: 1) integrating alternative test methods enhanced the sensitivity of the zebrafish embryo acute toxicity (FET) test, 2) integrated results provided a holistic understanding of potential mechanisms of action and effects, and 3) transcriptional PODs were protective of PODs from traditional long-term and short-term juvenile and adult fish toxicity tests reported in the literature. This integrated zebrafish embryo test is a sensitive, informative and protective chemical hazard screening tool.

Résumé

On s'inquiète de plus en plus du fait que les méthodes d'essai toxicologique actuelles sont trop lentes et trop coûteuses pour évaluer la sécurité des milliers de produits chimiques présents dans l'économie canadienne. Dans cette thèse, un nouveau test sur l'embryon de poisson zèbre, intégrant le comportement, la dépense énergétique et la transcription des gènes, a été utilisé pour évaluer le danger d'un ensemble diversifié de 29 produits chimiques. J'ai émis l'hypothèse que les points de départ (POD) du test intégré protégeraient des effets toxiques à long terme de ces produits chimiques. J'ai constaté que: 1) l'intégration de méthodes d'essai alternatives a amélioré la sensibilité du test de poisson, essai de toxicité aiguë au stade embryonnaire, 2) les résultats intégrés ont fourni une compréhension holistique des mécanismes d'action et des effets potentiels, et 3) les PODs transcriptionnels ont protégé les PODs des tests traditionnels de toxicité à long terme et à court terme sur les poissons juvéniles et adultes rapportés dans la littérature. Ce test intégré sur l'embryon de poisson zèbre est un outil sensible, informatif et protecteur de dépistage des risques chimiques.

Acknowledgements

I would like to extend my heartfelt thanks to Jason O'Brien and Jan Mennigen, who played a significant role in my academic journey. I deeply appreciate their unwavering support and mentorship over the last two years. Their encouragement and advice have been instrumental in shaping my research.

I am also indebted to Jason O'Brien, Doug Crump and Jan Mennigen for fostering welcoming and team-oriented laboratory environments. The Mennigen lab and ECCC Ecotoxicology laboratories were accommodating and supportive. Without the help of lab members and close colleagues, I would not be where I am today.

Tyler Nguyen from the ECCC EcoTox lab has been a pillar of support, especially during challenging times. Tyler has made countless sacrifices for this project and never turns down an opportunity to help others with his technical knowledge and expertise. Without Tyler, this project would not be possible. We spent countless hours in the laboratory and online doing RNA extractions, running exposure experiments and developing R scripts collaboratively. I want to express my gratitude for his unwavering assistance.

I would also like to thank Helina Gyasi from the ECCC EcoTox lab for her dedication to RNA extractions in the laboratory, a critical component of this research project. Her contributions have been indispensable, and I am thankful for her dedication.

Additionally, I would like to thank colleagues and collaborators who worked with me closely over the last two years: Hyojin Lee, Xia Pu, Divya Ramachandran and Giancarlo Talarico. Your teamwork and willingness to help enriched my research experience.

To the ECCC EcoTox lab, I want to convey my appreciation for maintaining the highest laboratory standards and providing an environment conducive to performing high-quality research. Your commitment to excellence has been invaluable to my academic journey.

I extend my gratitude to the esteemed members of my thesis defence committee, Professor Carol Yauk, Professor Stéphane Aris-Brosou and Professor Bill Willmore. I am grateful for their expertise and contributions to this academic milestone. Finally, I thank Compute Ontario and the Digital Research Alliance of Canada. As well as the Canadian Chemical Management Plan and NSERC for funding.

Table of Contents

List of Abbreviations and Definitions

[US] EPA	[United States] Environmental Protection Agency
[Z]FET	[Zebra]fish Embryo Acute Toxicity test
AIC	Akaike information criteria
ANOVA	Analysis of Variance
AOPs	Adverse Outcome Pathways
BMD	Benchmark dose
BMDExpress 3.0 (bioinformatics tool)	A tool for analyzing gene expression data that uses dose-response models from the United States Environmental Protection Agency Benchmark Dose Software and ToxicR. Outputs Benchmark dose values
BMD_L	Lower bound of the 95% confidence interval of the Benchmark Dose
BMD_U	Upper bound of the 95% confidence interval of the Benchmark Dose
BMR	Benchmark response
bp	base-pairs
BPA	Bisphenol A
BPAF	Bisphenol AF
CAS	Chemical Abstracts Service
CCAC	Canadian Council of Animal Care
cDNA	Complementary DNA
CEPA	Canadian Environmental Protection Act
CMP	Canadian Chemicals Management Plan
CPM	Counts per Million
DEGs	Differentially Expressed Genes

DESeq2 (R Bioconductor package)	Differential gene expression analysis based on the negative binomial distribution
DMSO	Dimethyl Sulfoxide
dpf	Days post fertilization
EC10	Effective concentration for 10% of the population
EC_{50}	Effective concentration for 50% of the population
ECCC	Environmental and Climate Change Canada
EDCs	Endocrine-disrupting chemicals
EtOH	Ethanol
$FADH_2$	Flavin adenine dinucleotide (hydroquinone form)
FastQC (bioinformatics tool)	A tool for quality control checking raw sequencing data
FDR	False discovery rate
$FMNH_2$	Flavin mononucleotide (reduced form)
GC bias	Uneven coverage of GC content
gDNA	Genomic DNA
GLM	Generalized linear model
GNU	GNU Project - GNU's Not Unix! (Collection of free software for operating systems)
GO (database)	Gene Ontology
GSEA	Gene Set Enrichment Analysis
HESI EnviroTox (database)	Health and Environmental Sciences Institute Environmental Toxicology database
hpf	Hours post fertilization
LC_{50}	Lethal dose for 50% of a population
LCRD tPOD	Lowest Consistent Response Dose Transcriptomic Point of Departure

LOEC	Lowest observed effective concentration
M [reads]	Million [reads]
MDCs	Metabolism-disrupting chemicals
MEPS	Mass Embryo Production System
mg/L (unit)	Concentration in milligrams per litre
min(s) (unit)	minutes
mm (unit)	millimeters
MOA	Mechanism of Action
mRNA	Messenger RNA
MSENR	Mexican Secretariat of Environment and Natural Resources
MultiQC (bioinformatics tool)	A tool that aggregates results from FastQC into a single report
NAAEC	North American Agreement on Environmental Cooperation
NAD[P]H	Nicotinamide adenine dinucleotide [phosphate]
$NaHCO_3$	Sodium bicarbonate
NAMs	New Approach Methodologies
nCov5	The number of genes with at least 5 mapped transcripts
Nextflow core RNAseq (bioinformatics pipeline)	Bioinformatics pipeline for analyzing RNA sequencing data with reference genomes and annotations
ng (unit)	Nanogram
nM (unit)	Nanomolar
NO[A]EL	No Observed [Adverse] Effect Level
NOEC	No Observed Effect Concentration
nSig80	The number of highly expressed genes comprising 80% of all quantified transcripts
NTP	National Toxicology Program of the US Departments of Health and Human Services

OECD	International Organisation for Economic Co-operation and Development
PAH	Polycyclic aromatic hydrocarbon
PCA	Principle component analysis
Phred Quality (Q) score	A measure of the quality of the identification of the nucleobases generated by automated DNA sequencing
POD	Point of departure
PolyA	Polyadenylation
preBMD	Preliminary BMD
Protective/Protectiveness	The term 'protective' and 'protectiveness' designates whether a POD is lower or equal to a particular effect level. Often this term is used to designate if a POD is lower or equal to a chronic toxicity effect level such as death, deformity, reproductive perturbations, cancer or endocrine disruption.
QC	Quality Control
R	Programming language for statistical computing and graphics
R-ODAF	Regulatory 'Omics Data Analysis Framework
RCurveP (R package)	Curve processing algorithm
RFUs (units)	Relative fluorescence units
RIN	RNA Integrity Number
rlog values	Regularized logarithm values
RLT	Qiagen RNeasy Lysis Buffer
RPE	Qiagen RNeasy Concentrated Wash Buffer
RPM	Revolutions per minute
rRNA	Ribosomal RNA
RW1	Qiagen RNeasy Wash Buffer 1

s (unit)	seconds
Salmon (bioinformatics tool)	A tool for fast transcript quantification from RNAseq data
SD	Standard Deviation
Sensitivity	The term 'sensitivity' is used in this thesis to designate the ability of a method to detect significant effects at lower concentrations relative to another method, not as the probability of a true positive test result
SortMeRNA (bioinformatics tool)	A local sequence alignment tool for filtering and mapping reads to ribosomal RNA database files
TDRM	Transcriptomic dose-response modelling
TG 236	International Organisation for Economic Co-operation and Development Test Guideline #236 (Fish Embryo Acute Toxicity test)
ToxRef (database)	United States Environmental Protection Agency Toxicity Forecast Reference database
tPOD	Transcriptomic Point of Departure
$tPOD_{10th}$	10^{th} Percentile of gene BMDs Transcriptomic Point of Departure
$tPOD_{20thGene}$	20^{th} gene BMD Transcriptomic Point of Departure
$tPOD_{mode}$	First mode in distribution of gene BMDs Transcriptomic Point of Departure
Trim Galore (bioinformatics tool)	A wrapper around Cutadapt and FastQC to consistently apply adapter and quality trimming to FastQ files
TSCA	US EPA Toxic Substance Control Act of 1976
Tukey's HSD	Tukey's honestly significant difference test
US EPA ECOTOX (database)	United States Environmental Protection Agency Ecological Toxicology database
WT	uOttawa Wildtype
µl (unit)	Microlitre

List of Tables

List of Figures

Statement of Contributions

Chapter 2. The Sensitivity of the Zebrafish Embryo Acute Toxicity (ZFET) Test is Increased with Added Swimming Behaviour and Oxidative Metabolism-dependent Energy Expenditure Endpoints

Study design	Jan Mennigen Jason O'Brien Florence Pagé-Larivière John Prindiville Doug Crump Phuoc Tyler Nguyen
Literature review	Jory Curry Florence Pagé-Larivière John Prindiville Phuoc Tyler Nguyen
Animal husbandry	uOttawa Animal Care and Veterinary Services
ZFET test (exposure, media preparation and renewal, observation scoring)	Jory Curry Tyler Nguyen Jan Mennigen
Swimming behaviour and alamarBlue® assays	Jory Curry Tyler Nguyen Jan Mennigen
Data analysis	Jory Curry
Manuscript preparation	Jory Curry Jason O'Brien Jan Mennigen

Chapter 3. Adding a Transcriptomic Endpoint to the ZFET Test Provides a POD that Protects Against Chronic Overt Toxicity Effects (death, deformity & reproduction) in Juvenile and Adult Fish

Study design	Jason O'Brien Jan Mennigen Florence Pagé-Larivière John Prindiville Doug Crump Phuoc Tyler Nguyen
Literature review	Jory Curry Florence Pagé-Larivière John Prindiville Phuoc Tyler Nguyen
Animal husbandry	uOttawa Animal Care and Veterinary Services
ZFET test (exposure, media preparation and renewal, observation scoring)	Jory Curry Tyler Nguyen Jan Mennigen
RNA Extraction	Jory Curry Helina Gyasi Phuoc Tyler Nguyen
RNA sequencing	Centre d'expertise et de services Génome Québec
Data analysis	Jory Curry Phuoc Tyler Nguyen
Manuscript preparation	Jory Curry Jason O'Brien Jan Mennigen

Chapter 1 – Introduction and background

1.1. The current state of toxicology testing

Importance

Chemicals are integral to the functioning of the economy, our homes, and our everyday lives. However, chemicals can harm us unknowingly, especially those with low-dose long-term effects from exposure. Chronic exposure to chemicals from everyday life threatens wildlife and humans alike. Therefore, assessors must efficiently and effectively characterize the risks associated with the rapidly expanding list of legacy and emerging chemicals. By characterizing the hazard of toxic substances, we can estimate how likely a chemical may be harmful and, thus, take the necessary steps to mitigate the harm to us and the environment. In the context of toxicology, the toxic risk is the likelihood that a particular substance will induce a hazardous effect given a specified exposure level. A crucial step of risk assessment is hazard characterization: identifying the potentially harmful effects that substances can induce. Identifying harmful effects is accomplished through toxicological testing.

The state of toxicological testing in North America

To assess the hazards and risks of toxic substances in North America, the United States of America, Mexico and Canada have collaborated through regulatory branches of government. North American departments include the United States Environmental Protection Agency (EPA), the Mexican Secretariat of Environment and Natural Resources (MSENR), and Health Canada & Environment and Climate Change Canada (ECCC). These government departments in North

America have made significant contributions to the field of toxicology since the Great Lakes Water Quality Agreement in 1972 (US and Canada) and the beginning of the North American Agreement on Environmental Cooperation (NAAEC) in 1994 (US, Canada & Mexico). The NAAEC has improved toxicity testing methods and protects vulnerable species through collaborative research and regulations (North American Agreement on Environmental Cooperation (NAAEC), 2012; Government of Canada and Government of the United States of America, 1972).

The US EPA Toxic Substance Control Act of 1976 (TSCA) contains a list of ~86,600 commercial chemicals as of August 2021. Until 2016, the TSCA list was estimated to grow by approximately 500-1000 chemicals annually on average since 1976 (Arnold, 2015; Basu *et al.*, 2019). In June 2016, the TSCA was amended and required the US EPA to update the legacy chemical inventory and report chemicals currently in use. Since then, the number of cases reported to the 'New Chemicals Review Program' has been growing by approximately 600-700 yearly, with approximately 42,000 commercial chemicals actively used today in the United States of America (US Environmental Protection Agency, 2019).

Similarly, the Canadian Chemicals Management Plan (CMP) is a government initiative to assess the risk of a list of 4,300 substances (triaged from a more extensive list of 23,000 substances in 2006) with the initial intent to have them assessed by 2020. The program has since been extended from the initial ending date of March 2021. According to the latest published progress report, over 3,600 chemicals have been assessed, and 456 were deemed harmful to the environment and human health (Government of Canada, 2019). The Chemistry Industry Association of Canada acknowledged the need to continue the CMP and, as of April 2021, proposed a budget of $476.7 million for a further five years to continue protecting Canadians and

the environment (Government of Canada, 2019). With a budget of just under $100 million per year for five years, it is apparent that it is very costly to conduct chemical risk assessments. The budget does not include the cost to the industry responsible for generating the toxicological test data used for risk assessments. A pesticide's average toxicity testing costs between $8 and $16 million (Price *et al.*, 2022).

The steep risk assessment cost is due primarily to the complexity of steps involved in a typical risk assessment under the Canadian Environmental Protection Act (CEPA):

1. Substance identification - information gathering on chemical properties and uses,
2. Hazard assessment - the focus of this thesis - assessment of environmental and health effects,
3. Exposure assessment - evaluation of potential routes and exposure levels,
4. Risk characterization - risk is characterized based on the hazard and exposure assessment results,
5. Risk management - efforts made to mitigate risk,
6. Monitoring and review - ensure risk remains at acceptable levels.

The large number of chemicals in commercial use globally, estimated to be over 350,000 by some researchers, highlights the urgent need for efficient and cost-effective methods of chemical hazard assessment (Woodruff *et al*., 2023). Traditionally, chemical hazard assessment methods rely heavily on animal testing. They are simply too costly and time-consuming to meet the global need to test existing and emerging chemicals and substances of concern. Alternative approaches to toxicity testing provide a quicker and more cost-effective way to assess the potentially toxic effects of these chemicals and reduce our reliance on animal use (Bernhardt *et al*., 2017; Harrill, Viant, *et al*., 2021; Harrill, Everett, *et al*., 2021).

Traditional test methods and chemicals of concern

The most common test methods currently used for ecological hazard assessments are based on overt toxicological effects (e.g. deformities and lethality) following short-term, high-dose exposures (i.e. acute toxicity tests). Acute toxicity experiments typically do not measure complex and subtle endpoints such as behavioural, reproductive, metabolic and sub-lethal developmental perturbations. Chronic toxicity experiments, on the other hand, are based on low-dose, long-term exposures and can provide rich and informative data that are more environmentally relevant. However, chronic tests are prohibitive in time and cost and are unethical because they require the long-term suffering of animals in laboratory settings (US EPA, 1994). For these reasons, collecting a large amount of data using these methods takes time, effort and money. The result is that most chemicals have very little or no data available regarding their long-term toxicity. Consequently, risk assessors must extrapolate chronic toxicity threshold values from traditional acute experiments (Malkiewicz *et al*., 2009). To account for the reduced sensitivity, risk assessors apply highly generalized "assessment factors" when extrapolating chronic toxicity from acute toxicity points of departure (Mount and Stephan, 1967).

An assessment factor is a numerical value used to account for uncertainties and variability in the data and to provide a margin of safety in risk assessment. These factors are based on the idea that long-term exposure to a substance may lead to additional effects even if the levels are below those that cause acute toxicity. Additional effects can be due to bioaccumulation or changes in response due to shifts in tolerance or sensitivity, for example (National Research Council, 2009; Canadian Environmental Protection Act, 1999). However, additional effects are complex. Additional effects are influenced by absorption, distribution, metabolism and excretion (i.e. toxicokinetics), dynamic interactions between a toxicant with a

biological target and its biological effects (i.e. toxicodynamics), exposure dose and duration, and individual susceptibility variability.

Because of the complexity of additional effects, there is no universal mathematical model for extrapolating results from acute tests to predict effects at lower concentrations after chronic exposures. Therefore, assumption factors must be made case-by-case based on empirical toxicity data. Thus, estimates of chronic toxicity based on assessment-factor extrapolations from acute toxicity tests are inaccurate (especially when empirical data is sparse) and may not be sufficiently protective of effects for specific chemicals.

Long-term effects are of particular concern for chemicals with low-dose chronic toxicity, such as the well-known endocrine-disrupting chemicals (EDCs) (e.g. bisphenol A – BPA). For example, consider the risk of BPA in aquatic environments: BPA has relatively low acute toxicity (adult zebrafish have an LC_{50} of 8.09mg/L), but BPA can have long-term endocrine disrupting effects as low as 1µg/L and neurodevelopmental effects as low as 1.5µg/L (Liu *et al.*, 2021; Kinch *et al.*, 2015). BPA is a concern given that environmental levels have been detected at concentrations greater than 1µg/L. For example, in Highland Creek, Beaverdams Creek, and Hamilton Harbour, Ontario, BPA has been detected at concentrations as high as 1.94 µg/L – 6.37µg/L (Environment Canada, 2020; Liu *et al*., 2021). In the case of BPA, an assessment factor of 10,000 is necessary to extrapolate a protective estimate for chronic toxicity from the acute test 50% lethality concentration (LC_{50}). Such high assessment factors are rarely applied in risk assessment. Therefore, more accurate estimates of chronic toxicity are required to ensure that chemical risk assessments are sufficiently protective.

Points of departure (PODs)

A point of departure (POD) is a dose at which a chemical begins to induce a biological effect in an exposed animal, cell, or tissue. Assessors use PODs to characterize chemical hazards in the calculation of risk quantitatively. PODs are derived for "apical" toxicity endpoints for acute toxicity testing, traditionally defined as lethality, deformities or reproductive failure *in vivo* (OECD, 2013). However, PODs can be determined for virtually any biological endpoint that can be measured, including molecular endpoints. PODs are calculated using various methods: the LC_{50} is a standard POD to describe lethality; the 50% effective concentration (EC_{50}) or lowest observed effect concentration (LOEC) are often used for non-lethal endpoints. More recently, the benchmark dose (BMD) approach has become the preferred method to derive PODs due to its more advanced characterization of the dose-response relationship (Izadi *et al*., 2012). Briefly, the BMD approach is a statistical method used in toxicology to estimate the dose-response relationship between chemical exposure and biological effects. The process involves modelling a dose-response curve to experimental data and estimating the dose associated with a specific benchmark response (BMR) threshold. Common BMRs are one standard deviation from the control group or a 10% effect threshold. The BMD POD is typically lower than traditional PODs, such as the LOEC, because it considers the shape of the dose-response curve, is not restricted to dose groups and provides a more accurate estimate of the dose at which a response occurs.

Apical endpoint PODs are generally used for quantifying the acute toxicity of chemicals. Apical endpoints are directly measurable whole-organism outcomes of exposure *in vivo*, generally defined by lethality, deformity, hatching success or reproductive failure (OECD, 2013). PODs for non-apical endpoints, such as molecular or behavioural perturbations, may be

more informative of potential mechanisms of action and environmentally relevant. Non-apical endpoints are an intermediate step at a level of biological organization below the apical endpoint, such as a biomarker, genomic/transcriptomic endpoint, behavioural perturbation, metabolic perturbation, or other sub-organism-level response (OECD, 2013). New approach methodologies (NAMs) are often employed to measure non-apical endpoints (e.g., behaviour perturbations, gene expression profiles, biomarkers) and derive non-apical PODs.

1.2. New approach methodologies – NAMs

Overview

In recent years, there has been a growing interest in using alternative testing methods, such as *in vitro* and computational models, to reduce animal testing and increase efficiency. In 2019, the EPA launched its new strategic plan for evaluating the safety of chemicals, emphasizing using alternative testing methods and prioritizing chemicals with the most significant potential for risk (US EPA, 2018). Similarly, CEPA has promoted alternative testing methods in Canada and supported research in this area since as early as 2001 (Canadian Environmental Protection Act, 1999). Recently, as of February 2022, with royal ascent in June 2023, an amendment to CEPA provided legislative backing for alternative testing methods in Canada (Bill S-5, 2023).

NAMs are any alternative method or strategy that avoids using intact animals to inform risk assessment and hazard identification, as well as different approaches for decision-making (US EPA, 2018). This broad definition encompasses different types of *in vitro*, *in silico* and *in vivo* methods that ultimately fall into four categories: 1) chemical characterization, 2) hazard identification and characterization, 3) dosimetry and 4) *in vitro – in vivo* extrapolation (Harrill,

Viant, *et al*., 2021; US EPA, 2018). The long-term motivation for the implementation of NAMs is to: 1) better adhere to the 3 R's (reduce, replace, refine) of humane animal experimentation, 2) accelerate the pace of chemical risk assessment for data-poor chemicals and fill in gaps for data-rich chemicals, and 3) improve or maintain the quality of scientific data (US EPA, 2018, 2019). Examples of NAMs include but are not limited to: read-across (expert-driven - toxicity prediction based on the expertise of toxicologists for similar chemicals and category-driven - toxicity prediction based on predefined groups/categories of toxic chemicals), transcriptomics (short-term early life stage *in vivo* assays or *in vitro* assays), high-throughput bioactivity assays and adverse outcome pathways (AOPs).

Limitations

Although there are undoubtedly many advantages of using NAMs (i.e. reducing costs, shorter experiments, using fewer animals, higher throughput assays), many limitations come along with developing any immature method. Namely, there needs to be more standardization and formal validation, which has hampered the adoption of NAMs for risk assessment applications. It is necessary to correlate both *in vitro* data and computational models to real and complex *in vivo* quantitative experimental toxicity data (Harrill, Everett, *et al*., 2021). However, there is potential for the widespread adoption of NAMs as they mature and reach their full potential. Two critical steps in helping NAMs to mature are generating data to support standardization and comparing the performance of NAMs to more accepted approaches (i.e. validation).

1.3. Zebrafish embryo model for twenty-first-century toxicity testing

The zebrafish (*Danio rerio*) embryo model refers to early-stage zebrafish from 0 to 3 days post-fertilization (dpf) and often also includes early larval stages up until 7dpf. The Canadian Council of Animal Care (CCAC) provides guidelines for the care and use of animals in research. Zebrafish embryos rely on their yolk sac and do not independently feed (at approximately 7dpf, larvae independently feed). Thus, they are exempt from CCAC regulations and considered an animal-alternative model, a major distinction from an animal-free model (Cassar *et al.*, 2020).

Zebrafish embryos have been a popular model for toxicity testing since the 1990s. When Charles B. Kimmel *et al*. from the University of Oregon published "Stages of embryonic development in the zebrafish" in Developmental Dynamics, the journal article had a massive impact on the popularity of the model species (Kimmel *et al*., 1995). The Zebrafish Genome Project sequenced and annotated the zebrafish genome (GRCz10) from 2001 to 2013 using the Tübingen strain (Howe *et al*., 2013). In the last decade, these major milestones in zebrafish embryo research have allowed toxicologists to identify genes and pathways involved in developmental toxicity and characterize developmental abnormalities and defects induced by toxicants.

There are several advantages to the zebrafish embryo model. First, they are inexpensive, easy to breed, develop rapidly and have high fecundity, lending to their high-throughput potential. Second, they are transparent during early development, making them easy to characterize when exposed to toxins to which they are sensitive. Third, they have a high degree of homology to other aquatic vertebrates and terrestrial animals, including humans, which makes them a suitable model for biological translation across species (Howe *et al.*, 2013; Sipes *et al.*,

2011; OECD, 2013). Zebrafish maintain a full range of conserved genes associated with chemical metabolism and biotransformation (i.e. cytochrome p450 genes), similar to humans, which is relevant for toxicology testing (Bambino and Chu, 2017). Fourth, compared to *in vitro* methods, the zebrafish embryo has the advantage of complex tissue and cell type interactions more representative of real biological interactions in nature. Finally, one of the greatest benefits of using the zebrafish embryo model is that it is a more humane model than using adult fish or mammals for toxicity experiments (because of the reduced experimental duration and the non-invasive nature of a bathing exposure).

However, one of the challenges with using zebrafish embryos as a screening tool in the regulatory context is a lack of experimental standards and designs, inconsistent data analysis strategies, and less complex tissues than adults (Hsieh *et al.*, 2019).

Fish embryo acute toxicity (FET) test

Developmental toxicity/embryotoxicity refers to a chemical's ability to perturb the normal development of an embryo. Embryotoxicity can manifest as malformations, growth retardation, behavioural perturbations, and mortality. The International Organisation for Economic Co-operation and Development (OECD) has collaborated with regulatory agencies to design standardized international guidelines for toxicity testing. The OECD is viewed as the international authority on toxicological test standardization. The Fish Embryo Acute Toxicity (FET) test (OECD TG 236) is an internationally standardized guideline released in 2013 that details methods to determine the acute toxicity of chemicals exposed to fish embryos using lethality/mortality as the main apical endpoint (OECD, 2013). The FET test is often conducted with zebrafish and is called the Zebrafish Embryo Acute Toxicity (ZFET) test. The ZFET test is a widely used alternative test method for the acute toxicity assessment of chemicals. The test

protocol involves exposing fertilized fish eggs to varying concentrations of chemicals of interest over 96 hours post-fertilization (hpf) during early development. Mortality is quantified each day, along with media renewal. The ZFET test is generally considered a more humane and cost-effective alternative to traditional acute toxicity tests (on juveniles and adults) since it requires few animals (i.e., only breeding pairs, which are not exposed to any substances) and less time. Traditional mortality and deformity endpoints of developmental toxicity are common and have established analysis methods (von Hellfeld *et al.*, 2020). However, non-traditional endpoint analyses such as swimming behaviour or oxidative metabolism-dependent energy expenditure are still under development (Hsieh *et al*., 2019; Renquist *et al*., 2013). Moreover, the ZFET (OECD TG 236) test does not include endpoints related to sub-lethal deformities and malformations. Quantifying both sub-lethal and lethal endpoints can substantially increase the sensitivity of the ZFET protocol (von Hellfeld *et al.*, 2020). There has been a growing interest in integrating neurodevelopmental assessment and transcriptomic techniques into the ZFET test.

In Chapters 2 and 3, I extended the exposure duration and development time to 120hpf instead of 96hpf because, at 120hpf, zebrafish embryos have full biotransformation capabilities, increased general activity, fully inflated swim bladders and fully developed primary and secondary islets of the pancreas (Kimmel *et al*., 1995, Ried *et al*., 2018). Because the zebrafish embryo is metabolically and cognitively competent at 120hpf (Kimmel *et al*., 1995, Ried *et al*., 2018, Renquist *et al*., 2013, Kozol *et al*., 2016, Tal *et al*., 2023, Klüver *et al*., 2015), we also expected to see more diversity in differentially expressed genes compared to 96hpf. At 120hpf rather than 96hpf, the zebrafish embryo is better suited for the assays incorporated in Chapters 2 and 3.

Neurodevelopmental toxicity assessment

Neurodevelopmental toxicity assessment is a crucial aspect of ecotoxicology that has received limited attention due to the challenges of conducting such tests. The current lack of attention in assessing neurodevelopmental toxicity is concerning, given that there are estimated to be over 350,000 chemicals and chemical mixtures in commerce globally, which is growing each year. Their potential effects on environmental health are mostly unknown (Woodruff *et al*., 2023). According to the US EPA's ToxRef database, fewer than 100 unique chemicals have been assessed for developmental neurotoxicity (Watford *et al.*, 2019; Tal, 2023). The lack of neurotoxicity assessment data in the ToxRef database indicates the pressing need for better methodologies that can be used to assess neurodevelopmental toxicity, which is especially critical considering the potential long-term health consequences that exposure to toxic chemicals may pose to ecosystems.

Developmental neurotoxicity assessments have clear guidelines for adult rodent models (OECD TG 426). However, as stated previously, these animal toxicity assays are time-consuming, expensive, and often raise important ethical concerns (OECD, 2007). Fortunately, the emerging zebrafish embryo model provides a promising solution to these challenges.

Zebrafish embryos have a fully developed and capable nervous system with a blood-brain-barrier within 72hpf and are a useful model to evaluate the ecological hazard of potential neurodevelopmentally active chemicals (Kimmel *et al*., 1995; Kozol *et al*., 2016). They are also diurnal, meaning homeostatic and circadian mechanisms regulate their behavioural responses. The highly conserved circadian system of zebrafish makes it possible to monitor stress-induced changes in their behaviour patterns as early as 3-4dpf when exposed to toxic chemicals and alternating photoperiods (Muriana *et al.*, 2022). Swimming behaviour is an essential aspect of

zebrafish embryos' development, and changes in swimming behaviour can indicate neurotoxicity (or other effects on bone, muscle, and fin development, for example). By using automated neurobehavioural tests that alternate light and dark periods to induce repeatable changes in locomotory behaviour, it is possible to observe changes in swim behaviour (hyper- or hypo-activity) in response to exposure to chemicals and determine the potential hazard to the developing nervous system. Observing swimming behaviour as an outcome of neurodevelopment is a major advantage of this model, especially when compared to *in vitro* methods because swimming behaviour is a complex and observable outcome impossible to emulate using *in vitro* methods. However, the OECD has yet to outline any clear internationally standardized technical guidelines on assessing developmental neurotoxicity in zebrafish embryos for chemical hazard assessment, leading to some reasonable doubt about the validity of such methods in the past (Tsuji and Crofton, 2012). Nevertheless, the OECD is actively reviewing the validity of the (light-dark) swimming behaviour assay, which is promising for the future of this animal-alternative developmental neurotoxicity model (Tal, 2023; Klüver *et al*., 2015).

Because of the lack of clear internationally standardized guidelines from the OECD, researchers are still developing assays to assess the neurotoxicity of hazardous substances in the environment. Recent research has demonstrated the utility of zebrafish embryos in neurotoxicity assessments. For example, several studies showed that exposure to a widely used herbicide - Glyphosate (the active ingredient in RoundUp®) - led to significant behavioural alterations and neurotoxicity in zebrafish embryos at concentrations as low as 0.01mg/L (Lanzarin *et al*., 2020; Bridi *et al*., 2017; Roy *et al*., 2016). Another study found that zebrafish embryos are sensitive to other neurotoxins such as benzo[a]pyrene, a polycyclic aromatic hydrocarbon (PAH) produced from incomplete combustion of organic material/fossil fuels that are ubiquitous around the globe.

This study identified a hyperactive larval photo-motor response phenotype from exposure to benzo[a]pyrene using the (light-dark) swimming behaviour assay (Knecht *et al*., 2017). Other researchers have found that zebrafish embryos exposed to the selective serotonin reuptake inhibitor Fluoxetine have their locomotor activity affected (reduced) at concentrations as low as 0.00088mg/L (Airhart *et al.*, 2007; Al Shuraiqi *et al*., 2021; De Farias *et al.*, 2019; Zindler *et al*., 2020). The evidence shows that swimming behaviour perturbation is a reproducible endpoint to measure neurodevelopmental toxicity.

Efforts have been made to develop a standardized method of applying the benchmark dose analysis to zebrafish embryo behavioural and developmental toxicity data. With the support of the National Toxicology Program of the US Departments of Health and Human Services (NTP), authors Hsieh *et al*. proposed a harmonized analysis approach. They developed a software package called 'RCurveP' - 'R Curve Processing' - for the popular open-source statistical software R (Hsieh *et al*., 2019). The proposed approach transforms the complex multidimensional data from the zebrafish behavioural and developmental toxicity test to fit common dose-response models such as the Hill equation model. A behavioural POD can be determined using this BMD method. It also provides a non-parametric modelling approach and a novel empirical approach to selecting the BMR threshold based on the intrinsic overall variance in the response endpoint (Hsieh *et al*., 2019). The authors propose that because this approach is generalizable across all endpoints, it may be suitable for standardization and systematic comparisons of data sets.

In conclusion, the zebrafish embryo model offers a promising approach to assessing neurodevelopmental toxicity. Embryos provide a valuable alternative to traditional animal models, with the advantage of a fully developed and entire nervous system that is observable

through swim behaviour. As demonstrated by recent research, the zebrafish embryo model provides a useful tool for evaluating the biological hazard of neurotoxicants (or other compounds affecting swimming behaviour) and will ultimately help risk assessors and regulators assess the risk of toxins and develop effective strategies to protect the environment.

Oxidative metabolism-dependent energy expenditure assessment in toxicology

Metabolism-disrupting chemicals (MDCs) such as pesticides, heavy metals, plastics, PAHs, EDCs, perfluoroalkyl substances, and nanoparticles can disrupt energy homeostasis in organisms, leading to changes in energy expenditure, which can ultimately affect their growth, and development, and reproduction (Sun *et al*., 2022). The mechanisms by which these chemicals disrupt metabolism are not fully understood (Sun *et al*., 2022). Therefore, continued research on the toxicology of MDCs is needed to understand their hazards better and develop effective strategies to protect the environment.

However, there are no standardized international guidelines for assessing metabolism-dependent energy expenditure disruption in zebrafish or other animals. Despite the lack of standardized guidelines, researchers have been developing alternative methods of assessing metabolic disruption for years. A decade ago, the high-throughput and non-invasive alamarBlue® assay (also known as the resazurin assay) was developed by two laboratory groups at the University of Arizona and Vanderbilt University School of Medicine to measure oxidative metabolism-dependent energy expenditure changes in response to chemical exposure. Although the alamarBlue® cell viability assay has been used for *in vitro* experiments since 1993, researchers developed a novel approach to screen for metabolic disrupting effects *in vivo* with zebrafish embryos (Page *et al*., 1993). The alamarBlue® assay *in vivo* was a significant

contribution to toxicology and biomedical research because it utilized a sensitive animal-alternative approach to monitor changes in energy expenditure in response to MDC exposure that was quick and high-throughput, unlike the primary rodent model at the time (Renquist *et al*., 2013; Williams and Renquist, 2016). Briefly, this assay quantifies energy expenditure by measuring changes in fluorescence over time by taking advantage of the properties of resazurin (alamarBlue®) and its reduced form resorufin via nicotinamide adenine dinucleotide (NADH). A few researchers from the University of Toronto and the University of Alabama at Birmingham made a significant effort to validate the zebrafish *in vivo* alamarBlue® assay in 2018. They found that the methods could effectively measure metabolic changes in zebrafish embryos and adults (Reid *et al*., 2018). The researchers demonstrated that the assay is inexpensive, non-invasive, accurate, quick, and much less labour-intensive than other methods of measuring metabolic rate, such as metabolic chambers.

Since the inception of the zebrafish embryo *in vivo* alamarBlue® assay, researchers have been making significant contributions to the field of toxicology. Several studies have shown that the zebrafish *in vivo* alamarBlue® assay is an effective screening tool for assessing whether substances are metabolically disruptive. Recently, the *in vivo* zebrafish alamarBlue® assay has been found to effectively detect changes in energy expenditure after developmental exposure to nonylphenol polyethoxylates (a type of surfactant used in consumer products), per- and polyfluoroalkyl substances (used in consumer products resistance to water, oil, stains), silver nanoparticles (used in paints), polystyrene nanoplastics, and flame retardant BDE-47 (Chackal *et al*., 2022; Tu *et al*., 2019; Mittal *et al*., 2021; Kassotis *et al*., 2022). However, the methods used in these studies could only calculate a statistically significant LOEC and not a BMD. Chapter 2 applies the 'RCurveP' BMD method to the energy expenditure endpoint.

Overall, assessing oxidative metabolism-dependent energy expenditure is essential in understanding the potentially toxic effects of environmental pollutants on animals. Using zebrafish embryos and the alamarBlue® assay in ecotoxicological studies provides a promising and efficient approach to assessing the effects of environmental pollutants on animal and environmental health while reducing the ethical considerations associated with traditional animal models.

1.4. Transcriptomics in toxicology

Transcriptomics involves analyzing a snapshot of the entire set of RNA transcripts (the transcriptome) at a given time in a cell or organism's life. Transcriptomics can characterize the toxicity of chemicals by identifying changes in gene expression that occur in response to exposure. Differential gene expression can provide insights into underlying molecular mechanisms of action to help us understand how a chemical may cause toxic effects. Other major advantages of transcriptomics in toxicological research are: 1) rapid data acquisition in bulk, 2) declining costs and increasing efficiency of using such technology, and 3) molecular effects are technically observable earlier than chronic apical effects. One of the most common transcriptomic techniques involves RNA sequencing (RNAseq) to quantify and compare the levels of RNA transcripts between different samples. Researchers can identify differentially expressed genes (DEGs) associated with toxicity by comparing gene expression profiles of exposed samples at varying concentrations. Transcriptomic PODs and biomarkers can help to inform the development of safer chemicals and help to predict the toxicity of similar chemicals that may affect the same pathways.

Transcriptomic approaches have become important toxicity test methods. Standard experimental designs, data analysis, and reporting frameworks are beginning to emerge (Harrill, Viant *et al.*, 2021). The NTP published a report in 2018 which outlines an approach for genomic dose-response modelling largely based on the original works of Russel Thomas *et al*. from 2013 (National Toxicology Program, 2018; Thomas *et al*., 2013). These articles outline a transcriptional approach to toxicology testing using dose-response changes in gene expression to derive a POD based on a benchmark dose analysis of DEGs.

Transcriptomic dose-response modelling

Transcriptomic dose-response modelling (TDRM) expands upon the BMD approach by mathematically describing genome-wide changes in gene expression across a range of dose groups. When designing an experiment for the TDRM method, a large range of doses is recommended to capture the entire shape of the dose-response curve and maximize the accuracy of the BMDs (National Toxicology Program, 2018). Gene expression data can be acquired through various methods (i.e., RNAseq, TempO-Seq, Affymetrix microarray, etc.). Briefly, the general method entails: 1) determining adequate signal in the data (ANOVA with Benjamini and Hochberg False Discovery Rate < 0.05), 2) filtering out genes that do not respond to the treatment (William's trend test with variable fold change filter), 3) fitting responsive genes to different dose-response models and calculating a BMD from the best fit 4) collectively evaluating all of the gene-specific BMD values to determine a genome-wide transcriptomic point of departure (tPOD). There are two main approaches to collectively evaluating the gene-specific BMDs to determine a tPOD: 1) grouping gene-specific BMDs into gene sets or 2) examining the overall distribution of gene-specific BMDs. These two approaches will be described in more detail below. Regardless of how it is calculated, a tPOD represents the minimal dose that

produces a detectable change in gene expression, which in theory, should be the lowest dose associated with any possible toxicological hazard. The tPOD may provide a more accurate and environmentally relevant estimate of chronic toxicity than traditional approaches that extrapolate from acute toxicity methods using assessment factors. There is growing evidence that tPODs from short-term exposure studies are highly correlated to PODs from traditional long-term toxicity tests, suggesting they might be suitable for estimating chronic toxicity in risk assessment (Johnson *et al*., 2022). The resulting data can also be used to elucidate genes, pathways, and biological effects of each chemical analyzed, hence estimating the hazard for each gene/pathway. Transcriptomics can reduce animal testing since molecular data can be collected *in vitro* or with other alternatives.

Gene-level distribution-based tPOD derivation

Gene-level tPOD derivation methods take advantage of the distribution of BMDs calculated for each gene and subsequently identify the dose at which global gene expression is significantly affected. These gene-level tPOD derivation methods use the distribution of all gene BMDs to determine a tPOD. Some examples include the 10th percentile of gene BMDs ($tPOD_{10th}$), 20th gene BMD ($tPOD_{20thGene}$), and the first mode of the BMD distribution ($tPOD_{mode}$) (Pagé-Larivière *et al*., 2019). Theoretically, these tPODs should protect organisms from all possible hazards if exposure does not exceed the tPOD. The gene-set tPOD derivation approach, discussed in the next section, evaluates the adverse outcomes of chemical exposure by modelling the dose-response relationship at the biological pathway level. Conversely, the pathway-agnostic distribution-based gene-level tPOD approach evaluates the chemical hazards by modelling the dose-response relationship at the biological gene level. This pathway-agnostic approach is expected to be more protective because the POD estimates the threshold level

beyond which the perturbation of global gene expression manifests, irrespective of specific pathways.

Interpreting these changes in gene expression from a pathway-agnostic perspective can be challenging, so interpreting biological effects through pathway analyses is often conducted in parallel. Note that although tPODs are experimentally derived and represent the estimated dose at which changes in gene expression might occur for a given chemical, they are not yet a prediction tool for specific adverse outcomes. TDRM is a tool to generate a POD that estimates the level of toxicity. Several studies have demonstrated a correlation between short-term changes in gene expression and long-term effects in adults (Farmahin *et al*., 2019; Pagé-Larivière *et al*., 2019; Thomas *et al*., 2013; Webster *et al*., 2015). Thus, tPODs can be extremely useful for risk assessment because they provide a single, highly protective POD for any analysis.

Gene set-level-based tPOD derivation

It is common for researchers to concurrently conduct gene set enrichment analyses (GSEA) to provide biological interpretations for the resulting gene expression changes after chemical exposure. However, it should be noted that the primary objective of the GSEA in this context is to derive tPODs. The primary objective is not to link enriched gene set observations to higher-order phenotypes such as swimming behaviour, overt toxicity, or oxidative metabolism-dependent energy expenditure. However, it should be acknowledged that linking changes in gene expression to adverse outcomes is likely necessary to accept transcriptomic methods in formal risk assessments (National Toxicology Program, 2018).

GSEAs can be performed using reference databases such as the Gene Ontology (GO) or REACTOME databases. The GO database contains annotated biological terms related to different biological ontology domains (i.e., cellular components, molecular functions and

biological processes). For example, the annotated GO term ‘heart contraction’ relates to the biological process domain. Likewise, the REACTOME database contains annotated biological pathways (e.g., metabolism, regulation of gene expression, transmission of signals, etc.). A common statistical method utilized for GSEAs is Fisher’s exact test. Briefly, GSEA with Fisher’s exact test involves 1) selecting a set of genes that share a specific biological function (e.g., stress response), 2) ranking genes based on how much expression changes with different chemical doses, 3) determining the number of genes from the selected gene set that are among the top-ranked genes, which respond the most to chemical exposure (minimum 3 genes with BMDs and 5% populated only), 4) using Fisher’s exact test ($\alpha = 0.05$) to see if the overlap between the selected gene set and the top-ranked genes is more significant than random chance, 5) filtering by p-value to identify enriched gene sets that potentially play a role in responding to chemical exposure (National Toxicology Program, 2018). The lowest-potency median enriched gene set BMDs is then used to derive the gene set-level tPOD.

Transcriptomics in the zebrafish embryo model

Using zebrafish embryos for TDRM and DEG analysis can provide a cost-effective, sensitive, and internationally recognized medium-throughput approach for studying the effects of toxic substances on gene expression and developmental outcomes with ecological and human relevance. Although it is beneficial to use TDRM to estimate toxic effect levels using zebrafish embryos to assess the hazard of toxic substances, there are many challenges.

First, there is a real challenge in translating biological effects in the embryo stage to the adult stage of animals. Translating biological effects from the embryo stage to the adult stage of animals is a complex task due to the multifaceted nature of development, temporal dynamics of gene expression, gene regulation, interactions between tissues, and environmental factors, for

example. The embryo stage is a period of rapid development, and it is more sensitive to the effects of chemical exposure than the adult life stage. Second, TDRM has been tested until now, mostly in mammals. This poses additional challenges that need to be addressed, such as how chemicals are absorbed into the body (fish are aquatic and how chemicals interact with organisms in different environments can affect gene expression). Third, RNA cannot be easily extracted from specific tissues since zebrafish embryos are very small. As a result, RNA is usually extracted from whole-embryo homogenates, making it challenging to discern tissue-specific effects. Fourth, extracting sufficient RNA from individual embryos for most transcriptomic applications is often challenging. Therefore, pooling individuals together in a homogenized sample for RNA extraction is often necessary. However, because whole embryos are homogenized and pooled, transcript abundance is (allometrically) biased by the most abundant cells and tissues (e.g., muscle cells, neurons, blood cells, epithelial cells) and the largest size or most developed individuals in the pool. Fifth, there is a need for a standardized international approach for conducting transcriptomic dose-response modelling in zebrafish embryos. Thus, it is challenging to compare results across studies, which limits the reliability and reproducibility of the data. Sixth, there is limited regulatory acceptance of transcriptomic dose-response modelling in general. As such, the integration of these methods needs to be validated before they can be widely accepted. Thus, further data is required to determine if TRDM using zebrafish embryos can provide suitably protective PODs for ecological risk assessment. Finally, due to an evolutionary genome duplication event, zebrafish and other teleost fish genomes contain paralogous genes (Glasauer and Neuhauss, 2014). Paralogous genes can have different functions (neo-functionalization), or each function as a sub-unit of a gene partition that controls a specific function (sub-functionalization). Therefore, to fully understand the differential gene

expression of paralogous genes, it is necessary to reference annotation databases such as Ensembl or RefSeq and group paralogous genes together. Researchers must exercise caution when interpreting gene expression changes because many factors affect the biological significance of results. Post-transcriptional mechanisms such as alternative splicing, mRNA stability and translation efficiency affect protein concentrations. Post-translational modifications and protein degradation regulation can also affect the concentration of proteins, which play a crucial role in cellular functioning and adverse outcomes (Vogel and Marcotte, 2012). Epigenetic factors affect gene expression through DNA methylation, histone modification, chromatin remodelling and non-coding RNA molecules (Choi, 2010). Therefore, differential gene expression analyses must be corroborated with multiple assays assessing biologically significant toxicity endpoints. Multi-technique approaches help to validate transcriptomic results and provide a more holistic understanding of adverse outcomes, which is necessary for regulatory science.

1.6. Thesis objective, hypothesis and predictions

I hypothesize that including swimming behaviour, oxidative metabolism-dependent energy expenditure and transcript-level gene expression endpoints in the ZFET test will make it more sensitive, protective and informative than when only standard mortality and deformity endpoints are considered. I predict that PODs from swimming behaviour, oxidative metabolism-dependent energy expenditure and transcript-level gene expression assays will be lower than experimental PODs from the overt toxicity (mortality and deformity) endpoint and provide information on the potential mechanisms of action. I also predict that PODs from the transcript-level gene expression assays will be lower than literature-based PODs from juvenile and adult

toxicity experiments measuring mortality, development, growth and reproduction. In other words, the experimental transcriptomics endpoint will provide a more protective POD than literature-based acute and chronic toxicity effects from traditional juvenile and adult fish toxicity tests. This project aims to demonstrate that the zebrafish embryo model, combined with overt toxicity, swimming behaviour, oxidative metabolism-dependent energy expenditure and transcriptomic assays, can be a sensitive, informative, and protective animal-alternative screening tool for chemical hazard assessment. Overall, an integrated approach to the ZFET test can provide valuable insights into the mechanisms of toxicity and potential health risks of environmental exposures and establish confidence in using transcriptomic methods for hazardous chemical screenings and formal risk assessments.

Chapter 2 - The Sensitivity of the Zebrafish Embryo Acute Toxicity Test is Increased with Added Swimming Behaviour and Oxidative Metabolism-dependent Energy Expenditure Endpoints

2.1. Abstract

High-risk persistent and bioaccumulative contaminants pose a growing concern due to their potential for chronic health issues, including neurotoxicity and disruptions in hormones and metabolism. New Approach Methods (NAMs) like the zebrafish embryo swimming behaviour and alamarBlue® assays provide ecological and human health-relevant alternatives to traditional juvenile and adult fish toxicity testing for hazardous chemical screening. Zebrafish embryos were exposed to 29 chemicals with varying mechanisms of action and potencies for 5 days (120 hours post-fertilization) per the Organisation for Economic Co-operation and Development Fish Embryo acute Toxicity (FET) test (TG 236) guidelines, with a 2 day extension. Death and deformity (overt toxicity), swimming behaviour and energy expenditure endpoints were measured. Points of departure (PODs) were derived via an ANOVA and post hoc Dunnett's test and benchmark dose-response modelling per the NTP 'RCurvep' bootstrap method. Swimming behaviour and energy expenditure PODs were compared to overt toxicity PODs to determine if swimming behaviour and energy expenditure PODs were lower. Furthermore, the analysis extended to assess whether experimental PODs were lower than literature-based PODs. Twenty-

four chemicals significantly affected zebrafish embryos after 5 days of exposure (i.e., 2,4-DMP, BPA, BPAF, SDS and TEG did not significantly affect embryos). PODs from the swimming behaviour and oxidative metabolism-dependent energy expenditure endpoints were, on average, 32X lower than the classical death and deformity endpoint for 7 chemicals (17b-estradiol, 4-TPP, 4-TOP, Clofibric acid, DES, Fluoxetine and Flutamide). PODs from 3 chemicals (4-TPP, Clofibric acid and Flutamide) protected the literature-based chronic POD from traditional juvenile and adult fish toxicity tests. Adding zebrafish embryo swimming behaviour and oxidative metabolism-dependent energy expenditure endpoints to the extended FET test enhanced sensitivity and revealed insights into potential mechanisms of action. This supplemented FET test is a valuable screening tool for neurotoxic, developmental and metabolism-disrupting chemicals.

2.2. Introduction

The issue of harmful substances in our environment is a matter of utmost importance that has global implications to ecological health. The widespread use of synthetic chemicals in various consumer goods and industrial products has led to the introduction of over 350,000 synthetic chemicals worldwide, with approximately 500-1000 new substances added yearly (Bernhardt *et al.*, 2017; Arnold, 2015; Basu *et al.*, 2019; Woodruff *et al.*, 2023). While Canada's current toxicity screening process is commendable, it is insufficient to keep pace with the rapid development of new substances; Canada can screen approximately 275 contaminants yearly using current toxicity screening methods, and a backlog of contaminants that require toxicity screening has formed (Government of Canada, 2019; Basu *et al.*, 2019). Furthermore, animal testing for traditional toxicity screens is not only a time-consuming and expensive process, but it

also raises ethical concerns. Consequently, toxicologists are faced with the challenge of identifying hazardous substances in an ethical, timely and efficient manner.

In an effort to reduce the use of animals in research, the field of toxicology has shifted towards New Approach Methodologies (NAMs). Many NAMs are based on cell culture or cell-free *in vitro* techniques or *in silico* simulations to avoid using animals. However, the biological relevance of these approaches may be challenging to interpret due to the lack of complex cellular and tissue-type cross-talk. For example, organismal behaviour emerges at higher levels of cellular organization and can not be easily assessed using *in vitro* or *in silico* methods. To address this challenge, the zebrafish embryo model offers a promising animal-alternative approach to toxicity testing, as it is exempt from the Canadian Council of Animal Care (CCAC) regulations and is a high-level-organism with complex cellular and tissue-type cross-talk (Achenbach *et al.*, 2020).

Due to the popularity of the zebrafish embryo model, the Organisation for Economic Cooperation and Development (OECD) has developed an internationally standardized Zebrafish Embryo Acute Toxicity (ZFET) test (OECD TG236). It has emerged as an internationally widespread animal-alternative toxicity test due to its cost-effectiveness and time-saving advantages over traditional adult fish toxicity tests. In comparison to traditional juvenile and adult animal toxicity tests (OECD TG203, OECD TG230 & EPA 850.1500), on average, the ZFET test saves approximately 79% in monetary cost, takes approximately 88% less time to conduct and increases throughput by approximately 87% (Mittal *et al.*, 2022). However, it is essential to note that the ZFET test has limitations, primarily in its ability to evaluate sub-lethal effects or identify critical mechanisms of action (Sobanska *et al.*, 2018; von Hellfeld *et al.*, 2020). As such, while it is a valuable tool for initial toxicity screening, its results should be

complemented with additional testing that can more comprehensively assess the potential risks of chemical exposure.

In addition to the emergence of NAMs, there has been a significant shift in the regulatory approach towards quantitative chemical hazard characterization by agencies such as Environment and Climate Change Canada (ECCC), Health Canada and the US Environmental Protection Agency (US EPA). These agencies have moved away from older quantitative points of departure (PODs), such as the statistically lowest observed effect concentration, LOEC, towards methods that consider the full dose-response relationship of test chemicals, such as the benchmark dose (BMD) approach. The primary objective of this shift is to enhance the reproducibility of results through the use of a consistent and biologically relevant benchmark response (BMR) threshold level and address shortcomings of the more traditional PODs. Some of the shortcomings of traditional PODs that the BMD approach addresses include but are not limited to, 1) a lack of consideration of the dose-response trend, 2) limitations to experimentally tested concentrations or doses and 3) no capability to provide certainty factors (Hsieh *et al.*, 2019).

In this chapter, a swimming behaviour assay and an oxidative metabolism-dependent energy expenditure assay were integrated into the standard zebrafish FET test to test the following hypotheses: 1) including behavioural and energy expenditure endpoints in the ZFET test produces PODs that are more informative and sensitive compared to only overt toxicity endpoints (i.e., dead and deformed) and 2) this integrated method is protective enough to substitute traditional animal experiments and act as an animal-alternative chemical hazard screening tool. The integrated ZFET test was used to screen 29 chemicals (Table 1) from various classes and mechanisms of action, exposing zebrafish embryos/larvae from 6-120 hours post

fertilization (hpf). PODs for each endpoint were determined using a novel bootstrapped BMD analysis that provides measures of certainty and confidence intervals, facilitating their comparison across all 29 test chemicals.

Overall, PODs derived from perturbed swimming behaviour and oxidative metabolic-dependent energy expenditure phenotypes were earlier indicators of toxicity than PODs derived from death and deformity endpoints. However, after comparing the experimentally derived PODs to published chronic and acute PODs from juvenile and adult fish in the literature, we found that the supplemented ZFET test lacked low enough PODs required to replace traditional toxicity test methods. Integrating additional endpoints, such as transcriptomics, is recommended to provide a more comprehensive assessment of the potential risks of chemical exposure.

2.3. Materials and Methods

2.3.0. Summary of approach

Zebrafish embryos were exposed to 29 chemicals with different mechanisms of action and potencies (Table 1 and Table 2) for 5 days following the internationally standardized guidelines of the OECD FET test (TG 236). The overt toxicity (mortality and deformity), swimming behaviour and metabolism-dependent energy expenditure endpoints were measured. Statistical LOECs and BMDs were calculated for each endpoint. A novel approach was used to empirically derive the BMR threshold level and the BMD (Hsieh *et al.*, 2019). The BMD analysis was further improved by calculating measures of certainty (certainty factors) associated with the BMD. Chemicals associated with a BMD with a certainty factor exceeding 50% were considered “active” (Hsieh *et al.*, 2019). Experimentally derived PODs were compared within and across endpoints and compared to chronic and acute PODs from juvenile and adult fish in

the literature. The relative sensitivity of the endpoints was determined by comparing their PODs. The endpoint with the lowest POD was considered the lowest.

The protectiveness of the supplemented ZFET test was also assessed to determine if the methods could substitute traditional toxicity tests. Experimental PODs are protective of a toxic outcome (e.g., chronic overt toxicity) when the POD is equal to lower than the POD of the toxic outcome (e.g., the lowest chronic toxicity POD from the literature).

2.3.1. Animal husbandry

Wildtype (WT) adult breeding zebrafish sourced from uOttawa stock were maintained in 8L and 3.5L tanks at five fish/1L density in Techniplast zebrafish housing and circulating system water. System water was treated with Instant Ocean sea salt to maintain a pH of 7.3 and a conductivity of 400µS. All zebrafish rooms were maintained at 28°C ± 1°C and had a 14:10 hour day-night cycle (lights on at 8 am EST). Breeding adults were fed twice daily with Skretting GEMMA Micro 300 and supplemented with frozen bloodworms.

Breeding occurred in the 60L Techniplast iSpawn mass embryo production system (MEPS). Approximately 30 to 40 zebrafish were bred over 2h, 30 min after the lights turned on in the morning at 8 am. When transferring to and from the MEPS, animals were carefully handled by slowly submerging the tanks and carefully netting them to reduce stress and increase spawning success. Fertilized eggs were collected into Petri dishes with system water, where they were counted. Coagulated and unfertilized eggs were removed using polyethylene transfer pipettes.

2.3.2. Chemicals

Chemical selection

29 Chemicals were selected with risk assessors from Environment and Climate Change Canada's New Substances Ecological Assessment Division. The rationale for risk assessors selecting the 29 chemicals was to select a diverse set of mechanisms of action and to select chemicals that Environment and Climate Change Canada is interested in assessing. To clarify, I did not participate actively in the chemical selection process. Table 1 contains a list of chemicals with descriptions of mechanisms of action and the rationale for including them in this study. All compounds tested were purchased at the highest percentage purity commercially available.

Table 1: List of 29 chemicals included in the study, their potential mechanism of action (MOA), and the rationale for including them in this research project.

CAS	NAME	MOA	RATIONALE
112-30-1	1-Decanol	Polar Narcosis/weaker surfactant	Related to octanol
111-87-5	1-Octanol	Polar Narcosis/weaker surfactant	In FET validation
112-27-6	(TEG) Triethylene Glycol	Polar Narcosis/weaker surfactant	In FET validation set (Busquet et al 2014). Chronic fish data available
151-21-3	(SDS) Sodium Dodecyl Sulfate	Polar Narcosis/stronger surfactant	In the 21d literature review report (negative)
140-66-9	(4-TOP) 4-t Octylphenol	Weak ER	Used in multigenerational validation/high ACR
80-46-6	(4-TPP) 4-Tert Pentylphenol	Weak ER	Related to 4-t-octylphenol/high ACR
105-67-9	(2,4-DNP) 2,4-Dimethylphenol	NA	Related to 4-t octyl and pentylphenol/high ACR
59-50-7	(4C3MP) 4-Chloro 3-Methylphenol	Perhaps very weak ER or perhaps non-endocrine	Related to 2,4 dimethyl phenol, used in multigenerational validation
67747-09-5	Prochloraz	Aromatase inhibitor/steroidogenesis inhibitor	Inlcuded in multigenerational study, in FET validation
50-28-2	17b-Estradiol	ER binder (estrogenic)	Used in multigenerational validation
67-68-5	(DMSO) Dimethyl Sulfoxide	Control	In validation set (Busquet et al 2014). High ACR. Chronic fish data
95-76-1	(3,4-DCA) 3,4 dichloroaniline	NA (negative control for endocrine disruption)	Positive control in FET, chronic data available
10161-33-8	(Trenbolone) 17B-Trenbolone	AR agonist	Used in multigenerational validation
50471-44-8	Vinclozolin	AR antagonist	Used in multigenerational validation
54910-89-3	Fluoxetine	Selective Serotonin Re-uptake Inhibitor	Chronic toxicity data available
122-14-5	Fenitrothion	Cholinesterase inhibitor (neurotoxic)	Similar to malathion
1071-83-6	Glyphosate	Herbicide, amino acid inhibition	NA
121-75-5	Malathion	Cholinesterase inhibitor (neurotoxic)	In FET validation set (Busquet et al 2014). Chronic fish data available. Related to fenitrothion
102676-47-1	Fadrozole	Aromatase inhibitor	NA
13311-84-7	Flutamide	Anti-androgen	NA
882-09-7	Clofibric Acid	PPAR alpha, and interferes with auxin	Chronic data available
51-28-5	(2,4-DNP) 2,4 Dinitrophenol	Respiratory un-coupler	In FET validation
62-53-3	Aniline	NA	Related to 3,4-dichloroaniline
68-12-2	(NN-DMF) NN-Dimethylformamide	NA	NA
57-63-6	(EE2) Ethinyl estradiol	ER binder (estrogenic)	Estrogenic Control
56-53-1	(DES) Diethylstilbestrol	ER binder (estrogenic)	Estrogenic Control
80-05-7	(BPA) Bisphenol A	ER binder (estrogenic)	Estrogenic BPA – Positive Control
1478-61-1	(BPAF) Bisphenol AF	ER binder (estrogenic)	Suspected estrogenic replacement
41481-66-7	(TGSH) Phenol, 4,4'-sulfonylbis[2-(2-propenyl)-	ER binder (estrogenic)	Suspected NON-estrogenic replacement

Literature review

A literature review was conducted to identify chronic and acute toxicity PODs for each test chemical in zebrafish and other fish species. A spreadsheet containing literature sources is available in the Supplementary Information. The inclusion criteria for the literature review were as follows: 1) toxicity tests must be conducted in adult or juvenile fish, 2) exposure methods must be via waterborne exposure, 3) the concentrations of doses must be maintained above 80% of the initial concentration throughout the exposure duration, 4) for the acute LC_{50}s (lethal concentration for 50% of population lethality), the exposure duration must be four or five days, 5) for the chronic toxicity POD, the duration must be twenty-one days or more, 6) for the acute LC_{50}s, the only endpoint considered must be lethality/mortality and 7) for the chronic toxicity POD, the endpoints must be related to mortality, development, growth, populations (such as sex ratios) or reproduction. PODs from the literature included: 1) the no observed adverse effect level (NOAEL), 2) the no observed effect level (NOEL), 3) the lowest observed effect level (LOEC) or 4) the 10% effective concentration (EC10).

The exclusion criteria for the literature review were to exclude studies if: 1) the toxicity test was conducted in embryos (<30dpf), 2) the toxicity test had a mixture of chemicals for the exposure (e.g., BPAF mixed with BPA), 3) the chemical of interest was less than 90% purity, 4) a conflict of interest is identified (e.g., influence from financial ties to industries that could affect the research findings), 5) an assessment factor has been applied to values and 5) an additional confounding factor was present, such as temperature, affecting the toxicity value.

The following sources were used for the literature review: 1) Health and Environmental Sciences Institute (HESI) 2021-2023 EnviroTox Database version 2.0.0, 2) 2021-2023 US EPA

ECOTOX Knowledgebase, 3) OECD Test Guideline validation reports, 4) and independent scientific literature found on Google Scholar, Web of Science and Scopus that was not captured by the previously mentioned databases and reports (See Literature_Review_Sources.csv in Supplementary Information).

Literature-based dose selection

Table 2 describes reported POD values from the literature and the highest and lowest doses used for each test chemical in this study. Dose ranges for the exposures were selected based on PODs from the literature and followed a $\log_{10}$ scale. The following criteria were used to determine the appropriate experimental dose range for each chemical:

I. If both an acute LC_{50} and chronic toxicity POD were available in the literature, the minimum of the dose range was determined by rounding the chronic toxicity POD value down to the nearest tenth and dividing by ten (e.g., if the chronic toxicity POD is 12mg/L round it to 10mg/L then divide it by 10. The lowest dose would be 1mg/L). The maximum of the dose range is equal to the lowest dose × 10^4 (e.g., the highest dose would be 1mg/L × 10^4 = 10,000mg/L). If the dose range reached or exceeded the acute LC_{50}, the entire dose range was divided by 10 (one order of magnitude). For example, if the chronic toxicity POD value was 12mg/L and the acute LC_{50} value was 5,650mg/L, the experimental dose range for the chemical exposure was adjusted to 0.1mg/L (lowest dose) - 1,000mg/L (highest dose).
II. If there was insufficient evidence available in the literature at the time of the review, a best-attempt at an effective range was chosen; chronic toxicity POD values fell in the range of the experimental dose range, and the acute LC_{50}s were avoided when possible.

Lethal concentrations were intentionally avoided to quantify only the sub-lethal effects of chemical exposure, as outlined above. If exposure experiments reached the lethal concentration, a lower dose group was added to a follow-up experiment.

Table 2: List of 29 chemicals included in the study with toxicity values.

CAS	NAME	CHRONIC TOX POD ZEBRAFISH	CHRONIC TOX POD FISH	ACUTE LC_{50} ZEBRAFISH	ACUTE LC_{50} FISH	HIGH DOSE	LOW DOSE
112-30-1	1-Decanol	0.43	NA	3.21	0.602	1	0.0001
111-87-5	1-Octanol	NA	0.75	19.8	12.2	10	0.001
112-27-6	(TEG) Triethylene Glycol	NA	NA	54800	5000	1000	0.1
151-21-3	(SDS) Sodium Dodecyl Sulfate	NA	1.357	3.6	1.39	0.1	1E-05
140-66-9	(4-TOP) 4-t Octylphenol	0.0046	0.012	0.462	NA	0.1	1E-05
80-46-6	(4-TPP) 4-Tert Pentylphenol	NA	0.057	NA	1.6	1	0.0001
105-67-9	(2,4-DNP) 2,4-Dimethylphenol	NA	0.131	NA	6.3	1	0.0001
59-50-7	(4C3MP) 4-Chloro 3-Methylphenol	NA	NA	NA	1.3	1	0.0001
67747-09-5	Prochloraz	0.03	0.063	5.6	NA	1	0.0001
50-28-2	17b-Estradiol	4.8E-06	4.2E-07	NA	1.28	0.1	1E-05
67-68-5	(DMSO) Dimethyl Sulfoxide	NA	0.006	20,964	34,000	10	0.001
95-76-1	(3,4-DCA) 3,4 dichloroaniline	0.002	0.0011	8.5	0.95	(1)0.1	(0.0001)0.00001
10161-33-8	(Trenbolone) 17B-Trenbolone	9.19E-06	1.5E-06	NA	NA	0.1	1E-05
50471-44-8	Vinclozolin	NA	0.05	4.86	13.6	1	0.0001
54910-89-3	Fluoxetine	0.0032	0.00054	0.25	NA	0.1	1E-05
122-14-5	Fenitrothion	NA	0.046	NA	0.2	1	0.0001
1071-83-6	Glyphosate	13.9	0.1	53.75	36.3	1	0.0001
121-75-5	Malathion	1	0.021	0.105	0.1	1	0.0001
102676-47-1	Fadrozole	0.1	0.0955	NA	NA	1	0.0001
13311-84-7	Flutamide	NA	0.101	NA	1.92	1	0.0001
882-09-7	Clofibric Acid	70	10.75	NA	526.5	10	0.001
51-28-5	(2,4-DNP) 2,4 Dinitrophenol	NA	0.208	NA	0.62	(10)1	(0.001)0.0001
62-53-3	Aniline	5.60	0.39	32	21.99	10	0.001
68-12-2	(NN-DMF) NN-Dimethylformamide	NA	0.1	NA	6,300	1000	0.1
57-63-6	(EE2) Ethinyl estradiol	5.0E-08	4.0E-08	1.7	0.001	0.001	1E-07
56-53-1	(DES) Diethylstilbestrol	1.0E-05	NA	NA	NA	0.1	1E-05
80-05-7	(BPA) Bisphenol A	0.2	0.05	8.04	4.6	1	0.0001
1478-61-1	(BPAF) Bisphenol AF	0.05	0.07	1.95	NA	1	0.0001
41481-66-7	(TGSH) Phenol, 4,4'-sulfonylbis[2-(2-propenyl)-	NA	NA	NA	NA	1	0.0001

The high and low experimental doses are included. Concentrations are in mg/L. Values in brackets in the high/low dose columns represent doses from experiments that reached lethal levels and were redone with lower doses. Chronic toxicological values include NOECs, NOAELs, EC10s and LOECs. The minimum chronic PODs and acute LC50s from the literature review are summarised in this table. For a full list of all toxicity values from the literature, please see the Supplementary Information

Stock solutions and solvents

Stock solutions were freshly prepared at 10X the highest dose concentration in glass bottles using system water (pH 7.3 conductivity of 400 µS) with 0.01% v/v Dimethyl Sulfoxide (DMSO). Some chemicals were hydrophobic and needed to be dissolved in DMSO after exhausting mechanical options of dissolving the test solution in the system water (e.g., 4-tert-Octylphenol/4-TOP). To dissolve chemicals in DMSO, they were added to 99% DMSO and serially diluted 1:10 in system water to obtain a final target concentration of 10X the highest dose and a final DMSO concentration of 0.01% v/v.

The concentration of 0.01% v/v DMSO (0.1mg/L) was chosen because, at that concentration, DMSO does not affect the permeability of the chorion and solubility and stability of test solutions that can lead to unreliable results (Kais *et al.*, 2013). Each chemical exposure experiment had an internal solvent control dose group of 0.01% v/v DMSO in system water to address the issue of using DMSO as a solvent to solubilize contaminants in water. DMSO was also included as a test chemical (with a system water-only control group) to elucidate any potential effects of the solvent control groups within each experiment.

2.3.3. Experimental Design

ZFET test & chemical exposure

A graphical description of the experimental exposure layout is available in Supplementary Figure 1. All embryo exposure experiments were conducted in a climate-controlled room (28°C ± 1°C) with a 14:10h light/dark cycle in covered glass Petri dishes.

At 1.5hpf, 20 embryos were individually pipetted with 720µL of the system water into each Petri dish for a total volume of 14.4mL per dish. Three replicate Petri dishes were prepared per dose group. Then, 1.6mL of the temperature-acclimated 10X target dose and 1.44µL of 99% DMSO were added to each dish to maintain a final concentration of 0.01% v/v DMSO. The final volume of exposure media was 16mL for each Petri dish.

On subsequent days, target concentrations were prepared in 0.01% DMSO system water and temperature acclimated in the exposure room. Semi-static media renewal was done daily by removing half of the previous day's exposure media and replacing it with freshly prepared exposure media to maintain the nominal test concentration throughout the experiment.

Lethal and sub-lethal developmental toxicity endpoints were scored daily, and dead embryos were removed. Details of the developmental toxicity endpoints scored at each timepoint throughout the experiment can be found in Supplementary Table 1 in the Appendix. Lethal developmental toxicity endpoints were described as gastrulation arrest (day 0), coagulation of the embryo, a lack of somite formation, lack of tail bud detachment from the yolk, a lack of heartbeat and finally, failure to hatch by 120hpf (OECD, 2013). Sub-lethal developmental toxicity endpoints (deformities) were described as spinal malformations, edema/swelling, a malformed heart, yolk deformation or reduced absorption of the yolk, eye malformations, and a lack of pigmentation (von Hellfeld *et al.*, 2020). Media renewal, endpoint scoring, and dead embryo removal continued until 120hpf, when the exposure concluded. I extended the exposure duration and development time to 120hpf instead of 96hpf, which is recommended in the OECD guideline. This is because, at 120hpf, zebrafish (eleuthero)embryos/larvae have full biotransformation capabilities, increased general activity, fully inflated swim bladders and fully developed primary and secondary islets of the pancreas (Kimmel *et al*., 2013; Reid *et al*., 2018).

The zebrafish embryo is metabolically and cognitively competent at 120hpf (Kimmel *et al*., 2013; Reid *et al*., 2018; Renquist *et al*., 2013; Kozol *et al*., 2016; Tal *et al*., 2023; Klüver *et al*., 2015). At 120hpf, the zebrafish embryo is better suited for the behavioural and metabolism assay.

Developmental toxicity endpoints from the ZFET test were summarized as the percentage of affected embryos. It is also referred to as the affected rate endpoint in the text and figures. The percentage affected response represented the total percentage of embryos that died or were deformed by chemical exposure for each dose group and replicated Petri dish. The percentage affected response was normalized for each chemical exposure by subtracting the mean control group baseline percentage from all percentage affected values because the baseline response was not zero across all experiments. Normalizing the percentage affected response values allows for comparisons across datasets and experiments. Normalizing and shifting the baseline response to zero also facilitated the benchmark dose analysis.

Swimming behaviour assay

Upon completion of the 5-day exposure assay, at 120hpf, three non-deformed zebrafish larvae from each replicate Petri dish were randomly selected and transferred to a clear 96-well plate (one larva per well, nine larvae per dose group, 6 doses, including controls, per chemical = 54 larvae per plate) with 200µL of exposure media. An example plate layout can be found in Supplementary Figure 2. The 96-well plate was placed in a ViewPoint Zebrabox to monitor the locomotor response of exposed larvae to cycling darkness and light conditions. The exact ZebraLab protocol (.vte file) for the swimming behaviour assay can be found on GitHub (https://github.com/JoryC/PFOS_OBS_F53B_Project_Zebrafish_Light_Dark_Behavioural_Response/blob/master/ZebraLab_Protocol.vte). Briefly, the light-dark swimming behaviour assay is a

30-minute assay that assesses the subtle effects of perturbed muscle function and the neurotoxic effects of chemical exposure. A visual description of the light-dark protocol can be found in Supplementary Figure 3. Before the assay, the fish were acclimated to the ViewPoint Zebrabox with the lights on for 20 minutes. Larvae were video recorded and tracked via infrared camera for 30 minutes. The 30-minute tracking period consisted of alternating 5-minute dark and light photoperiods. For each bin (60 seconds per bin), the total distance travelled in mm was measured using the ViewPoint Zebrabox tracking software, as well as numerous other variables such as the duration of inactivity/activity (s) and short and large swim movement distances travelled (mm). Other variables, such as velocity (mm/s), were calculated. However, the total distance moved response endpoint was selected to represent overall locomotor activity. Poor quality observations (less than 20s of movement trace per 60s bin) from errors in the ViewPoint Zebrabox were removed from the data set, and repeated one-sided Grubb's tests identified any other outliers.

Oxidative metabolism-dependent energy expenditure alamarBlue® assay

To measure the metabolic activity of zebrafish larvae, the alamarBlue® metabolism assay was used (Williams and Renquist, 2016; Renquist *et al.*, 2013). The alamarBlue® assay is dependent on an NADH-, NADPH-, $FADH_2$, and $FMNH_2$-based reduction of alamarBlue® (resazurin) and has been described as a proxy to quantify zebrafish oxidative metabolism (Chackal *et al.*, 2022; Williams and Renquist, 2016; Renquist *et al.*, 2013; Tu *et al.*, 2019). The 96-well plate layout for the alamarBlue® metabolism assay can be seen in Supplementary Figure 4. Upon completion of the 5-day exposure assay, at 120hpf, three non-deformed zebrafish larvae from each Petri dish were transferred to a black-walled 96-well plate with 100µL of exposure media and 100µL of 2X alamarBlue® media (1 larva per well, 9 larvae per dose group, 6 doses,

including controls, per chemical = 54 larvae per plate). Fresh 30mL aliquots of 2X alamarBlue® media (0.6mL Thermo Fisher Scientific alamarBlue®, 6.0mL of 4mM $NaHCO_3$, 0.06mL DMSO and 23.34mL of system water) were made for each experiment. Four wells of the black-walled 96-well plate were used as blank calibration controls and contained 100µL of 2X alamarBlue® media and 100µL of 0.01% DMSO in system water. Baseline fluorescence (in relative fluorescence units - RFUs) was measured at 120hpf using a Molecular Devices Gemini EM microplate reader with the parameters: λ excitation of 530nm and λ emission of 590nm. After baseline measurement, larvae were stored in covered 96-well plates in alamarBlue® media inside a dark incubator at 28°C ± 1°C. After 24h in alamarBlue® media (at 144hpf), the final fluorescence was measured using the same parameters. Measurements at each time point were recorded in duplicates, and each well's average was used as the final RFU. The change in fluorescence over 24 hours (**Δ**fluorescence) was calculated using the following method:

(Williams and Renquist, 2016)

$$\boldsymbol{\Delta} fluorescence = (\underline{24h\ RFU} - \underline{24h\ Blank\ RFU}) - (\underline{Baseline\ RFU} - \underline{Baseline\ Blank\ RFU}) \quad (1)$$

2.3.4. Traditional statistical POD approach: lowest observed effect concentration (LOEC)

Overt toxicity

Overtly toxic phenotypes (i.e., dead and deformed embryos) were recorded during the ZFET test. Overtly toxic phenotypes were summarised as the percentage of affected embryos. A one-way ANOVA was conducted for each exposure experiment to determine if the percentage

affected response differed across dose groups. Normality for each experiment was tested using quantile-quantile plots. Outliers were removed using repeated one-sided Grubbs' tests. With outliers removed from each experiment, the assumption of homoscedasticity was tested using Levene's test ($\alpha = 0.05$). P-values were adjusted using the false discovery rate (FDR) adjustment. A post hoc Tukey's honestly significant difference test (Tukey's HSD) was conducted to elucidate any potential subgrouping effects resulting from having dose replicates in different Petri dishes. Finally, a post hoc Dunnett's test was conducted to compare the means of the solvent control group to the mean of each treatment for each experiment. The overt toxicity LOEC was determined as the lowest dose at which a significant effect was observed.

Swimming behaviour

Immediately after the 5-day ZFET test, larval swimming behaviour was recorded over 30 minutes with alternating 5-minute light:dark phases. A one-way ANOVA was conducted for each exposure experiment, and phase to determine if the total distance moved (in mm) response differed across dose groups. Nine Individuals were assessed for each concentration for each chemical. The light and dark phases were pooled together (dark time=15 minutes total, light time=15 minutes total), and a one-way ANOVA was conducted for each pooled phase and chemical exposure experiment. The assumption of normality and homoscedasticity were tested for each experiment with quantile-quantile plots and Levene's test ($\alpha = 0.05$). Repeated one-sided Grubbs' tests were conducted to identify and remove outliers.

The swimming behaviour LOEC was calculated for each chemical and light/dark phase using a one-way ANOVA and post hoc Dunnett's test, comparing each treatment group to its corresponding internal control group. Because many statistical comparisons ($n = 290$) were performed with many observations per comparison ($n = 135$), a strict p-value cutoff of 0.001

after FDR adjustment was used to avoid Type I errors. The swimming behaviour LOEC for each chemical with significant results was the lowest dose at which a significant effect was observed grouped by either light or dark phase.

Oxidative metabolism-dependent energy expenditure

Immediately after the five-day ZFET test, larvae were prepared for the alamarBlue® metabolism assay, which measures larvae's oxidative metabolism-dependent energy expenditure over 24h. A one-way ANOVA was conducted for each exposure experiment to determine if the 24h-change in the fluorescence (**Δ**fluorescence) response was different for dose groups with different exposure concentrations. Nine Individuals were assessed for all concentrations of each chemical. The assumption of normality and homoscedasticity were tested for each experiment with quantile-quantile plots and Levene's test ($\alpha = 0.05$). Repeated one-sided Grubbs' tests were performed to identify and remove outliers.

The LOEC was calculated for each chemical using a one-way ANOVA and Dunnett's post hoc test. The LOEC for each chemical was the lowest dose at which a significant effect was observed.

2.3.5. Modelling approach: Benchmark dose (BMD) Analysis

Different models and data transformations were used to maximize the likelihood that an endpoint's dose-response data would fit a model(s) well enough to be considered further for analysis. The different data transformations will first be discussed, followed by the different models.

Pre-BMD modelling data transformation

Light-dark swimming behaviour

Supplementary Figure 5 describes transforming the total distance moved endpoint to similarity scores. After recording the swimming behaviour of zebrafish larvae over 30 minutes in the ViewPoint Zebrabox, the total distance moved (in mm, not including the 20-minute acclimation period) variable was transformed into similarity scores (Pearson's r correlation coefficient and Spearman's rho correlation coefficient) to prepare the data for BMD modelling. Previous studies showed that both similarity types metrics performed similarly in a BMD analysis (Hsieh *et al.*, 2019). For each fish, a similarity score was determined based on the correlation between its total distance swimming pattern and the swimming patterns of each of the 9 control fish. Thus, each treated fish had 9 similarity scores (one for each control fish), and the median similarity score was considered the representative similarity score response value. Each of the similarity scores was then normalized:

(Hsieh *et al*., 2019)

$$Normalized\ similarity\ score\ = \left(\frac{V_{treated}}{V_{solvent\ control}}\right) \times 100\ -100 \quad (2)$$

where $V_{treated}$ is the similarity score of the chemically treated larvae, and $V_{solvent\ control}$ is the median value of the similarity scores from all the larvae in the vehicle control.

After normalization, the baseline response of zero represents the median of the solvent control group. The normalized similarity scores of larvae below 0 and approaching -1 represent larvae with decreasing movement similarity compared to the control group's median. Similarity scores were multiplied by 100 to calculate the percent similarity to the median of the control group.

Oxidative metabolism-dependent energy expenditure

The oxidative metabolism-dependent energy expenditure of zebrafish over 24h was measured using the alamarBlue® assay and was summarised as the change in fluorescence over 24 hours (**Δ**fluorescence) response. The **Δ**fluorescence values were further transformed to z-scores relative to the mean and standard deviation of the control groups for each chemical:

$$Z\text{-}score = \frac{\mathbf{\Delta}fluorescence_{treated} - \underline{\mathbf{\Delta}fluorescence_{solvent\ control}}}{\sigma\ \mathbf{\Delta}fluorescence_{solvent\ control}} \quad (3)$$

where $\mathbf{\Delta}fluorescence_{treated}$ is the **Δ**fluorescence of the treated larva $\underline{\mathbf{\Delta}fluorescence_{solvent\ control}}$ is the mean **Δ**fluorescence of the control larvae and $\sigma\ \mathbf{\Delta}fluorescence_{solvent\ control}$ is the standard deviation of the **Δ**fluorescence of control larvae.

The z-score transformation was done before BMD modelling to normalize each experiment's response to the control group and shift the baseline response to zero. Normalizing the response values to the control group allows for comparisons across experiments in the dataset.

Benchmark Dose Modelling

This section outlines the bootstrapped non-parametric 'Curvep' and parametric Hill equation model approach to BMD modelling. 'Curvep' is a non-parametric noise filtering algorithm originally developed for processing cytotoxicity data that produces dose-response curves (Sedykh, 2016). It has major advantages over parametric models because of its robustness. The method is less sensitive to outliers and can handle different data types, such as ordinal, nominal or skewed. The method's robustness is important for noisy data with outliers

and slightly skewed data such as dose-response behavioural data. The bootstrap approach involves the repeated sampling of dose-response curves with replacement, which allows for a BMD to be calculated with associated confidence intervals and a certainty factor. One of the novel aspects of this approach is the ability to approximate the biologically significant BMR threshold level using data-driven heuristics.

In dose-response BMD analyses, selecting the BMR threshold level is a contentious process step. One of the default BMR threshold levels recommended by the US EPA is the 10% adverse health effect level (Hogan *et al.*, 2012). However, it is not likely that the 10% level is biologically relevant for all health data sets. The 'Curvep' approach to BMR selection approximates a data set's BMR threshold (Hsieh *et al.*, 2019). The BMR is a critical component of BMD modelling. The BMR is representative of a biologically significant level of response that is used as a threshold for calculating the BMD. The BMD is an estimate of the concentration associated with a biologically significant level of response based on the BMR. Common approaches to determining a BMR threshold use a fixed percentage change from the control group response or a standard deviation-based method that depends on the variability in the control group response. However, these fixed-level approaches are not always appropriate for all datasets or endpoints. For example, a fixed percentage change may not be biologically significant if the control group response is low, and a fixed standard deviation-based BMR is not always appropriate when background noise is non-normal. The heuristic 'Curvep' BMR calculation approach attempts to find the minimum threshold level where the variance of the potency measurements is stabilized for all of the chemicals in a dataset. It provides a flexible, data-driven approach to selecting a biologically significant BMR threshold. The heuristic

'Curvep' BMR calculation approach reduces the subjectivity of selecting a BMR threshold and improves the reproducibility of the results.

Curvep approach

Benchmark response threshold

This subsection will describe the methods for approximating the most appropriate BMR threshold of overt, behavioural and metabolic toxicity data sets. A bootstrap was conducted to estimate the BMR for the chemical exposure dose-response data set. One thousand responses at each dose were simulated by using sampling with replacement (Supplementary Figure 6). Each sampled response point from each dose was grouped by iteration of the bootstrap. The results were 1000 simulated sample curves with one response point at each dose. The non-parametric 'Curvep' response-noise filtering algorithm was used to process the sample curves (Sedykh, 2016). The algorithm functions by attempting to find the minimum set of corrections that would give a monotonic dose-response relationship. It makes corrections by applying the following rules to each point in the sample curve:

I. if the response point is below the noise threshold level, it is set to zero
II. if the response point is above the noise threshold level but violates the monotonicity assumption, the average of its adjacent points replaces it and
III. if the point is above the noise threshold level and does not violate the monotonicity assumption, it is left unchanged

'Curvep' is sensitive to the noise threshold level parameter and the maximum allowed deviation from the monotonicity parameter.

Preliminary BMDs (preBMDs) were calculated for each sample curve. The preBMDs were defined as the dose at which a sample curve passed through the noise threshold level (if a curve did not pass the noise threshold level, the preBMD was set to the maximum dose). Curve processing and preBMD calculations were repeated iteratively for incrementally increasing noise threshold levels (e.g., if responses ranged from 0% to 100%, noise threshold levels ranging from 5% to 95%, in increments of 5%, were tested). The maximum allowed deviation from the monotonicity parameter was set to 5% for all experiments.

Each chemical in the dataset was included in the bootstrap. Hundreds of thousands of preBMDs grouped by noise threshold level were calculated. The pooled variance of the preBMDs was calculated for every noise threshold level tested, plotted and fitted to an exponential-like curve (Supplementary Figure 7). The knee-point of the fitted exponential-like curve, representing the point of the curve with maximum variance stability, was considered the most appropriate BMR threshold value for the data set. Next, the final BMD was calculated using the empirically derived BMR threshold.

Benchmark dose

This subsection describes the methods used to calculate the final BMD of the overt, behavioural and metabolic toxicity data sets. The 'Curvep' bootstrap was repeated with one threshold BMR parameter set to the approximate BMR of the dataset (the knee-point) to calculate the BMD, BMD_L and BMD_U for each chemical in the dataset. The final BMD was defined as the median of the bootstrapped preBMDs; the BMD_L was defined as the lower bound of the 95% confidence interval of the BMD, and the BMD_U was defined as the upper bound of the 95% confidence interval of the BMD. Certainty factors were also reported for each chemical corresponding to the number of sampled curves crossing the BMR threshold level. Certainty

factors were treated as a biological activity measure resulting from chemical exposure. BMDs with a certainty factor above 50% were classified as "active", while BMDs with certainty factors below 50% were considered "inactive" (Hsieh *et al.*, 2019).

Classical Hill equation model approach

Benchmark dose

To facilitate a more balanced representation of BMDs from different methods, a classical parametric Hill equation model fit approach was also used in a bootstrap to calculate the BMD (and the associated certainty factor), BMD_L and BMD_U for each chemical with the most appropriate BMR threshold calculated with the 'Curvep' approach.

Similar methods to calculating the BMD using the bootstrapped 'Curvep' approach were used for the classical Hill equation model approach, with the only major difference being that the sample curves were fitted to the Hill equation model before determining BMD values.

2.4. Results

Throughout this section, chemical abbreviations are used. Please refer to Table 1 or Table 2 to see the full names of each chemical along with its associated Chemical Abstracts Service (CAS) registry number.

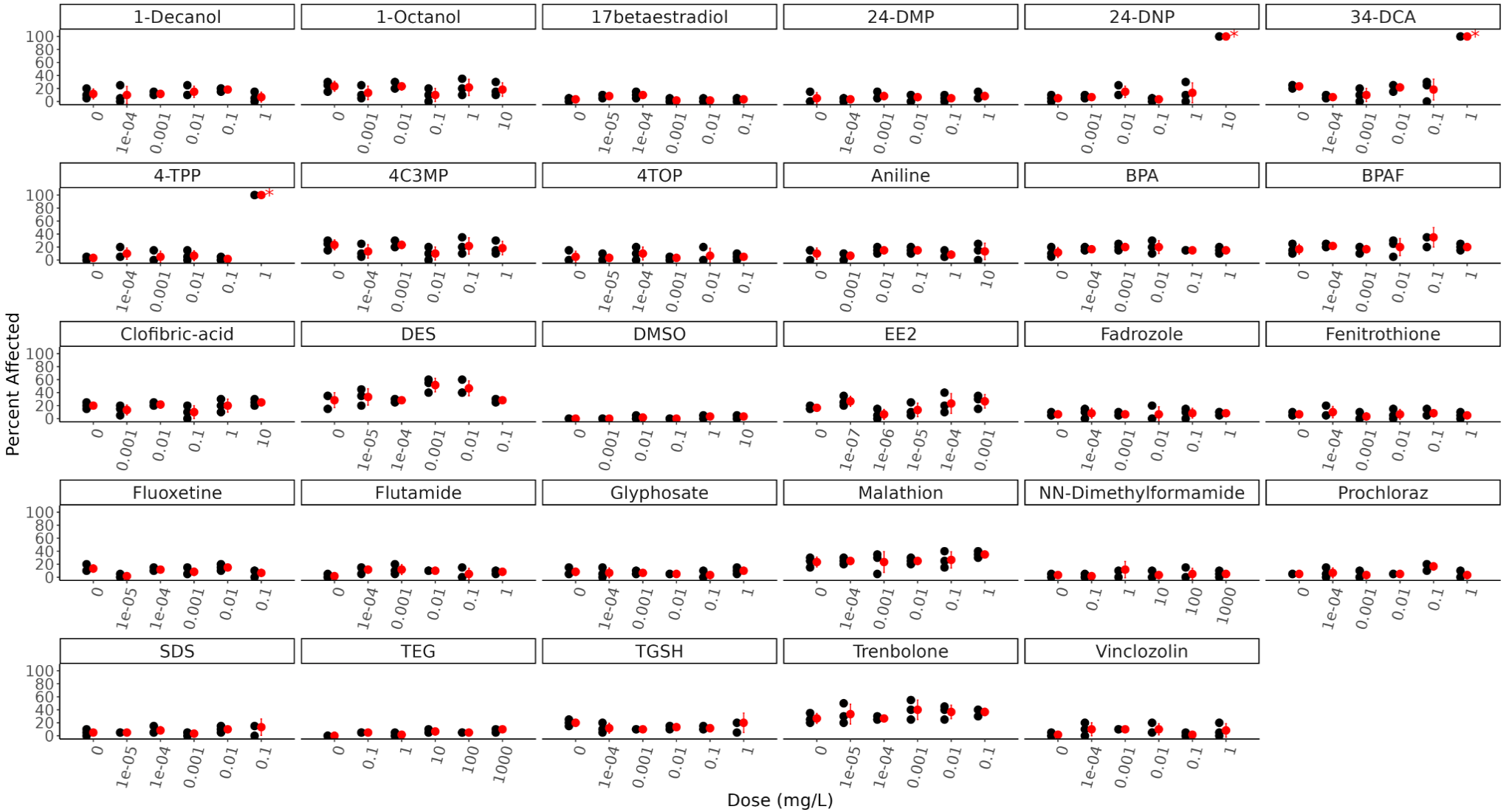
2.4.1. Statistical LOEC Analysis

The statistical LOECs for each chemical and endpoint were calculated to facilitate comparisons within and across experiments and between experimental LOECs and PODs from

the literature. The statistical LOEC analysis section is separated by endpoint (i.e., overt toxicity, oxidative metabolism-dependent energy expenditure and swimming behaviour).

Overt toxicity

Overtly toxic (i.e., dead and deformed) phenotypes were recorded during the ZFET test. Overtly toxic phenotypes were summarised as the percentage of affected embryos. After testing the model's assumptions and removing outliers, a one-way ANOVA was conducted for each exposure experiment to determine if the percentage affected response differed across dose groups. Five outliers were removed using repeated one-sided Grubbs' tests (BPAF - Dose 2-Group B, NN-Dimethylformamide - Dose 4-Group B, SDS - Dose 1-Group B, TEG - Dose 1-Group C, TGSH - Dose 1-Group B). All experiments had insignificant Levene's test results ($\alpha = 0.05$), indicating that variance was equal across all dose groups. Potential subgrouping effects were investigated before conducting the one-way ANOVA (Supplementary Figure 8). Tukey's HSD test revealed no statistically significant ($\alpha = 0.05$) subgrouping effects after an FDR p-value adjustment. The results of the one-way ANOVA revealed that the percentage affected response was significantly perturbed by chemical exposure after an FDR p-value adjustment for three exposure experiments: 2,4-DNP - [$p = 5.8 \times 10^{-6}$], 4-TPP - [$p = 7.6 \times 10^{-7}$] and 3,4-DCA - [$p = 1.9 \times 10^{-5}$] (Figure 1; Table 3). Finally, Dunnett's post hoc test results showed that the solvent control group was significantly different from the highest dose group for 3/29 chemical exposure experiments after an FDR adjustment (2,4-DNP - [$p = 3.17 \times 10^{-8}$], 3,4-DCA - [$p = 2.44 \times 10^{-4}$] and 4-TPP - [$p = 4.19 \times 10^{-13}$]).



Black dots represent the percent affected embryos from each Petri dish (n=3 Petri dish per dose). Red points represent the mean percent affected for each dose group. The red error bars are the standard deviation for each dose group. Red asterix represent statistical significance of the post-hoc Dunnett's test and false discovery rate adjustment (* - p < 0.001).

(title on next page)

Figure 1: The overt toxicity endpoint was summarized as the percentage of affected embryos over the 5-day duration of the chemical exposure experiments (i.e., dead and deformed embryos). An ANOVA and Dunnett's post hoc test revealed statistically significant ($p < 0.001$) effects from exposure to 2,4-DNP, 3,4-DCA and 4-TPP after an FDR p-value adjustment. 2,4-DNP and 3,4-DCA exposure affected the survival rate of zebrafish embryos/larvae in the highest dose group, and 4-TPP exposure affected only the rate of deformity (i.e., hypopigmentation) in the highest dose group.

Swimming behaviour

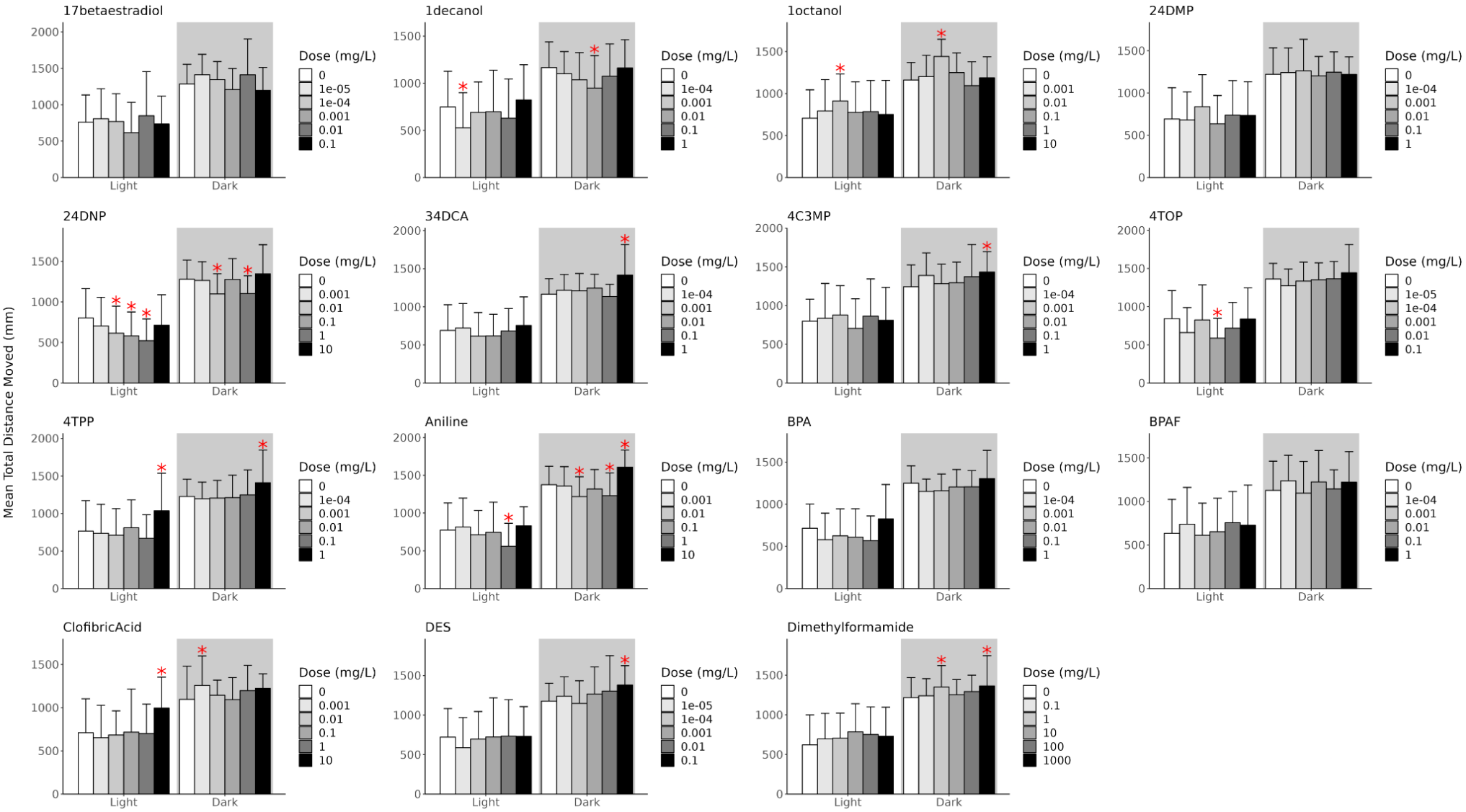
Immediately after the 5-day ZFET test, larvae swimming behaviour was recorded over 30 minutes with alternating 5-minute light and dark phases. After testing the model's assumptions and removing outliers, a one-way ANOVA was conducted for each exposure experiment and light phase to determine if the total distance moved (in mm) response differed across dose groups. In total, 73 outliers were removed using repeated one-sided Grubbs' tests, and 814 low-quality observations (wells detected as 'empty' > 20 seconds) were removed. Outliers and low-quality observations represented just 1.1% of all observations (78,300 total). All chemical exposure experiments showed approximately normal distributions. All but five exposure experiments (1-Octanol - $p = 0.87$, 2,4-DMP - $p = 0.71$, BPA - $p = 0.14$, SDS - $p = 0.09$, TEG - $p = 0.1$) had significant Levene's test results ($\alpha = 0.05$) indicating that variance was heteroscedastic across dose groups for 24/29 experiments. However, heteroscedasticity was considered minor because the sample sizes were equal for each dose group for every experiment, and the maximum group variance did not exceed four times the minimum group variance (Zimmerman, 2004). Heteroscedasticity was thus ignored and not considered a violation of the assumption of homoscedasticity for subsequent analyses (Zimmerman, 2004).

Before further tests, the potential subgrouping effect was investigated to ensure the validity of the one-way ANOVA model. Statistically significant subgrouping effects were observed in the total distance response data. However, only 39/522 (7.5%) of Tukey's HSD comparisons exceeded a 25% difference in the estimated value and estimated marginal mean response. Therefore, the magnitude of the subgroup effect from pseudo-replicates in the dose groups was minor for most (92.5%) samples. Thus, one-way ANOVA was a sufficient model to estimate the dose-response relationship for this data set.

With the validity of the one-way ANOVA model confirmed, one-way ANOVAs were conducted for each exposure experiment and light phase. The results of the ANOVAs revealed significant effects of chemical exposure on the total distance moved response for all chemicals. However, 2,4-DMP ($p = 0.56$) exposure resulted in no significant effects in the dark phase and TEG ($p = 0.64$) and TGSH ($p = 0.63$) exposure resulted in no significant effects in the light phase. The effects of chemical exposure on the total distance moved in mm behavioural toxicity endpoint are summarized in Figure 2. Please refer to Supplementary Figure 24 for the raw behavioural data.

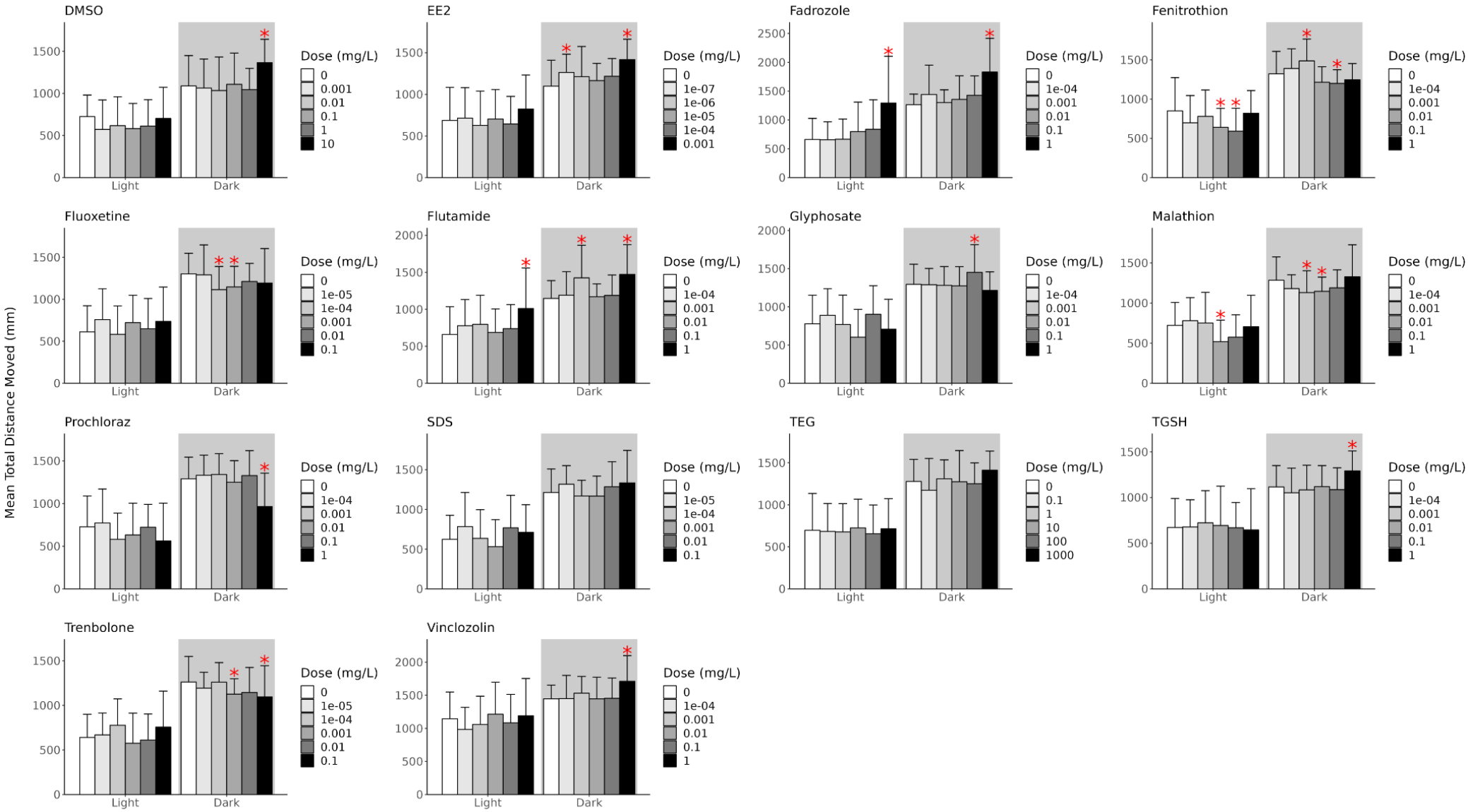
The swimming behaviour LOEC for each chemical exposure experiment can be found in Table 3. A Dunnett's post hoc test was conducted to determine statistical differences in the mean total distance response in the treatment dose groups (Figure 2). After an FDR adjustment and a strict p-value cutoff of 0.001, Dunnett's post hoc test resulted in 44/290 statistically significant ($\alpha = 0.05$) differences between the control and treatment groups. There were 14 significant differences in the light condition and 31 significant differences in the dark condition. All chemical exposures caused an effect in the total distance moved response in either light or dark conditions. Eighteen significant differences were in the Dose 1/Highest dose group, seven were in the Dose 2 group, seven were in the Dose 3 group, nine were in the Dose 4 group, and three were in the Dose 5/Lowest dose group. Hyperactive effects of chemical exposure were observed from exposure to 14 chemicals (1-octanol, 3,4-DCA, 4C3MP, 4-TPP, Clofibric Acid, DES, DMSO, Dimethylformamide, EE2, Fadrozole, Flutamide, Glyphosate, TGSH and Vinclozolin). Hypoactive effects of chemical exposure were observed from exposure to 7 chemicals (1-decanol, 2,4-DNP, 4-TOP, Fluoxetine, Malathion, Trenbolone and Prochloraz). 2 chemical exposures showed inconclusive activity effects because both hyperactive and

hypoactive effects were observed (Aniline and Fenitrothion). No significant effects of chemical exposure were observed in 6 chemicals after Dunnett's test (17b-estradiol, 2,4-DMP, BPA, BPAF, SDS and TEG).



(title on next page)

Figure 2: Swimming behaviour data measured by the total distance moved in mm response was recorded using the Viewpoint Zebrabox, summarized as the mean total distance moved in mm and categorized into light and dark phases. Error bars represent standard deviation. An ANOVA and post hoc Dunnett's test of the mean total distance moved endpoint from the swimming behaviour assay revealed statistically significant ($p < 0.001$) effects from chemical exposure on zebrafish larvae (120hpf) after an FDR p-value adjustment. Red asterisks depict statistically significant ($p < 0.001$) effects. The lowest concentration with statistically significant effects in either the light or dark phase was considered the LOEC.



(title on next page)

Figure 2 (cont'd): Swimming behaviour data measured by the total distance moved in mm response was recorded using the Viewpoint Zebrabox, summarized as the mean total distance moved in mm and categorized into light and dark phases. Error bars represent standard deviation. An ANOVA and post hoc Dunnett's test of the mean total distance moved endpoint from the swimming behaviour assay revealed statistically significant ($p < 0.001$) effects from chemical exposure on zebrafish larvae (120hpf) after an FDR p-value adjustment. Red asterisks depict statistically significant ($p < 0.001$) effects. The lowest concentration with statistically significant effects in either the light or dark phase was considered the LOEC.

Oxidative metabolism-dependent energy expenditure

Immediately after the 5 day ZFET test, the alamarBlue® assay was conducted to measure larvae's oxidative metabolism-dependent energy expenditure for 24h. After testing the model's assumptions and removing outliers, a one-way ANOVA was conducted for each exposure experiment to determine if the 24h changes in the fluorescence (**Δ**fluorescence) response differed across dose groups. Dunnett's post hoc test was conducted for each chemical to determine the effects of the treatment dose groups compared to the control group. In total, 44 outliers were removed using repeated one-sided Grubbs' tests. Outliers represented approximately 2.8% of all observations (1566 total). However, Grubb's test removed all response values from Aniline Dose 1 subgroup C. All experiments showed an approximately normal distribution. All but one experiment (EE2 - $p = 0.02$) had insignificant Levene's test results ($\alpha = 0.05$), indicating that variance was homoscedastic across dose groups for 28/29 experiments. However, because the sample sizes were equal across each dose group for every experiment, any heteroscedasticity was considered minor and was ignored (Zimmerman, 2004). Before conducting the one-way ANOVAs, the potential subgrouping effect was investigated.

One-way ANOVAs were conducted for each subgroup in each dose group and chemical to investigate the subgrouping effect. Tukey's HSD post hoc test was conducted to elucidate which subgroups differed from each other. Because subgroup C from the highest dose of the Aniline exposure experiment was removed, 521 comparisons were conducted instead of 522. Statistically significant subgrouping effects were observed in 10 chemicals in the **Δ**fluorescence response values after an FDR adjustment in 30/521 comparisons from Tukey's HSD post hoc test using a one-way ANOVA (Aniline Dose 1, BPA Dose 2, DES Doses 1-4, EE2 Dose 5, Fadrozole Dose 4, Fenitrothion Doses 1, 2 & 4, Glyphosate Dose 3 & 5, Malathion Dose 5, NN-

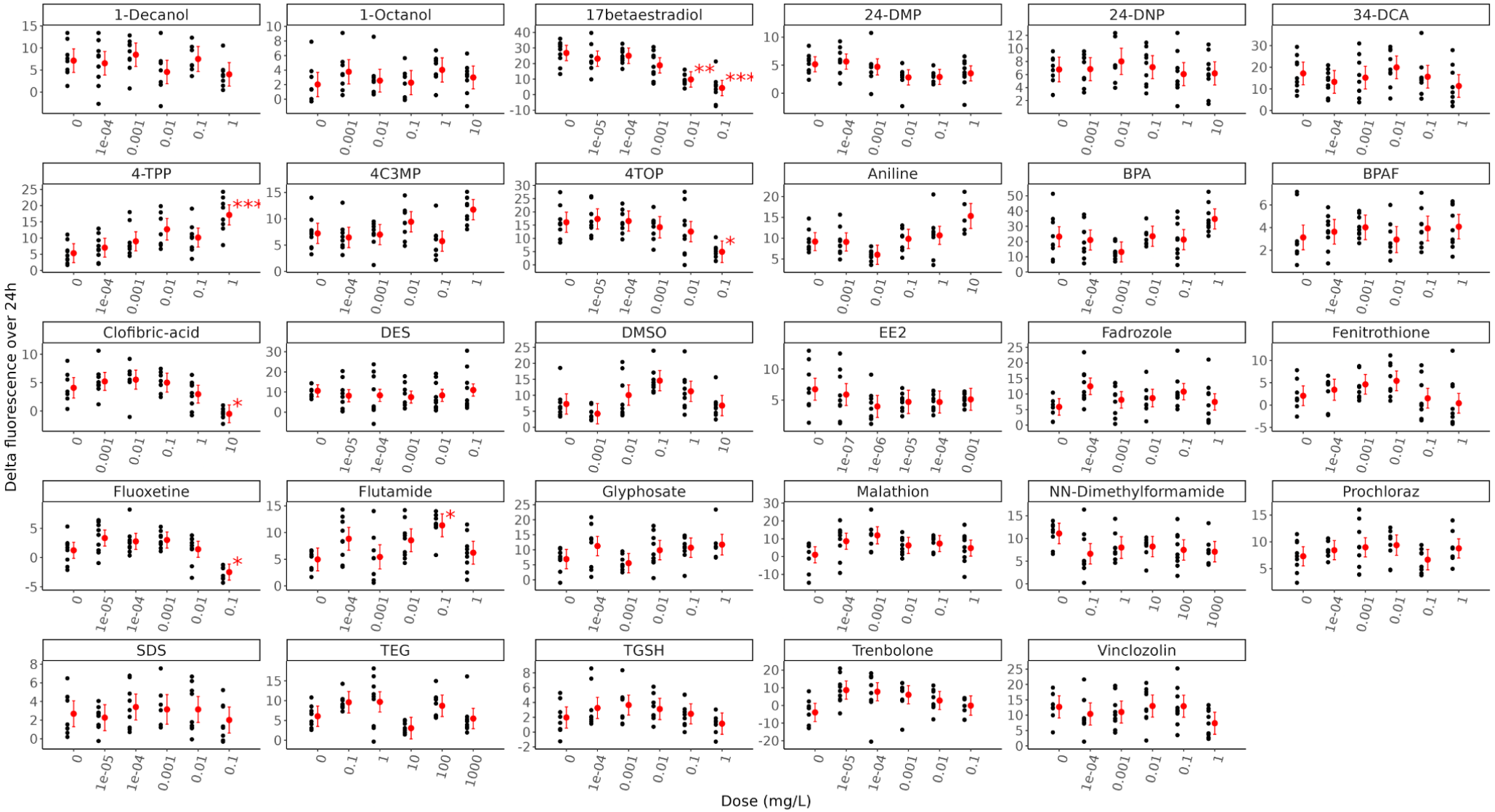
Dimethylformamide Dose 5, Prochloraz Dose 6). The magnitude of the difference between subgroups was investigated further by comparing the percentage difference in the estimate and the estimated marginal means between subgroups. The results showed strong differences in the modelled responses between subgroups for all 30 significantly different groups (> 25% difference). Potential interactions between the dose groups and subgroups were investigated further before conducting post hoc tests. A two-way ANOVA was conducted to investigate interaction effects between the subgroups and dose groups. The results of the two-way ANOVA revealed statistically significant interactions between the dose and subgroup variables for 4/29 chemicals after an FDR adjustment (EE2 [$p = 0.04$], Fentriothion [$p = 0.007$], Glyphosate [$p = 0.001$], NN-Dimethylformamide [$p = 0.01$]). The one-way ANOVA and two-way ANOVA models were compared for each chemical. Comparing these two models revealed significant differences between the models for 5/29 chemicals after FDR adjustment (DES - [$p = 5.39 \times 10^{-8}$], Fenitrothion – [$p = 7.08 \times 10^{-5}$], Glyphosate – [$p = 9.20 \times 10^{-4}$], Fadrozole - [$p = 0.01$] and NN-Dimethylformamide - [$p = 0.01$]). These results indicate that there could be significant interactions between the two independent variables (dose and subgroup) for the Fenitrothion, Glyphosate and NN-Dimethylformamide chemical exposure experiments. However, the interaction between the independent variables does not make sense from the experimental design; subgroups A, B and C were contained in one dose group and are theoretically identical.

The subgroup-dose group interaction effect is likely noise potentially from individual variation or resulting from inconsistencies in the nominal concentrations of chemicals in solution for the 24-hour period where larvae were in 96-well plates and alamarBlue® solution. Nonetheless, subgrouping effects were tested to validate the experimental design. The subgrouping effect could be prevented by increasing the sample size and analyzing the real

concentration of the exposure media in each Petri dish and 96-well assay plate, then incorporating the concentration variable in the statistical model. Therefore, ANOVA is a legitimate model for this data set (for most exposure experiments - 24/29) to estimate the response differences between dose groups.

With the model assumptions verified and subgrouping effects thoroughly investigated, the effects of chemical exposure on the **Δ**fluorescence response were investigated (Figure 3). The metabolic LOEC for each chemical is summarised in Table 3. The results of the one-way ANOVAs revealed significant effects of chemical exposure on the **Δ**fluorescence response in 12/29 experiments after an FDR adjustment (17b-estradiol, 4-TPP, 4C3MP, 4-TOP, Aniline, BPA, Clofibric acid, DMSO, Fluoxetine, Flutamide, TEG and Trenbolone). After an FDR adjustment, Dunnett's post hoc test revealed significant dose-response effects in 6/29 chemicals (17b-estradiol - Dose 1 and Dose 2 [$p = 2.25 \times 10^{-3}$ | $p = 1.08 \times 10^{-5}$], 4–TPP – Dose 1 [$p = 8.41 \times 10^{-4}$], 4-TOP - Dose 1 [p = 0.04], Clofibric acid - Dose 1 [p = 0.03], Fluoxetine - Dose 1 [p = 0.04] and Flutamide Dose 2 [p = 0.03]) (Figure 3).

Four chemical exposure experiments caused a decrease in metabolic activity (17b-estradiol, 4-TOP, Fluoxetine and Clofibric acid), and two chemical exposure experiments caused an increase in metabolic activity (4-TPP and Flutamide). The most frequently observed effects were in the highest-dose group, while others were in the second-highest-dose group.



Black dots represent response values (the change in fluorescence over 24h). Red points represent the estimated marginal means for each dose group (means adjusted for covariate in subgroups). Red error bars represent the 95% confidence intervals of the estimated marginal means. Red asterix represent statistical significance of the post hoc Dunnett's test (*** - $p < 0.001$, ** - $p < 0.01$, * - $p < 0.05$)

(title on next page)

Figure 3: The metabolic effects of chemical exposure were measured over 24 hours with the oxidative metabolism-dependent energy expenditure alamarBlue® assay. The metabolic response was summarized as the change in relative fluorescence units over 24 hours in alamarBlue® media. AlamarBlue® media changes from a blue colour to a pink colour as energy is consumed. An ANOVA and post hoc Dunnett's test of the **Δ**fluorescence (120hpf to 144hpf) response revealed statistically significant ($p < 0.05$) effects of 5-day chemical exposures in zebrafish larvae. If little-to-no energy is exerted over 24 hours in alamarBlue® media, **Δ**fluorescence is zero. Negative **Δ**fluorescence values are possible due to the normalization process, which corrects the response relative to the blank wells. The positive trend in the dose-response effect represents dose-dependent increases in metabolic activity. The negative trend in the dose-response effect represents dose-dependent decreases in metabolic activity.

2.4.2. BMD Analysis

BMR threshold

The heuristic BMR selection method calculated 5 thresholds in total. The five thresholds were derived from 1) the overt toxicity percent affected endpoint, 2) Pearson's r swimming behaviour correlation score endpoint, 3) Spearman's rho behaviour correlation score endpoint, and 4 & 5) the bi-directional z-scores from the oxidative metabolism-dependent energy expenditure endpoint.

The summarised overt toxicity responses ranged from 0 to 100%. The thresholds of 5% to 95% in increments of 5% were tested for the BMR approximation. The most appropriate BMR estimate was 15% for the overt toxicity data set. The summarised responses from the swimming behaviour assay ranged from 0% to -100%. Thresholds of -5% through -95% in increments of -5% were tested for the BMR approximation bootstrap. The BMR estimate for the Pearson swimming behaviour data set was -40% (Supplementary Figure 7). The estimated BMR for the Spearman swimming behaviour data set was also -40%. The summarised responses from the oxidative metabolism-dependent energy expenditure ranged bi-directionally from 0 to ±6. Thresholds of ±0.25 through ±5.75 in increments of ±0.25 were tested for the BMR approximation bootstrap. The BMR estimate for the negative direction was -1.75 (-30%). The BMR estimate for the positive direction was 2 (33%).

Overt toxicity

With a threshold level set to 15%, the BMD dose-response analysis was conducted in bootstrap using two methods: 1) the non-parametric 'Curvep' method (Supplementary Figure 10;

Table 3) and 2) the parametric Hill equation model method (Supplementary Figure 11; Table 3). 8/29 chemical exposures resulted in significant overt toxicity effects after processing the simulated dose-response curves (Supplementary Figure 9) with the 'Curvep' algorithm (2,4-DNP, 3,4-DCA, 4-TPP, DES, EE2, Malation, Trenbolone and Vinclozolin). 6/29 chemical exposures had significant effects after fitting the Hill model (2,4-DNP, 3,4-DCA, 4-TPP, DES, EE2 and Trenbolone).

However, after filtering for active chemicals (certainty factor > 50%), just 3/29 chemical exposures caused significant effects after processing the simulated dose-response curves with the 'Curvep' algorithm (2,4-DNP, 3,4-DCA, 4-TPP) and 4/29 chemical exposures resulted in significant effects after fitting the Hill equation model (2,4-DNP, 3,4-DCA, 4-TPP and DES).

Therefore, the BMD analysis of the o vert toxicity endpoint revealed that 4 chemicals (2,4-DNP, 3,4-DCA, 4-TPP and DES) caused significant overt toxicity effects.

Table 3: Summary of experimentally derived PODs for each chemical exposure experiment including BMDs with BMD_Ls and BMD_Us and LOECs.

	Overt toxicity (death and deformity rate)		
Chemical	Rcurvep	Hill	LOEC
1-Decanol	NA	NA	NA
1-Octanol	NA	NA	NA
17b-estradiol	NA	NA	NA
24-DMP	NA	NA	NA
24-DNP	**1.4384 (0.0056 - 1.4384)**	**1.2369 (1.0726 - 3.1086)**	10.00
34-DCA	**0.1569 (0.1569 - 0.1569)**	**0.2443 (0.1303 - 0.6562)**	1.00
4-TPP	**0.1429 (0.1429 - 0.1429)**	**0.2502 (0.141 - 0.6328)**	1.00
4C3MP	NA	NA	NA
4-TOP	NA	NA	NA
Aniline	NA	NA	NA
BPA	NA	NA	NA
BPAF	NA	NA	NA
Clofibric acid	NA	NA	NA
DES	0.0003 (0 - 0.0004)	**0.0011 (0.0001 - 0.041)**	NA
DMSO	NA	NA	NA
EE2	0.0007 (0 - 0.0007)	0.0005 (0.0001 - 0.001)	NA
Fadrozole	NA	NA	NA
Fenitrothion	NA	NA	NA
Fluoxetine	NA	NA	NA
Flutamide	NA	NA	NA
Glyphosate	NA	NA	NA
Malathion	0.791 (0.0791 - 0.791)	NA	NA
NN-Dimethylformamide	NA	NA	NA
Prochloraz	NA	NA	NA
SDS	NA	NA	NA
TEG	NA	NA	NA
TGSH	NA	NA	NA
Trenbolone	0.0003 (0 - 0.0003)	0.0012 (0.0002 - 0.0154)	NA
Vinclozolin	0.6602 (0.0001 - 0.6602)	NA	NA

Active BMDs with hit scores above 50% are highlighted in red. PODs reported in mg/L. Rcurvep is a non-parametric BMD model and Hill is a parametric BMD model. NA depicts that a model could not be fit.

Swimming behaviour

Pearson's r similarity score

With a threshold level set to -40%, the BMD dose-response analysis was conducted on the Pearson's r similarity scores in bootstrap using 1) the non-parametric 'Curvep' method (Supplementary Figure 13; Table 3) and 2) the parametric Hill equation model method (Supplementary Figure 14; Table 3). 23/29 Chemical exposures had significant swimming behaviour effects after processing the simulated dose-response curves (Supplementary Figure 12) with the 'Curvep' algorithm. 2/29 Chemical exposures resulted in significant effects after fitting the Hill equation model (4-TPP and Fluoxetine).

However, after filtering active chemicals (certainty factor > 50%), just one chemical exposure (4-TPP) caused significant effects ('Curvep' model).

Spearman's rho similarity score

With a threshold level set to -40%, the BMD dose-response analysis was conducted on the Spearman's rho similarity scores in bootstrap using 1) the non-parametric 'Curvep' method (Supplementary Figure 16; Table 3) and 2) the parametric Hill equation model method (Supplementary Figure 17; Table 3). 24/29 chemical exposures had significant swimming behaviour effects after processing the simulated dose-response curves (Supplementary Figure 15) with the 'Curvep' algorithm. 2/29 chemical exposures resulted in significant effects after fitting the Hill equation model (4-TPP and Fluoxetine).

However, after filtering just active chemicals (certainty factor > 50%), 2/29 chemical exposures caused significant effects after processing the simulated dose-response curves with the

‘Curvep’ algorithm (4-TPP and Fluoxetine), and the Fluoxetine exposure caused significant effects after fitting the Hill equation model.

Summary

Overall, the BMD analysis of swimming behaviour data revealed BMDs for two chemicals, 4-TPP and Fluoxetine, which suggests potential neurotoxic effects in zebrafish larvae after five days of exposure (Table 3).

Table 3 (cont'd): Summary of experimentally derived PODs for each chemical including BMDs with BMD_{LS} and BMD_{US} and LOECs.

	Swimming Behaviour				
	Pearson		Spearman		
Chemical	Rcurvep	Hill	Rcurvep	Hill	LOEC
1-Decanol	0.0071 (0.0029 - 0.6577)	NA	0.0322 (0.0009 - 0.6404)	NA	0.0001
1-Octanol	0.4151 (0.0484 - 3.349)	NA	0.8164 (0.0463 - 5.2809)	NA	0.01
17b-estradiol	0.0926 (0.0004 - 0.0926)	NA	0.0828 (0.0001 - 0.0828)	NA	NA
24-DMP	0.4625 (0.094 - 0.4625)	NA	0.3671 (0.0006 - 0.8289)	NA	NA
24-DNP	NA	NA	8.1708 (8.1708 - 8.1708)	NA	0.01
34-DCA	0.0069 (0.0001 - 0.0069)	NA	0.005 (0.005 - 0.005)	NA	1
4-TPP	**0.5348 (0.0048 - 0.9745)**	0.9836 (0.87 - 1)	**0.3912 (0.0058 - 0.9258)**	0.9481 (0.7478 - 0.9655)	1
4C3MP	0.3659 (0.0045 - 0.6001)	NA	0.4709 (0.0009 - 0.7955)	NA	1
4-TOP	NA	NA	NA	NA	0.001
Aniline	NA	NA	NA	NA	0.01
BPA	0.6078 (0.6078 - 0.9091)	NA	0.6618 (0.6618 - 0.6618)	NA	NA
BPAF	0.0085 (0.0001 - 0.397)	NA	0.3604 (0.0005 - 0.6842)	NA	NA
Clofibric acid	3.2695 (0.0312 - 5.0691)	NA	3.2239 (0.0303 - 6.6638)	NA	0.001
DES	NA	NA	0.0637 (0.0001 - 0.0637)	NA	0.1
DMSO	6.4094 (0.045 - 6.4094)	NA	NA	NA	10
EE2	0.0005 (0 - 0.0005)	NA	NA	NA	1E-06
Fadrozole	0.2586 (0.0042 - 0.4847)	NA	0.2643 (0.0073 - 0.6318)	NA	1
Fenitrothion	NA	NA	NA	NA	0.001
Fluoxetine	0.0497 (0.0007 - 0.0925)	0.0296 (0.0076 - 0.1)	**0.049 (0.0009 - 0.091)**	**0.0383 (0.0112 - 0.1)**	0.0001
Flutamide	0.4968 (0.0471 - 0.7997)	NA	0.4362 (0.0616 - 0.5195)	NA	0.001
Glyphosate	0.4638 (0.0047 - 0.7672)	NA	0.5331 (0.0042 - 0.9497)	NA	0.1
Malathion	0.3049 (0.3049 - 0.3049)	NA	0.323 (0.0001 - 0.323)	NA	0.001
NN-Dimethylformar	314.1705 (3.3675 - 878.9681)	NA	326.7297 (3.1801 - 930.8104)	NA	1000
Prochloraz	0.3806 (0.0086 - 0.3909)	NA	0.365 (0.0257 - 0.3812)	NA	1
SDS	0.0308 (0.0005 - 0.088)	NA	0.0065 (0 - 0.0348)	NA	NA
TEG	NA	NA	NA	NA	NA
TGSH	0.4568 (0.0003 - 0.4568)	NA	0.487 (0.0005 - 0.487)	NA	1
Trenbolone	0.0633 (0.0038 - 0.0669)	NA	0.0667 (0.005 - 0.0861)	NA	0.001
Vinclozolin	0.2919 (0.0006 - 0.7069)	NA	0.2722 (0.0005 - 0.736)	NA	1

Active BMDs with hit scores above 50% are highlighted in red. PODs reported in mg/L. Rcurvep is a non-parametric BMD model and Hill is a parametric BMD model. NA depicts that a model could not be fit. Pearson and Spearman are different methods of calculating similarity scores.

Oxidative metabolism-dependent energy expenditure

Negative direction - decreasing oxidative metabolism-dependent energy expenditure

With a threshold level set to -30%, the BMD dose-response analysis was conducted in bootstrap using 1) the non-parametric 'Curvep' method (Supplementary Figure 19; Table 3) and 2) the parametric Hill equation model method (Supplemental Figure 20; Table 3). 12/29 chemical exposures had significant energy expenditure effects after processing the simulated dose-response curves (Supplementary Figure 18) with 'Curvep' (17b-estradiol, 2,4-DMP, 2,4-DNP, 3,4-DCA, 4-TOP, Clofibric acid, DES, Fenitrothion, Fluoxetine, Flutamide, NN-Dimethylformamide, Vinclozolin). 5/29 chemical exposures had significant dose-response effects after modelling with the Hill equation (17b-estradiol, 4-TPP, DES, NN-Dimethylformamide and Vinclozolin).

However, after filtering active chemicals (certainty factor > 50%), 3/29 chemical exposures caused significant dose-response effects after processing the simulated dose-response curves with 'Curvep' (17b-estradiol, 4-TOP and DES), and 2/29 chemical exposures caused effects after modelling to the Hill equation (17b-estradiol and 4-TOP).

Positive direction - increasing oxidative metabolism-dependent energy expenditure

With a threshold level set to 33%, the BMD dose-response analysis was conducted in bootstrap using two methods: 1) the non-parametric 'Curvep' method (Supplementary Figure 22; Table 3) and 2) the parametric Hill equation model method (Supplementary Figure 23; Table 3). 14/29 chemical exposures had significant dose-response effects after processing the simulated dose-response curves (Supplementary Figure 21) with 'Curvep' (4-TPP, 4C3MP, Aniline, BPA, DES, DMSO, Fadrozole, Fenitrothion, Flutamide, Glyphosate, Malathion, Prochloraz, TEG, and

Trenbolone). 4/29 chemical exposures had significant effects after modelling the simulated dose-response curves to the Hill equation (4-TPP, 4C3MP, Flutamide and Trenbolone).

However, after filtering active chemicals (certainty factor > 50%), 2/29 chemical exposures caused significant dose-response effects after processing the dose-response curves with 'Curvep' (4-TPP and Flutamide), and 2/29 chemical exposures caused significant effects after modelling the Hill equation (4-TPP and Flutamide).

Summary

Overall, the BMD analysis revealed that exposure to 5 chemicals, 17b-estradiol, 4-TPP, 4-TOP, DES and Flutamide, induced metabolic disrupting/energy expenditure effects in zebrafish larvae after five days of exposure (Table 3). 17b-estradiol, 4-TOP, and DES exposure induced hypo-metabolic (negative) effects and exposure to 4-TPP and Flutamide induced hyper-metabolic (positive) effects.

Table 3 (cont'd): Summary of experimentally derived PODs for each chemical including BMDs with BMD_{LS}s and BMD_{US} and LOECs.

	Energy Expenditure (Metabolic)				
	Positive Direction		Negative Direction		
Chemical	Rcurvep	Hill	Rcurvep	Hill	LOEC
1-Decanol	NA	NA	NA	NA	NA
1-Octanol	NA	NA	NA	NA	NA
17b-estradiol	NA	NA	**0.0052 (0.0005 - 0.0376)**	**0.0029 (0.0009 - 0.0055)**	0.01
24-DMP	NA	NA	0.2761 (0.0027 - 0.9865)	NA	NA
24-DNP	NA	NA	5.1245 (0.4643 - 5.9652)	NA	NA
34-DCA	NA	NA	0.8036 (0.8036 - 0.8036)	NA	NA
4-TPP	**0.2196 (0.0003 - 0.7087)**	**0.025 (0.011 - 0.1266)**	NA	NA	1
4C3MP	0.9021 (0.0079 - 0.9125)	0.8844 (0.1542 - 0.9246)	NA	NA	NA
4-TOP	NA	NA	**0.0509 (0.001 - 0.0877)**	**0.0579 (0.046 - 0.0669)**	0.1
Aniline	2.7761 (0.2939 - 3.89)	NA	NA	NA	NA
BPA	0.9287 (0.9287 - 0.9287)	NA	NA	NA	NA
BPAF	NA	NA	NA	NA	NA
Clofibric acid	NA	NA	8.606 (5.8501 - 8.8115)	NA	10
DES	0.0207 (0 - 0.0332)	NA	**0.0006 (0 - 0.0765)**	0.0001 (0 - 0.0283)	NA
DMSO	0.0353 (0.0066 - 0.0849)	NA	NA	NA	NA
EE2	NA	NA	NA	NA	NA
Fadrozole	0.2843 (0.0001 - 0.2843)	NA	NA	NA	NA
Fenitrothion	0.0082 (0.0009 - 0.4159)	NA	0.8069 (0.0986 - 0.864)	NA	NA
Fluoxetine	NA	NA	0.0676 (0.0071 - 0.0833)	NA	0.1
Flutamide	**0.0056 (0 - 0.0398)**	**0.0037 (0.0003 - 0.2824)**	0.6795 (0.6795 - 0.6795)	NA	NA
Glyphosate	0.0993 (0.0048 - 0.2886)	NA	NA	NA	NA
Malathion	0.9282 (0.9282 - 0.9282)	NA	NA	NA	NA
NN-Dimethylformar	NA	NA	57.3868 (0.0488 - 499.6197)	0.1634 (0.0181 - 187.6191)	NA
Prochloraz	0.7616 (0.7616 - 0.7616)	NA	NA	NA	NA
SDS	NA	NA	NA	NA	NA
TEG	342.4766 (0.0454 - 342.4766)	NA	NA	NA	NA
TGSH	NA	NA	NA	NA	NA
Trenbolone	0 (0 - 0.0007)	0.0329 (0 - 0.1)	NA	NA	NA
Vinclozolin	NA	NA	0.5235 (0.0045 - 0.6855)	0.517 (0.1375 - 1)	NA

Active BMDs with hit scores above 50% are highlighted in red. PODs reported in mg/L. Rcurvep is a non-parametric BMD model and Hill is a parametric BMD model. NA depicts that a model could not be fit.Positive and negative directions describe the direction of effects on energy expenditure.

2.4.3. POD Summary

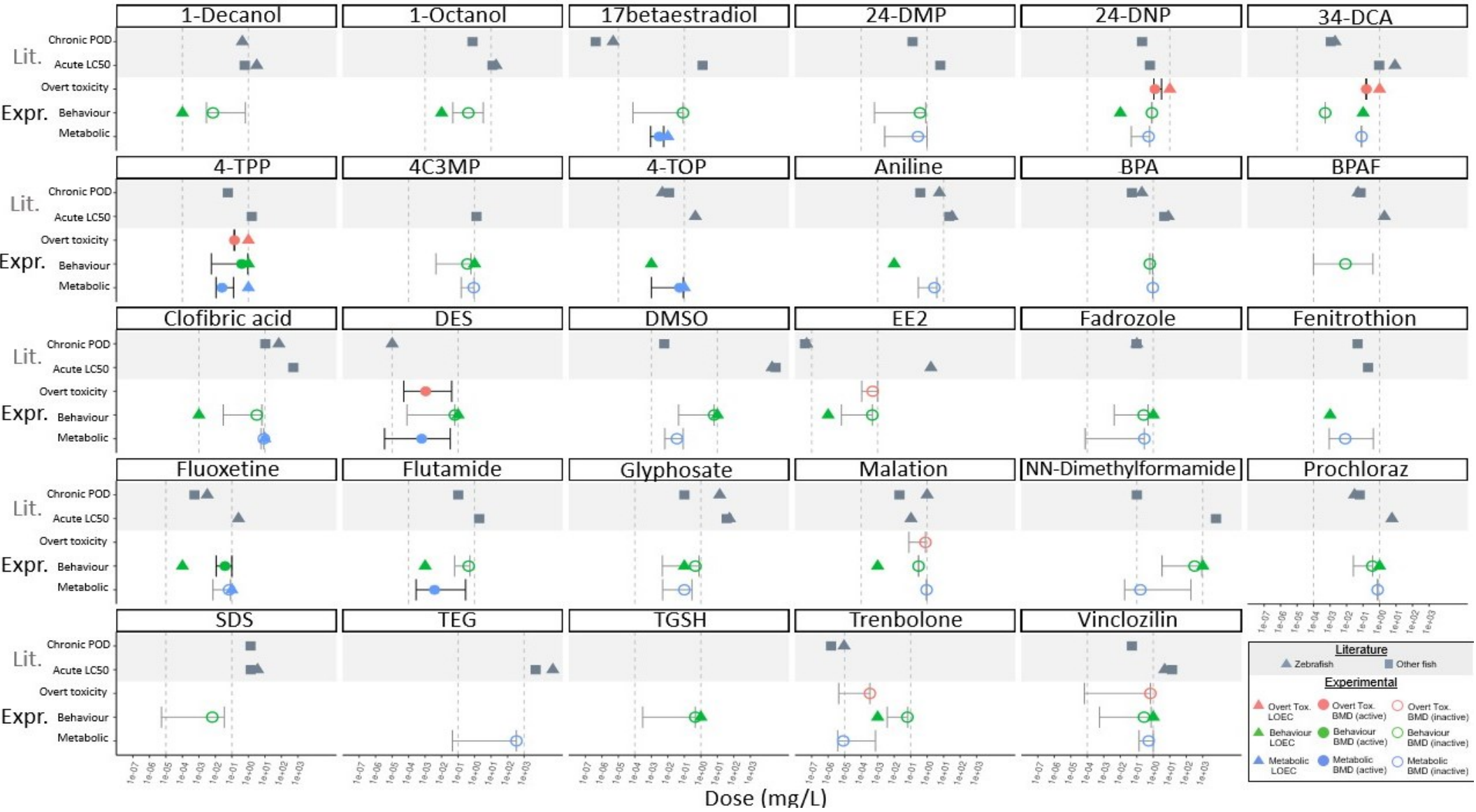
Experimental PODs

After completing the statistical LOEC and BMD analyses, the most sensitive PODs for each endpoint from each experiment were summarized in Figure 4.

All active (certainty factor > 50%) BMDs are summarized in Table 3, highlighted in red. Active BMDs in 3/29 chemicals revealed toxic metabolic effects related to oxidative metabolism-dependent energy expenditure when no overt toxicity effects were observed (17b-estradiol, 4-TOP, Flutamide). One active BMD from the Fluoxetine exposure experiment revealed potential neurotoxic effects related to swimming behaviour when no overt toxicity effects were observed. 4-TPP was the only chemical exposure with active BMDs in all 3 endpoints that induced overt toxicity, neurotoxicity and metabolic toxicity. DES exposure resulted in 2 active BMDs from the overt toxicity and metabolic activity endpoints.

All of the most sensitive statistical LOECs and BMDs from each endpoint are summarized in Figure 4. In total, the supplemented ZFET test identified 9 chemicals (17b-estradiol, 2,4-DNP, 3,4-DCA, 4-TPP, 4-TOP, Clofibric acid, DES, Fluoxetine and Flutamide) that induced toxic effects in zebrafish larvae after 5 days of exposure when excluding swimming behaviour LOECs (Figure 4). However, 2,4-DNP and 3,4-DCA were only overtly toxic. Including spurious swimming behaviour LOECs, 24 chemicals (1-Decanol, 1-Octanol, 17b-estradiol, 2,4-DNP, 3,4-DCA, 4-TPP, 4C3MP, 4-TOP, Aniline, Clofibric acid, DES, DMSO, EE2, Fadrozole, Fenitrothion, Fluoxetine, Flutamide, Glyphosate, Malathion, NN-Dimethylformamide, Prochloraz, TGSH, Trenbolone and Vinclozolin) induced toxic effects related to swimming behaviour and oxidative metabolism-dependent energy expenditure. Thus,

the swimming behaviour and oxidative metabolism-dependent energy expenditure PODs were more sensitive than the overt toxicity PODs since there were only 3 overt toxicity PODs overall (4-TPP, 2,4-DNP and 3,4-DCA) and 24 swimming behaviour and energy expenditure PODs.



(title on next page)

Figure 4: Summary of points of departure (PODs) from the literature review (highlighted with the grey background) and the most sensitive experimental PODs (chronic literature PODs include BMDs, LOECs, NOAELs, NOELs, and EC10s). BMDs with a hit rate above 50% were considered active and below 50% inactive. Overall, the FET test supplemented with behavioural and metabolic endpoints increased the sensitivity of the ZFET test even when deformities were included as an overt toxicity endpoint. However, the experimental PODs were not low enough to facilitate this methodology as a replacement for traditional toxicity test.

Experimental PODs vs. literature-based chronic and acute toxicity PODs

Experimental LOECs, BMDs and literature-based chronic and acute PODs from the literature review of juvenile and adult fish toxicity tests are summarized in Figure 4. Inactive BMDs were not considered biologically relevant and were ignored for the POD comparisons. Since experimental behavioural LOECs were spurious, POD comparisons were conducted with and without them.

When only overt toxicity was considered, no experimental PODs protected the chronic POD from adult and juvenile fish experiments in the literature. Only 2/29 experiments, 3,4-DCA and 4-TPP, protected the acute LC_{50} of traditional adult and juvenile fish toxicity tests in the literature.

When experimental PODs from the swimming behaviour and oxidative metabolism-dependent energy expenditure endpoint were considered (excluding swimming behaviour LOECs), 3/29 experimental PODs, 4-TPP, Clofibric acid and Flutamide, protected the chronic POD from traditional juvenile and adult fish toxicity tests. When swimming behaviour LOECs were not excluded, 11/29 experimental PODs protected the chronic POD from traditional juvenile and adult fish toxicity tests (i.e., PODs from 1-Decanol, 1-Octanol, 2,4-DNP, 4-TPP, 4-TOP, Aniline, Clofibric acid, Fenitrothion, Fluoxetine, Flutamide and Malathion).

When experimental PODs from the swimming behaviour and oxidative metabolism-dependent energy expenditure endpoint were considered (excluding swimming behaviour LOECs), 6/29 experimental PODs, 17b-estradiol, 4-TPP, 4-TOP, Clofibric acid, Fluoxetine and Flutamide, protected the acute LC_{50} of traditional adult and juvenile fish toxicity tests. However, with the swimming behaviour LOECs included, 19/29 experimental PODs protected the acute LC_{50} of traditional adult and juvenile fish toxicity tests (i.e., PODs from 1-Decanol, 1-Octanol,

17b-estradiol, 2,4-DNP, 3,4-DCA, 4-TPP, 4-TOP, Aniline, Clofibric acid, DMSO, EE2, Fenitrothion, Fluoxetine, Flutamide, Glyphosate, Malathion, NN-Dimethylformamide, Prochloraz, Vinclozolin).

2.5. Discussion

In this study, I integrated swimming behaviour and oxidative metabolism-dependent energy expenditure endpoints into the ZFET test to test the hypothesis that: 1) including behavioural and energy expenditure endpoints in the ZFET test produces PODs that are more sensitive and informative compared to only overt toxicity endpoints (i.e., dead and deformed) and 2) this integrated method is protective enough to substitute traditional animal experiments and act as an animal-alternative chemical hazard screening tool. I found that, on average, PODs from the swimming behaviour and oxidative metabolism-dependent energy expenditure endpoints were 32X more sensitive (had lower PODs) than the classical death and deformity endpoint. I also found that integrating information from both endpoints helped to elucidate potential mechanisms of action for neuroactive and metabolic-disrupting chemicals. Finally, I found that PODs from 3 chemicals protected the literature-based chronic POD from traditional juvenile and adult fish toxicity tests.

Swimming behaviour and oxidative metabolism-dependent endpoints increase the sensitivity of the ZFET test

Overt toxicity

The statistical LOECs from the overt toxicity endpoint only detected strong high-dose effects related to severe deformities and death. The overt toxicity BMDs were always more

sensitive (lower) than the overt toxicity statistical LOECs. Overt toxicity effects were statistically significant for the positive control chemical, 3,4-DCA. These results validate the exposure methods. Although there were only 3 overt toxicity LOECs (2,4-DNP, 3,4-DCA, 4-TPP) and 4 active (certainty factor > 50%) BMDs (2,4-DNP, 3,4-DCA, 4-TPP, DES) for the data set, it was expected that overt toxicity would not be frequently observed due to the selected dose ranges used for each chemical exposure. Overt toxicity PODs served as a reference to assess if the additional endpoints (swimming behaviour and oxidative metabolism-dependent energy expenditure) increased the sensitivity of the ZFET test.

Swimming behaviour

Generally, the statistical LOEC of the behavioural endpoint was the lowest POD overall. These results indicate that zebrafish larvae are responsive to the neurotoxic effects of chemical exposure at 5 days post-fertilization. The most frequently observed statistically significant swimming behaviour toxicity effects were in the dark phase, and the most frequently observed statistically significant dose group was the highest. A mix of hyperactive and hypoactive effects was observed due to chemical exposure. Hyperactive effects were 2X more common than hypoactive effects from the tested set of chemicals. However, the statistical LOECs were spurious. LOECs sometimes differed greatly between the light and dark phases (e.g., the light phase LOEC for Clofibric Acid was the highest dose while the LOEC for the dark phase was the lowest). In contrast, behavioural BMDs were not spurious, but the BMDs were less sensitive than the statistical LOECs and mostly inactive. Fundamental differences in the analyzed data sets likely drive the stark contrast between the BMDs and statistical LOECs. The statistical LOEC analysis binned the light and dark phases separately, while the BMD analysis considered the entire experiment duration regardless of phase. Despite the LOECs being spurious, visually

comparing the raw swimming behaviour data in Supplementary Figure 24 to the summarized swimming behaviour data with LOECs in Figure 2, it is evident that the effects are present.

Generally, the swimming behaviour LOECs and BMDs were lower than the overt toxicity LOECs and BMDs. In all instances where swimming behaviour effects were observed, LOECs and BMDs were lower than the overt toxicity LOECs and BMDs 21/23 times (i.e., 4-TPP and DES overt toxicity LOECs and BMDs were lower than swimming behaviour LOECs and BMDs).

There were 2 experiments where the overt toxicity BMD was lower than behavioural LOECs and BMDs (4-TPP & DES experiments). However, when excluding spurious LOECs and inactive BMDs, swimming behaviour active BMDs were lower than overt toxicity BMDs and LOECs 1/2 times (i.e., only 4-TPP and Fluoxetine had active BMDs). Overall, the swimming behaviour LOECs and BMDs were lower than overt toxicity LOECs and BMDs, but the swimming behaviour analysis methods were not robust.

Improving the sensitivity of the swimming behaviour BMDs and recommendations for the future

Although the swimming behaviour LOECs were the most sensitive POD for the entire experiment, the results were spurious, and statistical LOECs lacked 'activity' classifications (i.e., the > 50% certainty factor for the BMDs). Thus, I do not have much confidence in these statistical LOEC results. The standard deviations of the mean total swimming distance were large. Differences in individual fish swimming behaviour could contribute to the high variability in the data. Allowing fish to acclimatize longer to the 96-well plate in the Viewpoint Zebrabox environment could reduce variability in the data.

The swimming behaviour LOEC and BMD are not concordant most of the time. As previously mentioned, the non-concordance of the swimming behaviour LOECs and BMDs is

likely due to the nature of the data sets for each method; the statistical LOEC method divided the behavioural data sets by light or darkness phase, while the BMD method did not split the data set by light or darkness phase. One of the major challenges with this BMD analysis method is that it requires improvements to identify true positives more reliably. Fundamentally, the overall variability in the data greatly influences the sensitivity of the BMD analysis methods. A potential solution to this problem is splitting the data set up by light or dark conditions before transforming the data to similarity scores. However, this has been found to increase the identification of false negatives (Hsieh *et al.*, 2019). As well, the statistical LOEC analysis methods require improvements. This study used a p-value cut-off of 0.001 for the swimming behaviour LOEC to reduce the chances of false positives. However, applying a magnitude of effect filter may be more appropriate than a stricter p-value cutoff (e.g., a 10% difference from the control group or a % difference equal to the BMR).

A BMR threshold level of 25-30% is typical when the 'Curvep' BMD dose-response analysis technique is applied to zebrafish swimming behaviour data (Hsieh *et al.*, 2019); however, in this study, the BMR threshold was 40%. Increasing the number of doses and including lethal doses for all experiments would likely improve the sensitivity of the supplemented ZFET test and provide better overt toxicity PODs for the sensitivity comparisons. Conceptually, the reasoning behind this recommendation is that including a larger number of dose groups and a lethal dose group theoretically captures the entire shape of the dose-response curve. Including more dose groups and higher concentrations would improve the shape of dose-response curves, thus improving the dose-response models.

Adjustments could be made to the swimming behaviour assay protocol to improve the BMR threshold level and reduce variability in the data. Since writing this, other Mennigen lab

members have seen improved variability by adjusting the protocol acclimation time to 30 instead of 20 minutes. Reducing batch effects would also likely decrease the background noise from genetic variations in the breeding pairs. However, this is, in some cases, limited to fish husbandry and exposure logistics. Variations in swimming behaviour from individuals could also contribute to the background noise of the behavioural data. To better account for these variations, increasing the sample size of every dose group (n = 16 instead of n = 9) could reduce these effects. Finally, the nominal concentrations of the exposure media could be measured to validate the results and further improve the statistical model to capture the true concentration-response effects. These method adjustments may reduce the background noise response in the data, decrease the behavioural BMR threshold to a more reasonable 25-30% threshold, and improve the statistical results. Decreasing the BMR threshold level would likely increase the sensitivity and certainty of the methods.

Oxidative metabolism-dependent energy expenditure

Generally, the oxidative metabolism-dependent energy expenditure endpoint BMDs were lower than the statistical LOECs. The non-parametric 'RCurvep' modelling derived approximately the same amount of active BMDs (certainty factor > 50%) as the Hill equation modelling method (5 vs. 4, respectively). The non-parametric 'RCurvep' modelling method was less variable than the parametric Hill equation modelling method, and the Hill equation modelling method was more sensitive than the 'RCurvep' modelling method. However, the BMD analysis (both modelling methods) failed to detect active (certainty factor > 50%) metabolic effects in two chemicals, Clofibric acid and Fluoxetine, where a statistical LOEC was observed (Figure 4 & Table 3).

Generally, the oxidative metabolism-dependent energy expenditure (blue) BMDs and LOECs were more sensitive (lower) than the overt toxicity (red) BMDs and LOECs. When oxidative metabolism-dependent energy expenditure effects were observed, they were more sensitive than overt toxicity effects 7/7 times (including LOECs and active BMDs). The oxidative metabolism-dependent energy expenditure endpoint was less variable than the behaviour endpoint, resulting in more certain/active BMDs.

Improving the sensitivity of the metabolic BMDs and recommendations for the future

Although the BMD was generally the most sensitive POD for the oxidative metabolism-dependent energy expenditure endpoint, there were sensitivity issues, as mentioned previously (Clofibric acid and Fluoxetine exposures caused statistically significant effects in the highest dose group, but BMDs were inactive). By decreasing the background noise level in the oxidative metabolism-dependent energy expenditure data set, the BMR threshold level would likely decrease, and the incidence of positive results (active BMDs) would improve. In addition to adjustments mentioned in the above section on improving the swimming behaviour data (i.e., reducing batch effects, increasing the sample size to reduce individual variation and validating the nominal concentrations of the exposure media), an effort should be made to reduce the background noise resulting from the stability of the reagents used in the assay. Reagents should be made fresh each day from the same stocks. Potential fungal contamination in the system water could also contribute to biological response noise. If possible, fresh 96-well plates (used in this chapter) and anti-fungal media (not used in this chapter) should be utilized for the alamarBlue® assay. However, preliminary results from other members of the Mennigen lab showed that using anti-fungal media interferes with the alamarBlue® assay because the hue of commonly used anti-fungal aquaculture media (methylene blue) is also blue.

Summary - Sensitivity

By conducting a bootstrapped BMD analysis and adding swimming behaviour and oxidative metabolism-dependent energy expenditure endpoints to the ZFET test, we improved the sensitivity of the standard ZFET test and provided certainty measures for the results. Figure 4 shows that swimming behaviour toxicity and oxidative metabolism-dependent energy expenditure toxicity PODs are more sensitive than overt toxicity PODs despite the analysis methods needing improvements (green, blue and red, respectively in Figure 4).

On average, the most sensitive swimming behaviour or oxidative metabolism-dependent energy expenditure POD (including LOECs and excluding inactive BMDs) was ~32X lower than the overt toxicity POD when present (24-DNP, 34-DCA, 4-TPP, and DES). Swimming behaviour LOECs were included because active BMDs were unavailable for 2,4-DNP and 3,4-DCA to compare PODs quantitatively. For 2,4-DNP, the swimming behaviour LOEC (0.01mg/L) was 120X more sensitive than the lowest overt toxicity BMD (1.2mg/L). For 3,4-DCA, the swimming behaviour LOEC (0.1mg/L) was 1.6X more sensitive than the lowest overt toxicity BMD (0.16mg/L). For 4-TPP, the oxidative metabolism-dependent energy expenditure BMD (0.025mg/L) was 5.6X more sensitive than the overt toxicity BMD (0.14mg/L). Finally, for DES, the oxidative metabolism-dependent energy expenditure BMD (0.0006mg/L) was 1.7X more sensitive than the lowest overt toxicity BMD (0.001mg/L).

In many instances, significant swimming behaviour and oxidative metabolism-dependent effects were observed even when significant overt toxicity effects were not observed (20 times), providing further evidence that these endpoints enhanced the sensitivity of the ZFET. Including spurious swimming behaviour PODs, the swimming behaviour and oxidative metabolism-dependent energy expenditure PODs were more sensitive than the overt toxicity PODs for 24/24

chemicals (i.e., 2,4-DMP, BPA, BPAF, SDS, and TEG did not significantly affect embryos). Excluding spurious swimming behaviour LOECs, swimming behaviour and oxidative metabolism-dependent energy expenditure PODs were more sensitive than overt toxicity PODs for 7/9 chemicals (i.e., 17b-estradiol, 4-TPP, 4-TOP, Clofibric acid, DES, Fluoxetine and Flutamide were the only chemicals that significantly affected embryo swimming behaviour and oxidative metabolism). 2,4-DNP and 3,4-DCA were the only chemicals where overt toxicity PODs were the most sensitive. Overall, the evidence supports the hypothesis that including swimming behaviour and oxidative metabolism-dependent energy expenditure endpoints in the ZFET test produces more sensitive (lower) PODs than when only overt toxicity endpoints are considered.

The swimming behaviour LOECs were the most sensitive POD from the entire experiment. The LOECs were spurious and non-concordant with the active (certainty factor > 50%) swimming behaviour BMDs. The active (certainty factor > 50%) swimming behaviour BMDs were not sensitive and were sparse due to high variability in the response data (i.e., 4-TPP and Fluoxetine were the only chemicals with active BMDs). Thus, caution should be exercised when interpreting the swimming behaviour analysis results. However, despite the analysis methods needing improvements, swimming behaviour effects can be seen in the raw total swimming distance (mm) data in Supplementary Figure 24 (e.g., 3,4-DCA, 4-TPP, Aniline, Clofibric acid, DMSO, Fadrozole, Fluoxetine, Flutamide, Prochloraz & Vinclozolin all show visually stark differences in the highest dose group vs control groups).

The oxidative metabolism-dependent energy expenditure PODs were generally the best-performing overall. The methods of calculating the metabolic PODs did well in balancing sensitivity and robustness. The BMDs were overall lower than the statistical LOEC, except for

the Clofibric acid and Fluoxetine exposures. In every instance where metabolic effects were observed, they were more sensitive than the effects of overt toxicity. Overall, the metabolic endpoint was more sensitive than overt toxicity endpoints. This is the first time anyone has done a BMD analysis with oxidative metabolism-dependent energy expenditure data from the alamarBlue® assay.

Overall, including swimming behaviour and oxidative metabolism-dependent energy expenditure endpoints in the ZFET test improves the protocol's sensitivity compared to when only the overt toxicity endpoints are considered.

Swimming behaviour and oxidative metabolism-dependent energy expenditure endpoints increase the informativeness of the ZFET test

Many chemical exposure experiments induced neurotoxic (behavioural) and metabolic-disrupting (energy expenditure) effects. 17b-estradiol, 4-TOP, Clofibric acid, Fluoxetine and Flutamide disrupted the oxidative metabolism-dependent energy expenditure or normal swimming behaviour of fish unrelated to overt toxicity (i.e., no overt toxicity PODs were derived, but swimming behaviour and oxidative metabolism-dependent energy expenditure PODs were derived). Known metabolism-disrupting chemicals/endocrine-disrupting chemicals such as DES, Flutamide and the primary female sex hormone 17b-estradiol were all found to cause metabolic effects after five days of exposure (Table 1) (Sun *et al*., 2022; Papalou *et al*., 2019).

Adding the swimming behaviour and oxidative metabolism-dependent energy expenditure endpoints to the ZFET test provided more mechanistic information than is standard for the ZFET test. To illustrate this point, one can consider the Fluoxetine chemical exposure

experiment as an example (Table 1). In the past, researchers have found that zebrafish embryos exposed to Fluoxetine have their locomotor activity affected at concentrations as low as 0.00088mg/L (Airhart *et al.*, 2007; Al Shuraiqi *et al*., 2021; De Farias *et al.*, 2019; Zindler *et al*., 2020). De Faris *et al*. proposed that potential mechanisms of action were the inhibition of cholinesterase activity, which can cause myopathy of skeletal muscles and loss of motility, inhibition of serotonin transporter activity known to affect spontaneous swimming activity, and induction of hypercortisolism. This chapter found that zebrafish larvae's average total swimming distance (mm) was significantly affected by Fluoxetine exposure at the dose of 0.0001mg/L, and oxidative metabolism-dependent energy expenditure was also significantly affected by Fluoxetine exposure at the dose of 0.1mg/L (Figure 4). These results support the hypothesis that Fluoxetine exerts its toxic effects by inhibiting cholinesterase and the serotonin transporter activity (Airhart *et al.*, 2007; De Farias *et al.*, 2019; Diaz-Camal *et al*., 2022; Zindler *et al*., 2020). Inhibition of these mechanisms could directly or indirectly affect zebrafish larvae's swimming behaviour and oxidative metabolism-dependent energy expenditure. Exposure to Fluoxetine may be directly neurotoxic to zebrafish larvae through inhibiting serotonin transporters (Airhart *et al.*, 2007; De Farias *et al.*, 2019; Diaz-Camal *et al*., 2022; Zindler *et al*., 2020). It may also indirectly disrupt the swimming behaviour and oxidative metabolism-dependent energy expenditure of zebrafish larvae through myopathy related to the inhibition of cholinesterase activity (De Farias *et al*., 2019). Although the Fluoxetine exposure experiment is a good example to illustrate the point of supplementing the standard ZFET test with swimming behaviour and oxidative metabolism-dependent energy expenditure improves the information provided, the methods used in this chapter are only capable of detecting effects associated with

swimming behaviour and energy expenditure. Additional methods would be required to detect other mechanisms of action.

Improving the informativeness of the supplemented ZFET test and recommendations for the future

In the future, to verify these conclusions regarding the mechanisms of action, transcriptomics can measure transcript abundance levels related to these mechanisms. This may involve assessing transcripts like those related to serotonin transporters or cholinesterase to correlate the upregulation or downregulation of transcripts to the biological effects of chemical exposure.

Additionally, the swimming behaviour assay provides alternative endpoints that can be analyzed to further corroborate known mechanisms of action. For example, Fluoxetine affects the spontaneous swimming behaviour of fish (De Farias *et al.*, 2019). To elucidate whether Fluoxetine exposure significantly affects spontaneous swimming activity, alternative endpoints, such as the duration of small bursts of swimming behaviour, can be analyzed rather than the total swimming distance. Analyzing more nuanced swimming behaviour endpoints can provide more specific information about perturbations in swimming behaviour.

To improve the ability of the alamarBlue® assay to detect changes in oxidative metabolism-dependent energy expenditure, integrating the swimming behaviour assay with the alamarBlue® assay provides more specific information about perturbations in metabolism. It addresses the confound of overall swimming activity, which the swimming behaviour assay can also detect.

Supplementing the ZFET test swimming behaviour and oxidative metabolism-dependent energy expenditure endpoints alone may not be a suitable substitute for juvenile or adult fish acute or chronic overt toxicity tests

Protectiveness

To determine if the methods are sensitive enough to substitute traditional juvenile and adult fish toxicity testing, the protectiveness of the experimental early-stage zebrafish PODs were assessed (Figure 4). The experimental PODs were compared to acute and chronic juvenile and adult fish overt toxicity PODs (death, deformity and reproduction). The acute toxicity PODs were LC_{50}s (lethal concentration for 50% of the population), and the chronic toxicity PODs were a mix of BMDs, LOECs, No observed [adverse] effect concentrations/levels (NOECs/NOELs), and Effective concentrations for 10% of the population (EC_{10}s). An experimental POD protects an outcome (e.g., chronic toxicity) when the POD is equal to or more sensitive (lower) than the literature-based outcome POD (e.g., the most sensitive chronic toxicity POD from the literature).

Inactive (< 50% certainty factor) BMDs were not considered biologically relevant and were ignored for the POD comparisons. Since experimental behavioural LOECs were spurious, POD comparisons were made with and without them.

When experimental PODs from the swimming behaviour and oxidative metabolism-dependent energy expenditure endpoint were considered (excluding swimming behaviour LOECs), 3/29 experimental PODs protected the chronic POD from traditional juvenile and adult fish toxicity tests (4-TPP, Clofibric acid and Flutamide) and 6/29 experimental PODs protected the acute LC_{50} of traditional adult and juvenile fish toxicity tests (17b-estradiol, 4-TPP, 4-TOP, Clofibric acid, Fluoxetine and Flutamide).

When swimming behaviour LOECs are not excluded, 11/29 experimental PODs protected the chronic POD from traditional juvenile and adult fish toxicity tests (1-Decanol, 1-Octanol, 2,4-DNP, 4-TPP, 4-TOP, Aniline, Clofibric acid, Fenitrothion, Fluoxetine, Flutamide and Malathion) and 19/29 experimental PODs protected the acute LC_{50} of traditional adult and juvenile fish toxicity tests (i.e., PODs from 1-Decanol, 1-Octanol, 17b-estradiol, 2,4-DNP, 3,4-DCA, 4-TPP, 4-TOP, Aniline, Clofibric acid, DMSO, EE2, Fenitrothion, Fluoxetine, Flutamide, Glyphosate, Malathion, NN-Dimethylformamide, Prochloraz, Vinclozolin).

Therefore, including metabolic and behavioural endpoints improved both the sensitivity of the ZFET test and the protectiveness of the PODs. However, the methods were not protective enough (even when spurious swimming behaviour LOECs were considered) to substitute traditional chronic or acute juvenile and adult fish toxicity tests for 10 of the chemicals tested.

Improving protectiveness and recommendations for the future

Two ways to improve the protectiveness of the PODs are: 1) improve the sensitivity of the experimental PODs by improving the protocol and analysis methodologies and reducing variability, and 2) improve the literature review methodology (especially important for data-poor substances). It was discussed previously how improvements can be made to the protocol and analysis methodologies that may increase the sensitivity, so they will not be discussed further. However, a refined systematic literature review could be conducted to collect relevant PODs with a more fine-toothed comb. Systematic literature reviews focus on reviewing the literature relevant to a narrow research question. Whereas a more systematic literature review may have only included directly comparable PODs (i.e., LOECs and BMDs) in the search results, the literature review conducted in this study considered BMDs, NOELs/NOECs, LOECs and EC10s because they were commonly available in the literature. However, the literature review approach

used in this thesis likely resulted in more conservative (higher potency) literature-based PODs. NOELs/NOECs are inherently more sensitive than the LOEC because NOELs/NOECs will be the highest tested concentration that was not statistically significant. It is also true that instead of comparing experimental LOECs to NOELs/NOECs and EC10s in the literature, experimental NOELs/NOECs could be compared instead. Regardless of the approach taken, a more direct comparison of PODs could improve the protectiveness of the assay.

2.6 Conclusion

Overall, supplementing the internationally standardized ZFET test (OECD TG 236) with swimming behaviour and oxidative metabolism-energy expenditure endpoints increased the sensitivity of the toxicity test while also providing helpful information about the potential mechanisms of action of the tested compounds. At best, the early-stage zebrafish experimental PODs were protective of acute toxicity effects in juvenile and adult fish for 19/29 chemical exposure experiments and protective of chronic toxicity effects in juvenile and adult fish for 11/29. However, when excluding spurious LOECs, PODs were protective of acute toxicity effects in juvenile and adult fish for 6/29 chemical exposure experiments and protective of chronic toxicity effects in juvenile and adult fish for 3/29. The supplemented ZFET test may be a helpful chemical hazard screening tool for exploring chemicals' potential neurotoxic, developmental and metabolism-disrupting effects. Although supplementing the standard ZFET test with additional endpoints improved the sensitivity of the toxicity test, the results were not protective enough of acute or chronic toxicity effects in juvenile and adult fish to substitute traditional juvenile and adult fish toxicity tests on their own. However, incorporating additional endpoints, such as transcriptomics, could improve the protectiveness of the early-stage zebrafish

experimentally derived PODs and enrich information about potential chemical mechanisms of action. The inclusion of a transcriptomic endpoint will be explored further in Chapter 3.

Chapter 3 - Adding a Transcriptomic Endpoint to the ZFET Test Provides a POD that Protects Against Chronic Overt Toxicity Effects from Juvenile and Adult Fish

3.1. Abstract

New chemicals and legacy substances require comprehensive hazard and risk assessment, but traditional fish toxicity tests are resource-intensive and unethical. The standard zebrafish Organisation for Economic Co-operation and Development Fish Embryo acute Toxicity (FET) test (TG 236) is a potential alternative to adult animal testing. Environment and Climate Change Canada and Health Canada are interested in determining whether short-term changes in zebrafish embryo gene expression can be predictive or protective indicators of chronic toxicity effects. Here, zebrafish embryos were exposed to 10 chemicals for 5 days (120 hours post fertilization), flash-frozen, and sequenced using Illumina whole-transcriptome sequencing. Differentially expressed genes were benchmark dose-response modelled, and transcriptional points of departure (PODs) were derived per the 'Omics Data Analysis Framework for Regulatory applications and standard US National Toxicology Program guidelines. Experimental PODs from Chapters 2 and 3 and literature-based PODs were compared to assess sensitivity and protectiveness. Transcriptomic PODs were derived despite the BMDs from Chapter 2 being non-active. These results suggest that the transcriptomic endpoint was the most sensitive in the integrated zebrafish FET test. Transcriptomic PODs protected chronic overt toxicity effects in juvenile and adult fish 6/7 times (excluding EE2, Trenbolone and TGSH, which were omitted

from the POD comparison). In conclusion, transcriptomic PODs are strong protectors of environmental and human health. This integrated New Approach Methodology is valuable for chemical hazard screening, providing critical insights into potential mechanisms of action for informed decision-making by risk assessors and regulators to protect the environment and well-being of Canadians.

3.2. Introduction

Hundreds of new substances are introduced to global consumer markets yearly, and thousands of legacy substances remain in the environment and are constantly re-purposed (Arnold, 2015; Basu *et al.*, 2019; Bernard *et al.*, 2017; Woodruff *et al.*, 2023). These substances require thorough safety testing. However, traditional test methods that use adult and juvenile animals are low throughput. They are too slow and costly to address the many chemicals requiring risk assessment (Basu *et al.*, 2019; Government of Canada, 2019). Recently, regulatory agencies, including those in Canada, have taken steps to modernize the chemical hazard screening process and adopt New Approach Methodologies (NAMs), which are typically cheaper, high-to-medium throughput, rapid and more ethical (Bill S-5, 2023).

The Zebrafish Embryo Acute Toxicity (ZFET) test is a cost-effective, time-saving way to test toxicity. It involves exposing zebrafish embryos and larvae for 5 days and scoring developmental effects. The Organisation for Economic Cooperation and Development (OECD) and regulatory agencies standardized the ZFET test (TG 236). However, the standard ZFET test is best suited for detecting acute toxic effects (i.e. overt toxicity from short-term exposures). It is generally not sensitive enough to detect the more subtle toxic effects caused by low-dose long-term exposures (i.e. chronic effects) (Sobanska *et al.*, 2018; von Hellfeld *et al.*, 2020).

Transcriptomics provides a high-throughput means of producing large data sets of gene expression that cover a broad range of biological responses to hazardous chemical exposure (Reardon *et al.*, 2023; Johnson *et al*., 2022). Transcriptomics can detect molecular effects soon after the initiation of chemical exposure, at much shorter times and lower concentrations than is required to detect acute effects. These molecular effects can be analyzed to predict whether a substance will likely induce chronic toxicity. Further, a transcriptional point of departure (tPOD) can be calculated, which can be used in quantitative risk assessment to estimate concentrations likely to cause chronic effects. Benchmark dose (BMD) modelling of a gene's dose-response relationship is applied to derive tPODs. Transcriptional PODs correlate with adverse outcome levels from traditional toxicity tests, such as cancer outcome levels (Thomas *et al.*, 2013; Webster *et al.*, 2015; Pagé-Larivière *et al.*, 2019).

This chapter investigates whether including transcriptomics in the ZFET test can generate tPODs that are protective (i.e. lower) of chronic toxicity in adult fish. Chapter 3 includes a standard BMD analysis that derives gene-level and gene set-level tPODs (National Toxicology Program, 2018). The lowest tPOD is then compared to acute and chronic overt toxicity PODs in adult fish from the literature to assess their protectiveness. In addition to comparing the tPODs to the overt toxicity PODs in the literature, this study also compared novel behavioural and metabolic PODs (as described in Chapter 2) to assess each method's sensitivity.

3.3. Materials and Methods

The same materials and methods used in Chapter 2 were used in Chapter 3. Please see the Animal Husbandry and Chemicals sections of Chapter 2 for more information on those topics. This chapter included the following chemicals in the transcriptomic analysis: 1-Decanol, Aniline,

EE2, Fadrozole, Fenitrothion, Glyphosate, Malathion, Prochloraz, TGSH and Trenbolone. Please see Table 1 and Table 2 from Chapter 2 for more information on the mechanisms of action of each chemical and the concentrations used in the exposure experiments. These 10 chemicals were chosen for Chapter 3 in late 2022 based on preliminary data from Chapter 2. A diverse set of mechanisms of action were selected through these 10 chemicals to investigate whether the use of transcriptomics is a good chemical hazard screening strategy. Due to the financial constraints of whole-transcriptome RNA sequencing, only 10 chemicals were chosen, and not all 29.

ZFET test and chemical exposure

The standardized ZFET test (OECD TG236) protocol was followed as previously described in the Chapter 2 Methods section (Experimental design - ZFET test and chemical exposure). Briefly, embryos were exposed to the test compounds from 0dpf to 5dpf with daily semi-static renewal of media (system water + dissolved test compound). Five dose groups (10-fold serial dilution) plus a control group were tested for each compound in triplicate. Each replicated comprised a pool of 5 embryos. Pooled replicates were collected at the end of the exposure period with approximately 150µL exposure media and were flash-frozen on dry ice and stored at -80°C. Backup samples (also containing 5 zebrafish embryos each) were collected, flash-frozen and stored to minimize the risk of accidental sample loss and in case quality control standards were not met by the first 'primary' sample.

RNA Extraction and Quality Control (QC)

Total RNA was extracted using RNEasy Mini Kits (Cat. No. / ID: 74104) and RNase-free DNase treatment (Cat. No. / ID: 9256) to deplete DNA and reduce the risk of genomic contamination. Briefly, zebrafish embryos were homogenized in 350uL RNeasy Lysis Buffer

(RLT) buffer for 4 minutes and centrifuged to extract RNA from a sample. The supernatant was transferred to an RNase-free 2mL tube, and 350µl of 70% EtOH was added. 700µl of the sample was transferred to an RNase-free spin column. 80µl of RNase-free DNase incubation mix was added to the spin column and incubated on the benchtop for 15 minutes. The sample was centrifuged at 10,000 revolutions per minute (RPM) for 15s, and the flow-through was discarded. The spin column was washed once with 350uL of RNeasy wash buffer 1 (RW1) and twice with 500uL of RNeasy concentrated wash buffer (RPE). The first two washes were spun at 10,000 RPM for 15s, and the final spin was done at 10,000 RPM for 120s. The flow-through was discarded. Extracted RNA was eluted into 30µl of RNAse-free water by adding the water to the spin column and centrifuging at 10,000 RPM for 1 minute. Sample quantity was measured using the Qiagen QiAxpert Instrument (Cat. No. / ID: 9002340). Sample quality was measured using the QIAxcel Instrument and QIAxcel RNA Quality Control Kit 2.0 (Cat. No. / ID: 929104). The concentration and RIN of each sample were recorded (Supplementary Information - Copy of RNA Extractions) before diluting the samples to 33.3ng/µl (as requested by Genome Quebec for library preparation inputs). A minimum RIN threshold of 6 was used.

500ng of total RNA (15 µl at 33.3 ng/µl) was sent to Genome Quebec for Illumina mRNA stranded Whole transcriptome paired-end RNA sequencing (100bp) with Polyadenylation (PolyA) tail capture and ligation-based addition of adapters and indexes on a NovaSeq 6000 system and S4 flow cell. Genome Quebec quality-checked each sample again using Nanodrop and Tapestation systems before mRNA stranded library preparation. After mRNA stranded library preparation, samples were quality control checked again using a Bioanalyzer system. If library concentrations were below 3nM, they were removed from the

study. Samples passing quality control measures were paired-end sequenced using a Novaseq 6000 system and S4 flow cell.

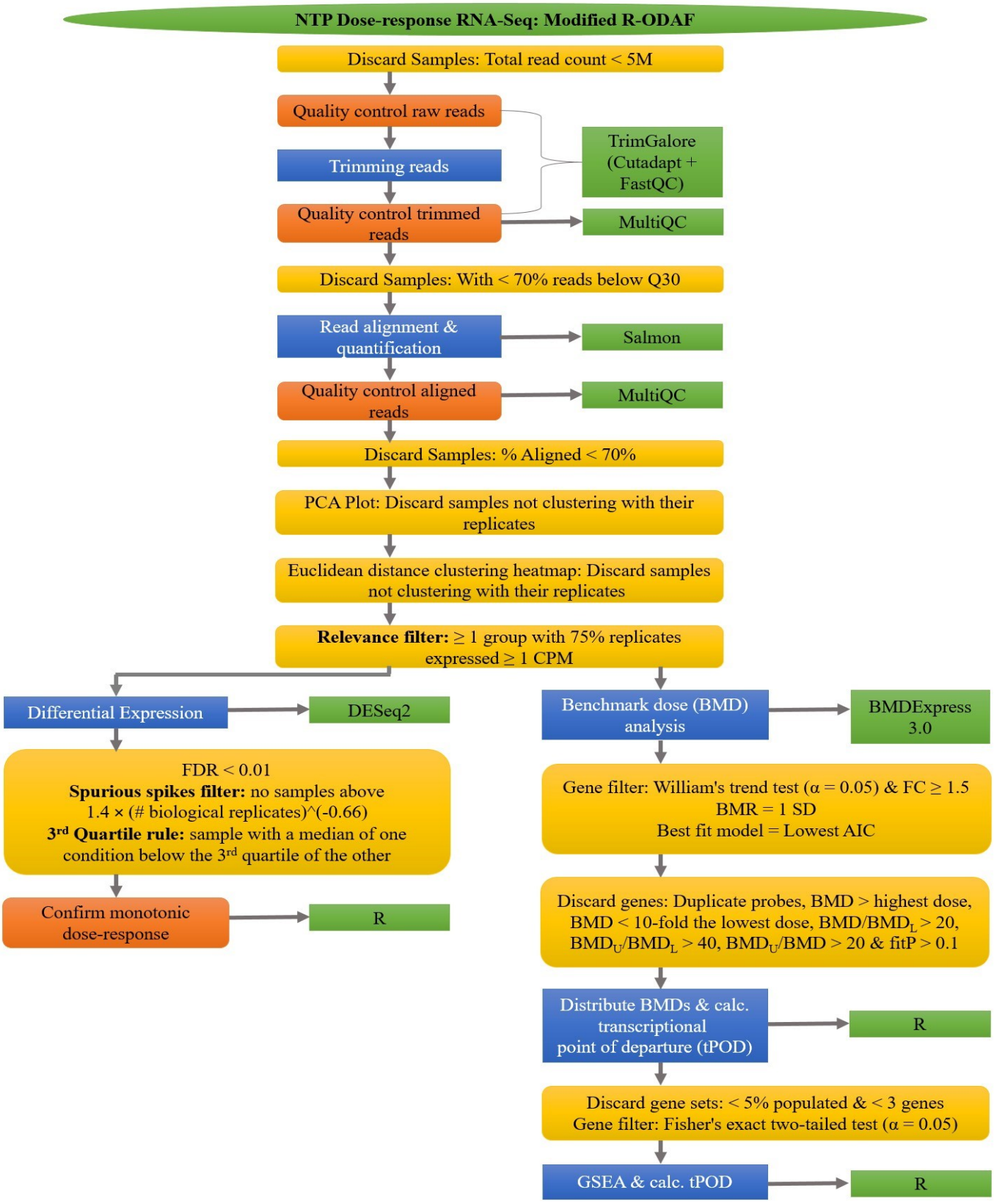
RNAseq pipeline

FastQ files were downloaded from Genome Quebec's Nanuq portal servers using GNU Wget. The Nextflow core RNAseq bioinformatics pipeline was used to process, pseudo-align, quantify and quality-check RNAseq FastQ files (Patel, 2023). The upstream processing of the RNAseq data via Nextflow was enabled in part by support provided by Compute Ontario (https://www.computeontario.ca/) and the Digital Research Alliance of Canada (https://alliancecan.ca).

The Nextflow RNAseq bioinformatics pipeline is a workflow framework for integrating programs and tools into one cohesive pipeline that is 1) portable, 2) reproducible, 3) flexible, and 4) checkpointed (Ewels *et al.*, 2020). 1) Portability: the Nextflow RNAseq pipeline is portable to Windows (with extra steps), Mac and Linux operating systems. 2) Reproducibility: stable versions of the Nextflow RNAseq pipeline are available on GitHub, and the pipeline comes with docker containers, making installation easy. Nextflow is readily available on Digital Research Alliance of Canada High-Performance Compute clusters natively as a module, making installation simple for academic researchers and sponsors. Additionally, RNAseq pipelines can be shared precisely as UNIX-like command arguments or configuration files. 3) Flexible: software tools can be swapped easily, and steps can be added simply via configure files or as command arguments. Thus, the Nextflow RNAseq bioinformatic pipeline can be customized to regulators' latest data analysis framework recommendations. 4) Checkpointed: snapshots of application states are periodically saved throughout the workflow so that applications can restart

from snapshots in case of failure. This is especially useful for pipelines that can take long to run and if systems are error-prone.

The following steps were followed during the upstream processing of the RNAseq data. A visual summary of the data analysis framework is provided in Figure 5.



(title on next page)

Figure 5: This image and figure title are adapted from the R-ODAF publication (Verheijen *et al.*, 2022). This image depicts a standard RNAseq dose-response framework incorporating a modified R-ODAF bioinformatics pipeline. The orange and blue boxes represent stages in the pipeline. The orange boxes specifically represent quality control stages in the pipeline. The green boxes represent options and software tools. The yellow boxes represent filtering criteria. NTP=National Toxicology Program, R-ODAF=Regulatory 'omics data analysis framework, Q30=bases with a Phred quality score about 30, CPM=Counts per million, FDR=False discovery rate, SD=Standard deviation, BMR=Benchmark response, BMD=Benchmark dose, BMD_U=Upper confidence level of the BMD, BMD_L=Lower confidence level of the BMD, GSEA=Gene set enrichment analysis

First, FastQ read file strandedness was defined. Several library preparation methods result in sequencing reads with different characteristics. For example, reads can be single-end or paired-end and be unstranded or strand-specific. Each library preparation method can also have different fragment orientation characteristics. For example, it is possible for strand-specific paired-end fragment reads to face inward toward each other, where read 1 comes from the reverse strand and read 2 comes from the forward strand (Patro *et al.*, 2017). The sample libraries were of the Inward Stranded Reverse library type meaning that libraries were prepared in a stranded paired-end protocol where the reads face inward towards each other, read1 comes from the reverse strand, and read2 comes from the forward strand.

Second, Trim Galore (a wrapper tool around Cutadapt and FastQC) was used to remove nucleotides with a Phred quality score of less than 20, remove Illumina Nextera adapter sequences ('CTGTCTCTTATA') from the 3' end of reads and remove paired reads less than 20bp in length (Kreuger, 2023). Per the R-ODAF (Regulatory 'Omics Data Analysis Framework) recommendations, samples with an average Phred quality score of less than 30 were removed (Figure 5) (Verheijen *et al.*, 2022).

Third, SortMeRNA was used to quantify the proportion and percentage of reads mapping to rRNA from different domains of life (i.e., archaea, bacteria and eukaryotes) (Kopylova *et al.*, 2012).

Fourth, a reference GRCz11 decoy-aware (decoy sequences are regions of the genome with sequences similar to annotated transcripts) transcriptome index was built via Salmon using the entire GRCz11 genome for accurate quantification estimates (Patro *et al.*, 2017). Decoy-aware transcriptome indexes help to prevent unannotated genomic loci bearing sequence

similarity to annotated transcripts from falsely mapping to the annotated transcript (Srivastava *et al.*, 2019).

Finally, paired reads were pseudo-aligned to the transcriptome index using Salmon's quasi-mapping mode and quantified (Patro *et al.*, 2017). As recommended in the regulatory-focused R-ODAF workflow, samples with less than a 70% mapping rate were removed from the analysis (Figure 5) (Verheijen *et al.*, 2022). After quantifying each transcript, the quantified transcripts were transformed to gene level counts data. One of the major benefits of using Salmon for quantification is that it offers a parameter, `--gcBias`, which corrects for fragment-level GC biases in the reads data. Additionally, downstream in the analysis, DESeq2 offers normalization that takes into account the average transcript length when scaling the counts with `DESeq2::counts(dds, normalized=TRUE)` with gene counts data quantified via Salmon with the `--gcBias` parameter enabled (Soneson *et al.*, 2015; Love *et al.*, 2014). Salmon's quasi-mapping mode, rather than performing base-to-base alignments, maps read fragments to potential transcripts based on loci alignment scores and a quasi-mapping algorithm. The quasi-mapping algorithm uses suffix arrays originating from the transcriptome and hash tables mapping each k-mer occurring in the transcriptome to its suffix array interval (Srivastava et al., 2016, 2019). An explanation of how quasi-mapping functions is beyond the scope of this thesis. Please see Salmon's online documentation for more information on how the quasi-mapping algorithm, selective alignment, and quantification steps function.

One common pitfall of RNAseq analyses is contamination from rRNA because of its abundance in RNA samples. As mentioned previously, rRNA was removed as part of the RNAseq library preparation (PolyA enrichment of the mRNA minimizes the presence of rRNA). However, rRNA is so abundant that some can remain in RNAseq samples even after removal.

Thus, reads mapping to rRNA at the gene level were removed before summarising quality control metrics and normalizing the gene level count data. Ribosomal RNA was not removed at the read level in the upstream pipeline because it is much quicker to remove ribosomal RNA at the gene level using this methodology. Downstream quality control was done using summary statistics. Briefly, the total read count, number of genes detected, nCov5 (the number of genes with at least 5 mapped transcripts) and nSig80 (the number of highly expressed genes comprising 80% of all quantified transcripts) were summarised to assess sequencing depth and quality (Harrill *et al*., 2021).

The R-ODAF post-processing regulatory framework was followed to strengthen the validity of the normalized gene counts data and final DEG list for each chemical (Verheijen *et al.*, 2022). Briefly, principal component analysis (PCA) plots and sample-to-sample comparison heatmaps were generated from clustering by Euclidean distances between DESeq2 regularized logarithm (rlog) values for each sample via the nf-core/rnaseq deseq2_qc.r script (https://github.com/nf-core/rnaseq/blob/master/bin/deseq2_qc.r). Using a PCA plot and the clustering heatmaps, outlier samples were removed visually, as shown in Supplementary Figures 32 and 33. A maximum of two outliers per dose group could be removed for a chemical to remain valid for analysis. Finally, the gene-level counts data were normalized using the DESeq2 R package (Love *et al.*, 2014). DESeq2's normalization process includes a size factor correction which facilitates comparisons across samples which may have been sequenced at different depths (Anders and Huber, 2010). Between Salmon's `--gcBias` correction mentioned earlier and DESeq2's size factor correction, gene counts were normalized for both sequencing depth across samples and transcript length (Soneson *et al.*, 2015; Love *et al.*, 2014; Anders and Huber, 2010).

Next, the R-ODAF 'relevance filter' was applied. Only genes consistently expressed in more than one dose group and 75% of replicates with a counts-per-million (CPM) value of 1 or more were filtered for further analysis (Figure 5). The normalized and 'relevance' filtered gene counts data were then used in the downstream dose-response analysis. Before the dose-response analysis, the R-ODAF differential expression analysis was done to ensure that each experiment yielded differentially expressed genes from chemical exposure.

Differential Expression Analysis

The source code for R-ODAF's DEG analysis can be found on GitHub (https://github.com/R-ODAF/Main/blob/main/scripts/R_ODAF_DEGs.R). It should be noted, however, that R-ODAF's DEG analysis is not designed for dose-response analyses. The R-ODAF DEG analysis does, however, facilitate regulatory-compliant biological interpretations because of how strict the filtering criteria are. Therefore, the primary standard dose-response analysis used in this chapter is described in the next section, 'Transcriptomic Dose-response Modelling'. The DEG analysis in this section is strictly complementary and is not used to derive tPODs.

Following normalization and the 'relevance' filter, the DESeq2 R package was used to identify DEGs via the `DESeq2::DESeq()` function, which applies negative binomial generalized linear model (GLM) fitting and the Wald test for the GLM coefficients (Love *et al.*, 2014). The model fit p-values were adjusted for multiple testing via the FDR adjustment. Genes with adjusted p-values greater than 0.01 were removed per the R-ODAF. Furthermore, a spurious spike in the expression filter was applied (Figure 5).

(Verheijen *et al.*, 2022)

$$Spurious\ Spike\ Threshold\ = 1.4\ \times (\#\ of\ biological\ replicates)^{-0.66} \quad (4)$$

If any samples exceeded the spurious spike threshold for a gene, the gene was removed. This function excludes genes with replicates representing more than 66% of the total read counts (with triplicate biological replicates). Spurious spikes in expression are defined as a single sample representing the majority of reads of a dose for a gene (Verheijen *et al.*, 2022). Finally, a 3rd quartile rule filter was applied (Figure 5). It is common for DEG analyses to apply a fold-change filter. However, a fold-change filter can lead to excluding genes that are differentially expressed but have high expression levels. To replace the fold-change filter, the 3rd quartile rule was used to remove genes from the gene list for which the median of a dose group is not higher than the 3rd quartile of the control group (Verheijen *et al.*, 2022). The list of unique genes from all dose group comparisons remaining after the before-mentioned filters were considered DEGs responsive to treatment (i.e., all dose groups were treated as individual treatment groups and were compared to the control group. Each comparison yielded lists of DEGs. The DEGs were summarized into one list of unique DEGs).

Transcriptomic dose-response modelling

The National Toxicology Program (NTP) Approach to Genomic Dose-Response Modelling guideline was followed for the remainder of the transcriptomic dose-response analysis, including the gene level and gene set level analyses (National Toxicology Program, 2018). This standardized guideline from the NTP was followed to ensure the reproducibility of the results and maximize the likelihood of regulatory acceptance of the methods. The NTP approach to 'omic dose-response modelling is specifically tailored to derive tPODs for dose-

response analyses, unlike the R-ODAF. The normalized and 'relevance' filtered gene counts data were imported to BMDExpress 3.0 for the dose-response analysis (Phillips, 2022). BMDExpress 3.0 is a dose-response modelling program that takes gene level counts data as input and outputs a benchmark dose (BMD) for each treatment-responsive DEG identified by William's trend test with a fold-change filter. In BMDExpress 3.0, the normalized and 'relevance' filtered gene level counts were filtered further using William's trend test ($\alpha = 0.05$) and an effect size filter (fold change of ≥ 1.5) to isolate only DEGs responding to the chemical exposure treatment monotonically. A BMD analysis was done for each DEG with a benchmark response (BMR) threshold of 1 standard deviation (SD) from the control group. The BMR of 1 SD is consistent with the United States Environmental Protection Agency (US EPA) recommendations and the standardized guidelines for genomic dose-response modelling as outlined by the NTP (National Toxicology Program, 2018; Hogan *et al.*, 2012). The 1 SD level is appropriate for normally distributed data when the biologically significant threshold level is unknown and is theoretically concurrent with the 10% effect level (National Toxicology Program, 2018). Several models were fitted to the dose-response data, including a classical Hill model, Power model, linear model, 2^{nd}-order polynomial model (quadratic), 3^{rd}-order exponential model and 5^{th}-order exponential model (Supplementary Figure 39). Parametric models are used over non-parametric models because of their simplicity and computational efficiency, especially considering there can be thousands of DEGs for every exposure experiment (n = 10 experiments) (National Toxicology Program, 2018). Many different models were used to maximize the likelihood that a gene's dose-response data would fit a model(s) well enough to be considered further for analysis (National Toxicology Program, 2018). The best model fit was determined as the model with the lowest Akaike information criteria (AIC) value. AIC estimates each model's fitness by considering its

complexity and how well it fits the data. The complexity of the model is dictated by the number of independent variables used to build the model - excess complexity is penalized to avoid overfitting. The fitness of the model is determined by the maximum likelihood estimate of the model. Finally, the potency value derived from the best-fit model for each gene is the BMD. The BMD is the estimated dose that produces a predetermined change in the response rate of a biological response (the BMR) (National Toxicology Program, 2018). Before calculating tPODs, the best-fit gene models were filtered based on various criteria: 1) Ensembl gene IDs were translated to Entrez gene IDs, and duplicates were removed, 2) the best-fit gene models with BMDs greater than the highest dose were removed, 3) the best-fit gene models with a BMD/BMD_L ratio greater than 20 were removed, 4) the best-fit gene models with a BMD_U/BMD_L ratio greater than 40 were removed, 5) the best-fit gene models with a BMD_U/BMD ratio greater than 20 were removed, 6) the best-fit gene models with a BMD value less than 10-fold the lowest dose were removed and 7) the best-fit gene models with a fit p-value < 0.1 determined by a likelihood ratio test were removed (Figure 5) (National Toxicology Program, 2018).

Distribution-based gene level tPOD

The distributions of gene level BMDs induced by each chemical were plotted using a histogram in R. Several methods were used to calculate distribution-based tPODs for each chemical: 1) the first mode tPOD, 2) the 10^{th} percentile tPOD and 3) the 20^{th} gene tPOD. The 10^{th} percentile tPOD is the 10^{th} percentile gene BMD in the distribution, the 20^{th} gene tPOD is the 20^{th} gene BMD in the distribution, and the first mode tPOD is the dose at which the first mode occurs in the distribution of the BMDs. The first mode is determined in R by a simple mode-finding algorithm that takes a set of BMD values, the minimum density threshold (i.e., the

minimum mode size), the bandwidth smoothing function (i.e., the Sheather and Jones bandwidth method) and the minimum bandwidth size (i.e., bin size) as input and outputs the first mode. A bootstrap was conducted to calculate lower and upper confidence intervals for the distribution-based tPODs. Briefly, the BMD values were sampled 2000 times with replacement, and the 50^{th} (median), 2.5^{th} and 97.5^{th} percentile values were calculated. The median 50^{th} percentile was the distribution-based gene level tPOD, the 2.5^{th} percentile was the lower confidence interval of the distribution-based gene level tPOD, and the 97.5^{th} percentile was the upper confidence interval of the distribution-based gene level tPOD.

Classification-based gene set level tPOD

Classification-based gene set level tPODs were also determined based on the enrichment of gene sets according to two classifications: 1) REACTOME biological pathways and 2) gene ontology (GO) biological function terms. For the Gene set enrichment analysis (GSEA), only enriched gene sets that were minimally 5% populated (based on the total number of annotated genes in the gene set) with at least 3 genes were selected. A Fisher's exact two-tailed test ($\alpha = 0.05$) was conducted to filter gene sets with expression levels significantly different from the background expression levels (Figure 5) (National Toxicology Program, 2018). The median BMD of enriched gene sets was plotted in an accumulation plot using R. The classification-based gene set level tPOD was defined as the median BMD of the first enriched gene set. A bootstrap was conducted to calculate the lower and upper confidence intervals of the gene set level tPODs. Briefly, the BMD values of the first gene set (minimum 3 BMDs) were sampled 2000 times with replacement. The median 50^{th} percentile was the classification-based gene set level tPOD, the 2.5^{th} percentile was the lower confidence interval of the classification-based gene set level tPOD,

and the 97.5th percentile was the upper confidence interval of the classification-based gene set level tPOD.

POD Comparisons

The tPODs were compared to one another to assess the sensitivity of each method. The most sensitive tPOD from each chemical group was then used to compare PODs to assess the protectiveness and sensitivity of the supplemented ZFET test. Overall, the ZFET test was supplemented in Chapter 2 with a swimming behaviour assay, an oxidative metabolism-dependent energy expenditure assay and an overt toxicity (death and deformity) assay. Finally, this chapter used a dose-response transcriptomic gene-expression assay to supplement the ZFET test. To summarize the results of both Chapter 2 and this chapter (Chapter 3), a summary figure was made to compare the experimental PODs and PODs from the literature.

3.4. Results

Total RNA sample and cDNA library quality control

In-house RNA extractions conducted with Qiagen RNEasy mini kits yielded an average of 2520ng of total RNA (84ng/µl RNA in 30µl RNase-free water) as measured by the Qiagen QIAxpert Instrument. The average RIN of RNA samples, measured by the Qiagen QIAxcel instrument, was 6.7 (min 4.4 & max 8.7) (Supplementary Information - RNA_Extractions).

Genome Quebec quality checked each RNA sample to ensure good sample quality upon delivery. 3µl of each sample was used for quality control. On average, each sample contained 478ng of total RNA (approximately 40ng/µl RNA in 12µl RNase-free water) with an average RIN score of 8.7 (min 5.9 & max 10) before library preparation (Supplementary Figure 25).

However, after cDNA library preparation using Illumina mRNA stranded kits, 3 samples (Aniline dose 0.01mg/L replicate C, Aniline dose 1mg/L replicate B and Glyphosate dose 1mg/L replicate B) failed quality control screening with the Bioanalyzer and were thus omitted from sequencing (libraries concentrations were too low for sequencing - <3nM).

It should be noted that RNA samples from the TGSH and EE2 chemical exposures (n=36) underwent three freeze-thaw cycles, and the remaining RNA samples were only freeze-thawed once before RNA sequencing. The distribution of RIN scores for each chemical exposure experiment is shown in Supplementary Figure 25. The RNA samples from the TGSH and EE2 chemical exposure experiments were freeze-thawed three times because they were used for two previous experiments, which are not included in this thesis.

RNAseq quality control

On average, samples had 17.1M reads. The maximum number of reads was 23.1M, and the minimum number of reads was 12.7M. No samples had a read count less than 5M (R-ODAF's cutoff threshold) (Supplementary Figure 34).

On average, 2.4% of the bp were removed from reads during trimming (Supplementary Figure 29). No samples had an average Phred quality score of less than 30. The average Phred quality score by sample (n = 360 samples) was 36, and the average percentage of reads above the Phred quality score of 30 was 98.1% (Supplementary Figure 31). Thus, no samples were removed for having an average Phred quality score of less than 30.

SortMeRNA was used to quantify rRNA contamination. The percentage of overall rRNA contamination in samples grouped by chemical is shown in Supplementary Figure 27. The proportion and percentage of reads mapping to rRNA broken down by life domain (i.e., archaea, bacteria and eukaryotes) is shown in Supplementary Figure 28. Since a PolyA enrichment of

mRNA was included in the library preparation protocol, rRNA contamination levels were low. On average, 95.6% of rRNA was removed. Under the assumption that ~90% of a biological sample is rRNA, and considering, on average, 95.6% of rRNA was removed, it can be expected that ~28.4% of aligned reads will be rRNA molecule fragments and ~71.6% of aligned reads will be of the transcriptome and therefore relevant to the study (Thermo Fisher Scientific, 2022). This value was used as a reference to roughly estimate the efficiency of read mapping. Reads mapping to rRNA mapped to eukaryotic rRNA, on average, 83.8% of the time (Supplementary Figure 28). There were 2 sample outliers with higher than average rRNA contamination (Aniline dose 0 replicate C and Malation dose 0.1 replicate B). TGSH and EE2 samples had the highest rRNA contamination levels (Supplementary Figure 27).

The percentage of read duplication was measured using FastQC and is shown in Supplementary Figure 30. On average, reads were 45% duplicated across all samples with two outlier reads (Aniline dose 0 replicate C read 1 and read 2). Generally, TGSH and EE2 samples had the highest levels of read duplication (Supplementary Figure 30).

Salmon's quasi-mapping mode was used to quantify the number of paired reads mapping to the decoy-aware (decoy sequences are regions of the genome with sequences similar to annotated transcripts) GRCz11 zebrafish transcriptome (Figure 5). The percentage of mapped paired reads is shown in Supplementary Figure 26. On average, 78.6% of all paired reads were mapped to the zebrafish transcriptome. The percentage of mapped paired reads (78.6%) was approximately concurrent with the estimated percentage of aligned reads (71.6%). Thus, there were no issues with the overall alignment methods. The real percentage of aligned paired reads was 7% better than estimated, on average. There were 2 sample outliers with poorer mapping results (Aniline dose 0 replicate C and Malation dose 0.1 replicate B). Overall, 13 samples did

not meet the 70% R-ODAF mapping threshold (Figure 5 and Supplementary Figure 26) (Verheijen *et al.*, 2022). 12 of the 13 samples were from the TGSH chemical exposure experiment. The 1 other sample was from Aniline control dose 0mg/L replicate C.

After quantification via Salmon, Nextflow's RNAseq DESeq2 quality control script (https://github.com/nf-core/rnaseq/blob/master/bin/deseq2_qc.r) was used to create PCA plots and clustering heatmaps to identify outliers samples. The PCA plot and an example of the clustering heatmaps are in Supplementary Figures 32 and 33. Two outlier samples (EE2 dose 0.0001mg/L replicate C & Fadrozole dose 1mg/L replicate A) were identified and discarded.

Based on all of the above quality control metrics, a total of 24 samples were removed from the analysis. After removal from the original sample size of n=180, the final sample size was n=156. The removed samples included: all (n = 18) of the TGSH samples due to not meeting the 70% mapping filter; 3 Aniline samples (Aniline control dose 0mg/L replicate C, Aniline dose 0.01mg/L replicate C, and Aniline dose 1mg/L replicate B) from errors in library preparation (the library concentration was below 3nM - too low for sequencing), a high percentage of read duplicates and the 70% mapping filter, 1 Glyphosate sample (Glyphosate dose 1mg/L replicate B) from errors in library preparation (the library concentration was below 3nM - too small for sequencing), 1 EE2 sample (EE2 dose 0.0001mg/L replicate C) from the PCA/sample clustering filter and 1 Fadrozole sample (Fadrozole dose 1mg/L replicate A) from the PCA/sample clustering filter.

Furthermore, gene counts data from the remaining samples were filtered using the R-ODAF 'relevance' filter (Figure 5). The R-ODAF 'relevance' filter removed 3,049 (13.4%) of the genes assessed from all samples, leaving an average of 19,705/22,754 genes for further downstream analysis (R-ODAF_DEG_Report.html - Pipeline output text block).

Differential Expression

The R-ODAF differential expression analysis was conducted via the DESeq2 R package. The number of DEGs filtered can be seen in R-ODAF_DEG_Report.html. After adjusting the `DESeq2::DESeq()` negative binomial GLM model fit p-values with an FDR adjustment (α = 0.01), on average, ~293 DEGs per chemical were identified. However, DEGs were filtered further, with the 3rd quartile rule, removing ~21 DEGs per chemical on average. Finally, DEGs were again filtered further with the spurious spike filter, removing ~1 DEG per chemical on average. The average number of remaining DEGs (including duplicate DEGs from multiple dose group comparisons to the control) across all samples after all filtering was ~271 DEGs per chemical. The number of unique DEGs per chemical is shown in Figure 6, which also shows the dose-response relationship of DEGs for each chemical. The full table of unique DEGs from this analysis and a separate non-standard dose-response analysis can be found in Supplementary Information (R-ODAF_DEG_Report.html). Aniline, Fadrozole, Glyphosate and Trenbolone had relatively weak responses, with less than 100 DEGs across all doses. 1-Decanol, EE2, Malathion and Prochloraz had moderate responses, each having between 100 and 300 DEGs. Fenitrothion had the strongest response with over 900 DEGs. Most of the chemicals showed a clear dose-dependent increase in DEGs, though there were few minor departures from a monotonic response. One exception is that Fenitrothion had an unusually large number of DEGs on one of the lower doses (0.001 mg/L).

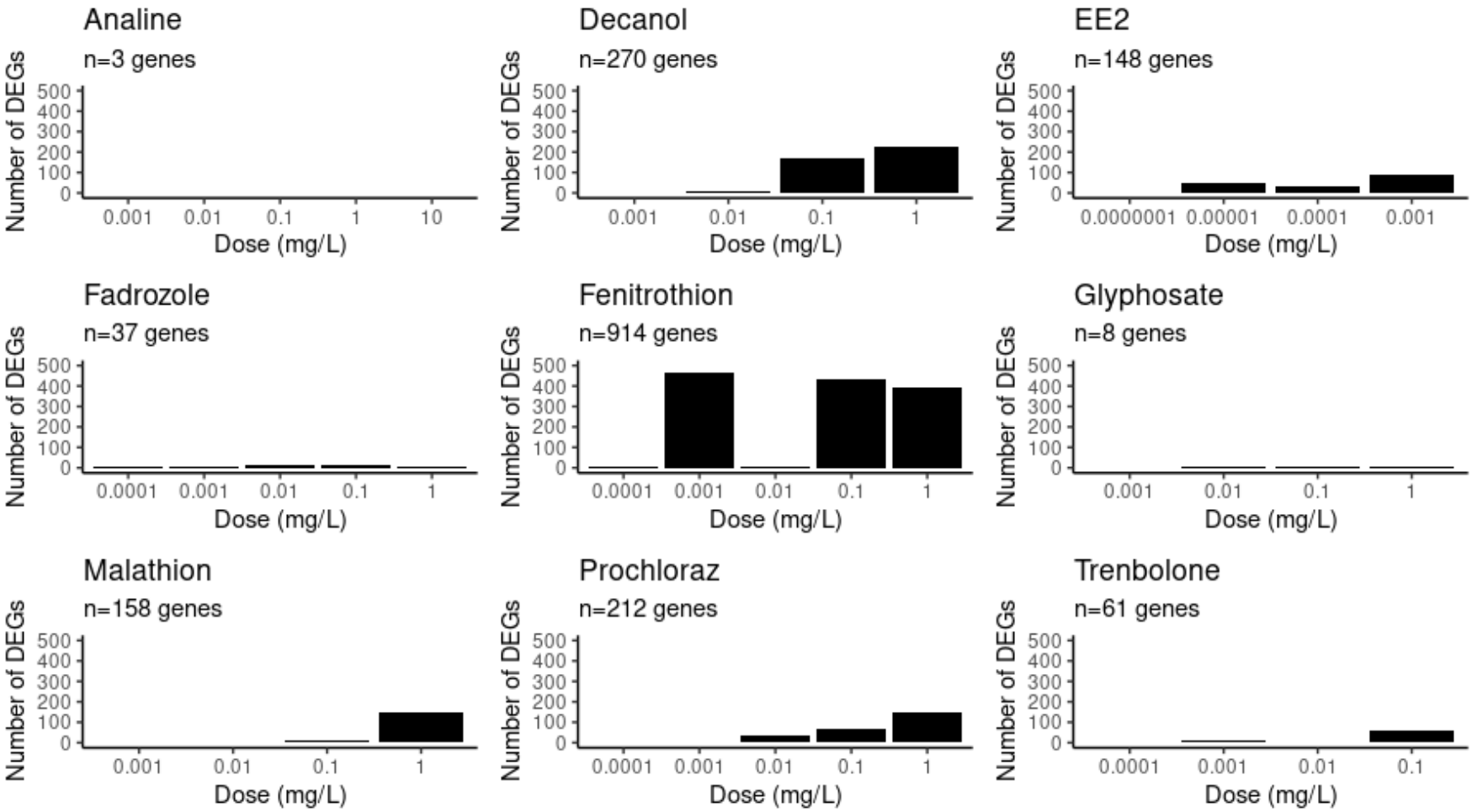
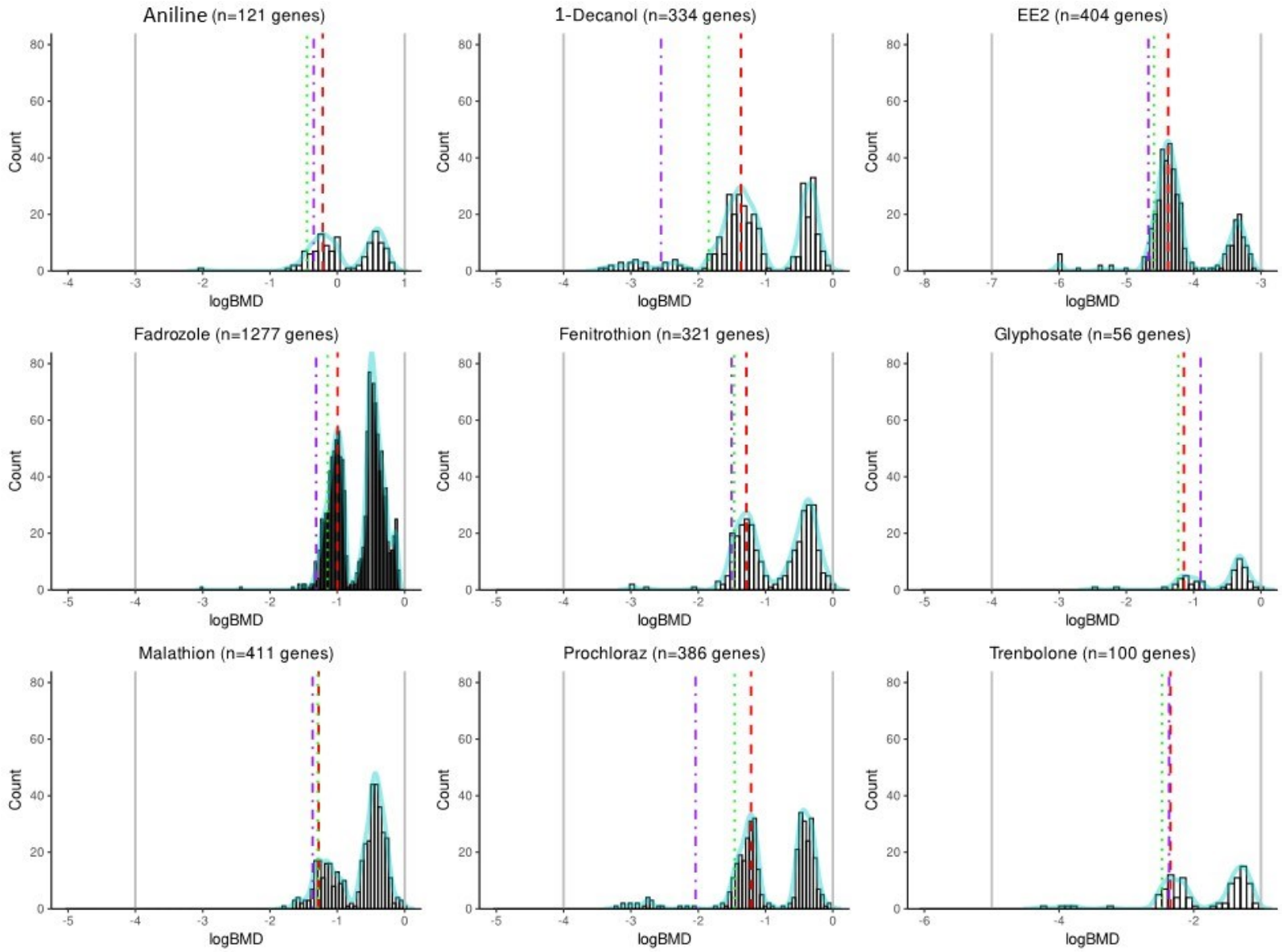


Figure 6: Per the R-ODAF pipeline, the DESeq2 R package was used to identify DEGs in each treatment group from the 5 day zebrafish embryo exposure experiment. This image depicts dose groups on the x axis, the number of DEGs in each treatment group on the y-axis and the number of unique DEGs across all dose groups in the subtitle of each facet. Under the assumptions of a monotonic dose-response, it is expected that the number of DEGs will increase as the dose increases. The results show that responses in gene expression for 1-Decanol, EE2, Malathion, Prochloraz and Trenbolone behave monotonically. Gene expression responses to Fadrozole and Fenitrothion exposure behave non-monotonically. The dose-response relationship of Aniline and Glyphosate are inconclusive due to the few number of DEGs.

Transcriptomic dose-response modelling

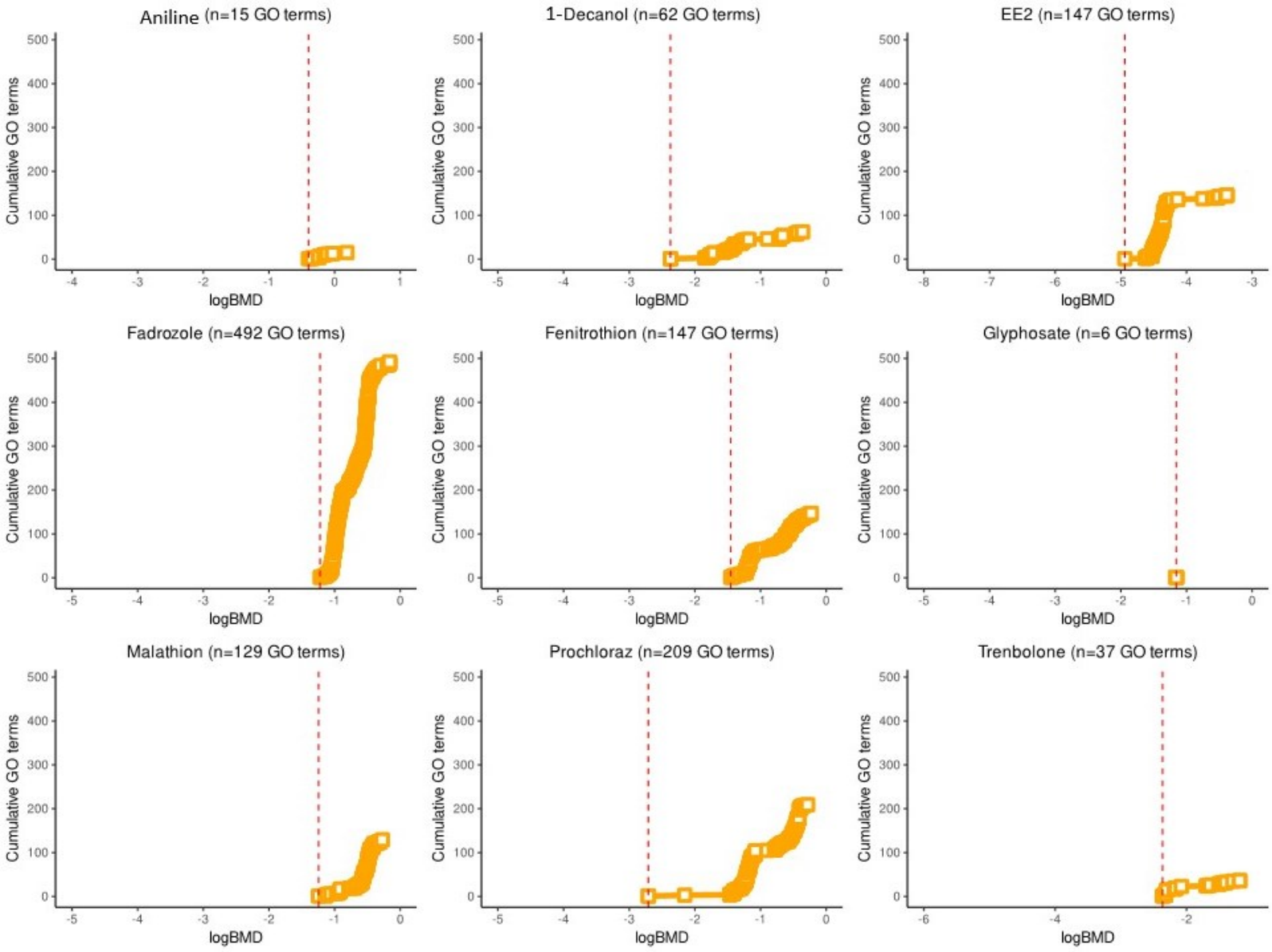
Next, the gene level BMD analysis was conducted via BMDExpress 3.0 using normalized and 'relevance' filtered gene counts data from each chemical and dose group (Figure 5). The number of genes passing the BMD analysis filtering criteria described in Figure 5 is shown in Figure 7 next to the chemical names. In this analysis, only Glyphosate had fewer than 100 dose-responsive genes. All other chemicals had between 100 and 500 dose-responsive genes, except Fadrozole, which had over 1200 dose-responsive genes. The full list of genes and associated BMDs are available in the Supplementary Information (All_Chemicals_BMD_Data_for_Histogram.csv).



(title on the next page)

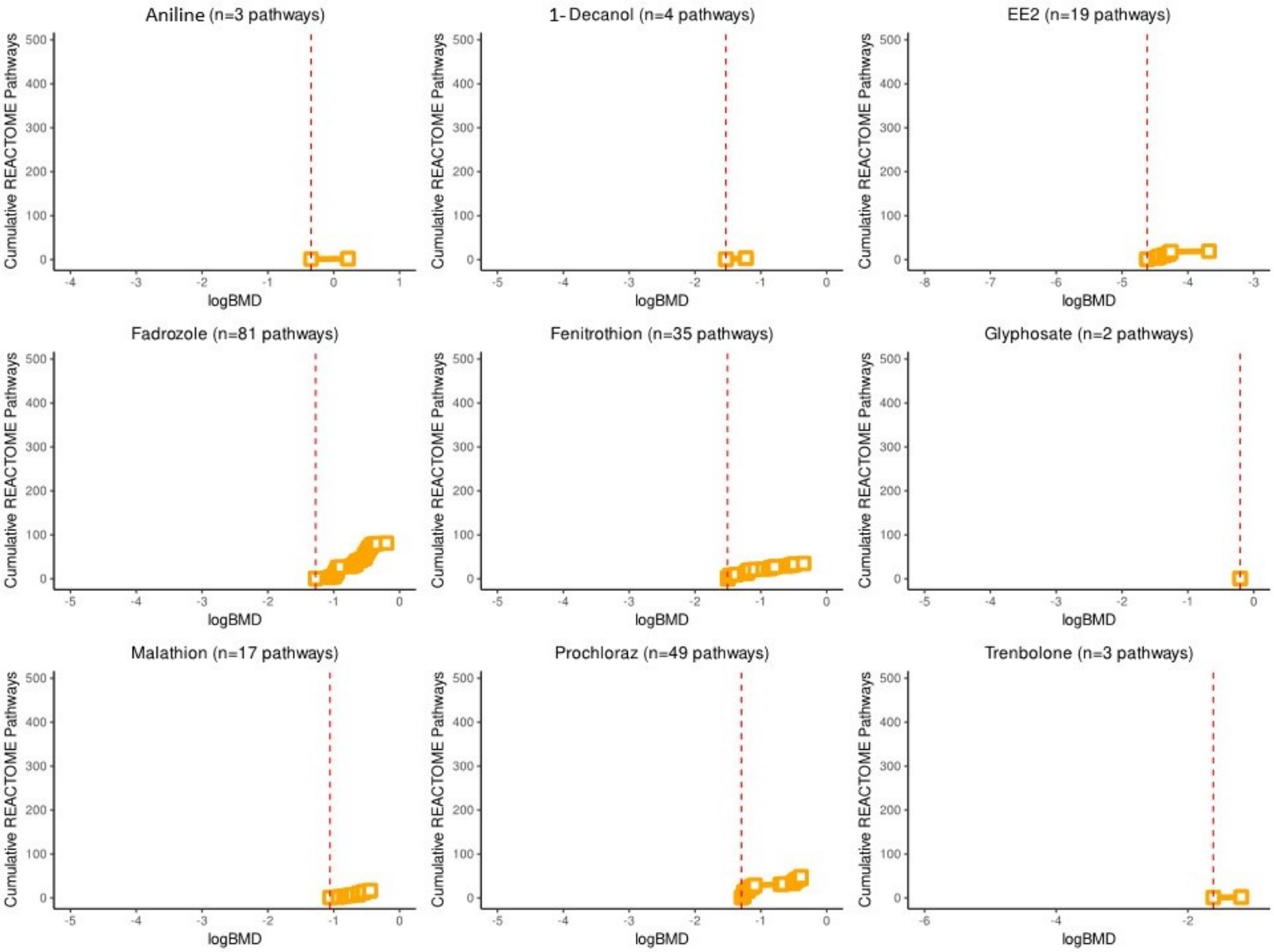
Figure 7: Per the NTP guidelines on 'omic dose-response modelling, BMDExpress 3.0 was used to conduct a BMD analysis of gene expression at the gene level from the 5 day zebrafish exposure experiment. To calculate tPODs, the gene BMDs were distributed along the x-axis. The bandwidth of the density smoothing function is determined by the Sheather and Jones bandwidth function. The Sheather and Jones bandwidth function is based on the concept of the mean integrated squared error to measure the quality of the density estimation. The bin widths of the histograms are equal to the bandwidth. The first mode tPOD, the red dashed vertical line, is the first mode/peak of the distribution of gene BMDs. It is representative of the most potent change in gene expression. The green vertical dotted line is the 10^{th} percentile tPOD. The purple dotted-and-dashed vertical line is the 20^{th} gene tPOD. The 10th percentile and 20th gene tPOD are typically more sensitive than the first mode tPOD, thus, they are useful for potential hazard assessment. However,the first mode tPOD often infers an underlying mechanism of action which is useful for biological interpretations of the results.

Figure 8 and Figure 9 show the accumulation of enriched GO term BMDs and REACTOME pathway BMDs, respectively. The vertical dashed red line also shows the corresponding tPODs for each chemical in each figure. A list of all the enriched GO terms and REACTOME pathways can be found in the Supplementary Information (All_Chemicals_GO_Data.csv and All_Chemical_Reactome_Data.csv, respectively).



(title on next page)

Figure 8: Per the NTP guidelines on 'omic dose-response modelling, BMDExpress 3.0 was used to conduct a BMD analysis of gene expression at the gene set level from the 5 day zebrafish exposure experiment. To calculate the gene set level tPODs, GO terms (from the Gene Ontology database) were enriched with genes that possessed BMDs. The median BMD of each enriched GO term was plotted along the x-axis of the accumulation plot. The increasing y-axis shows the total number of enriched GO terms. The median BMD of the most potent enriched GO term was determined as the GO term tPOD.



(title on next page)

Figure 9: Per the NTP guidelines on 'omic dose-response modelling, BMDExpress 3.0 was used to conduct a BMD analysis of gene expression at the gene set level from the 5 day zebrafish exposure experiment. To calculate the gene set level tPODs, REACTOME pathways (from the REACTOME pathway database) were enriched with genes that possessed BMDs. The median BMD of each enriched REACTOME pathway was plotted along the x-axis of the accumulation plot. The increasing y-axis shows the total number of enriched REACTOME pathways. The median BMD of the most potent enriched REACTOME pathway was determined as the REACTOME pathway tPOD.

POD Comparisons

All of the gene level and gene set level tPODs were compared to one another (Figure 10). Generally, most of the tPODs within a chemical were within 10-fold of each other. However, they varied in stability (as indicated by their 95% confidence intervals). Overall the 10^{th} percentile and 20^{th} gene tPODs had the lowest variability, followed by the first mode. The gene set-based tPODs (REACTOME and GO Term) were the most variable.

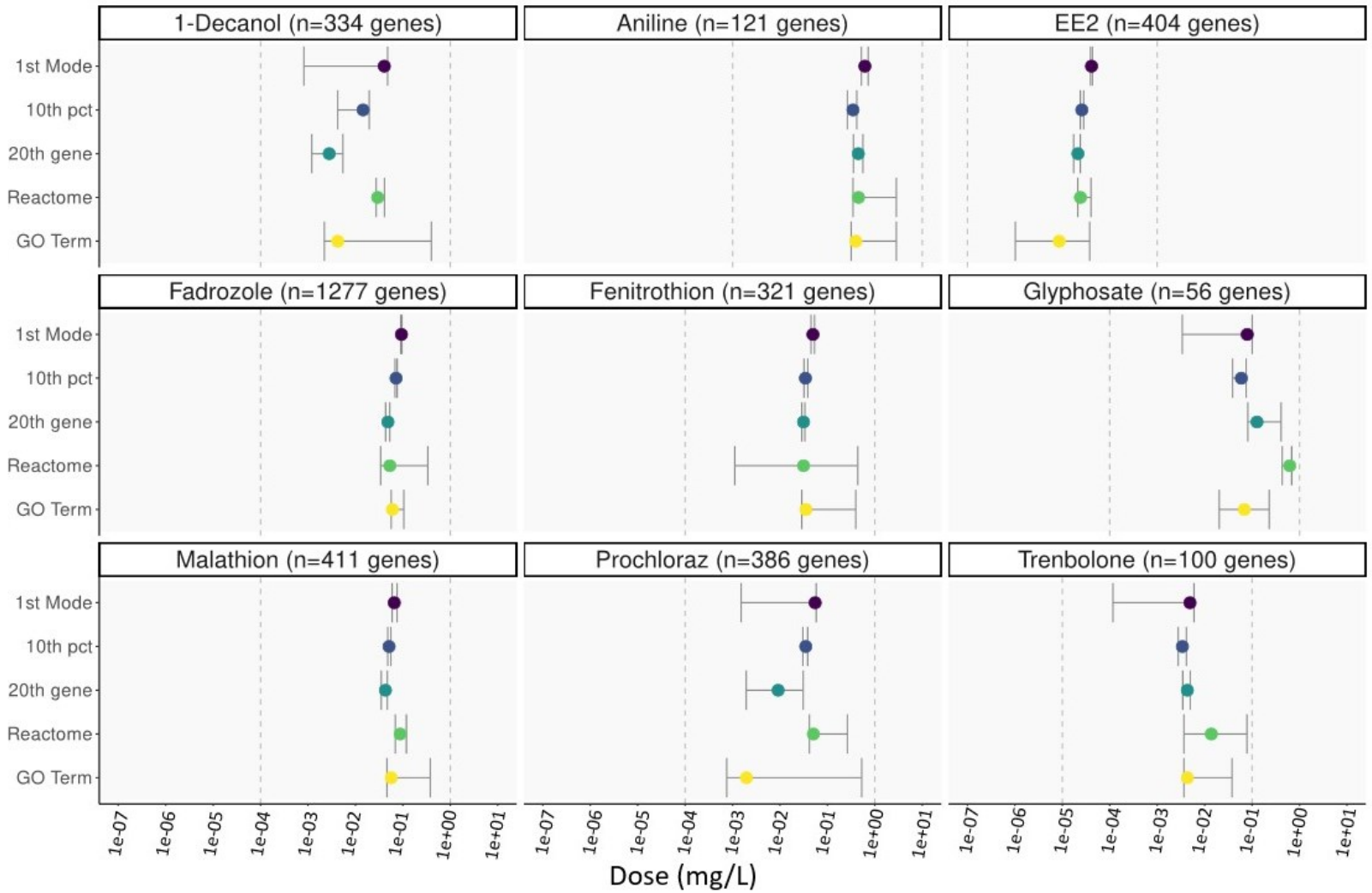
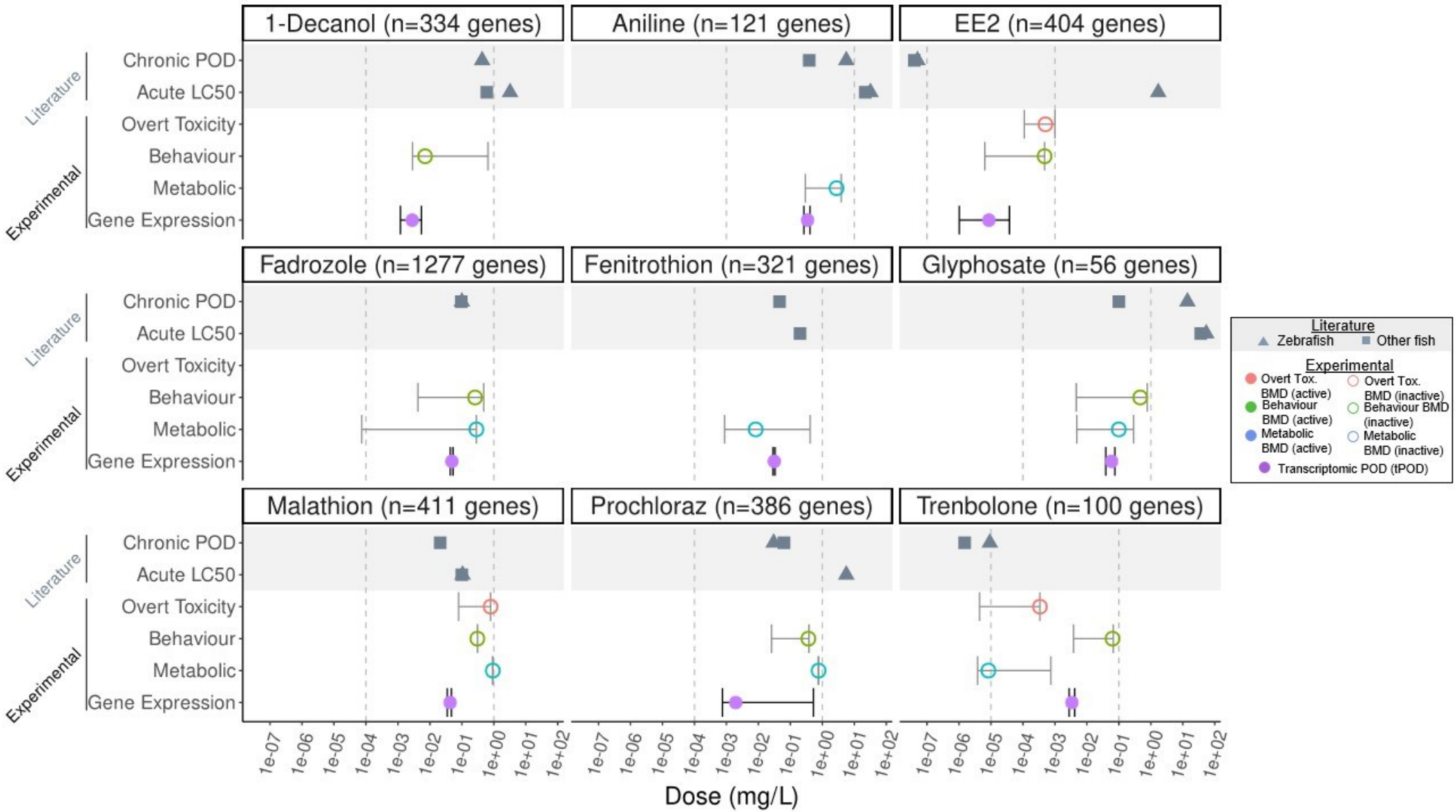


Figure 10: The different tPODs derived from the gene level and gene set level analyses are compared in this figure. The number of DEGs passing all filtering criteria are shown in each title facet along with the respective chemical exposed to the 5 zebrafish embryos over 5 days (5dpf). A bootstrap (n = 2000 times) was conducted to determine the 2.5% lower and 97.5% upper confidence intervals (CIs) of each of the tPODs. These CIs are shown as error bars around the median bootstrapped tPOD. Vertical grey dotted lines represent the dose limits of the experiments.

Finally, the RNAseq-based tPODs were compared to PODs from Chapter 2 (Figure 11), related to swimming behaviour, metabolism-dependent energy expenditure and overt toxicity, and PODs reported for juvenile and adult fish from the literature. This figure compares only the most sensitive POD from each assay. The statistical LOECs were not included in the POD comparisons to simplify the comparison. Generally, the lowest tPOD was the lowest of all experimental PODs, including those from Chapter 2 (it is the lowest POD for the 1-Decanol, Aniline, EE2, Fadrozole, Glyphosate, Malathion, and Prochloraz experiments). The lowest tPOD was not the lowest experimental POD in two cases, the Fenitrothion and Trenbolone experiments. In both exceptional cases, the inactive oxidative metabolism energy expenditure POD was the lowest. Generally, the lowest tPOD was lower or within 10-fold of the chronic overt toxicity juvenile/adult fish POD from the literature (it is lower than or within 10-fold of the chronic overt toxicity juvenile/adult literature POD for the 1-Decanol, Aniline, Fadrozole, Fenitrothion, Glyphosate, Malathion and Prochloraz experiments). The two exceptional cases were the EE2 and Trenbolone experiments, where the chronic overt toxicity of juvenile/adult POD was lower than the lowest tested exposure concentration. Thelowest tPOD was always lower than the acute overt toxicity juvenile/adult fish POD from the literature (except for the Trenbolone experiment, for which acute overt toxicity PODs for juvenile/adult fish were unavailable in the literature).



(title on next page)

Figure 11: The experimental PODs and the literature-based PODs are summarised and compared in this figure. Only the most sensitive PODs and associated CIs are shown. The number of DEGs after exposing zebrafish embryos for 5 days (5dpf) are shown in the facet titles along with the associated chemical. PODs from Chapter 2 are distinguished as inactive or active depending on the certainty factor after bootstrapping the samples (n = 1000 times). The most sensitive PODs from the literature review of fish toxicity tests in Chapter 2 are shown in the gray panel. Experimental PODs more potent than the literature-based PODs are considered protective.

3.5. Discussion

Transcriptomic endpoints increase the sensitivity of the ZFET test

Compared to every other endpoint included in the supplemented ZFET test in Chapter 2, transcript-level gene expression was the most sensitive endpoint 100% of the time, excluding the TGSH experiment (Figure 11). As a reminder, the hollow circles in Figure 11 represent BMDs classified as inactive in Chapter 2. There were no active overt, swimming behaviour or oxidative metabolism-dependent energy expenditure toxicity PODs from the 9 chemicals shown in Figure 11. Thus, since there were no active PODs from overt, behaviour or metabolism-dependent energy expenditure toxicity, but there were tPODs calculated for every chemical, the evidence in Figure 11 supports the hypothesis that supplementing the ZFET test with additional endpoints, especially transcript-level gene expression endpoints, improves the sensitivity of the ZFET test greatly.

When comparing the tPODs at the gene level and the gene set level, the gene level tPODs were lower than the gene set level tPODs most of the time. This was observed 7 out of 9 times. Specifically, the lowest gene level tPODs were the 10th percentile (Aniline, Glyphosate, and Trenbolone) and 20th gene tPODs (Decanol, Fadrozole, Fenitrothion, Malathion). The second lowest tPOD was the first enriched GO term median BMD tPOD (EE2 and Prochloraz).

Generally, all BMD distributions were multimodal at the gene level, as expected (Pagé-Larivière *et al.*, 2019). In previous experiments and analyses conducted in the O'Brien lab, the first mode tPOD performed best when the number of BMDs in the distribution was above ~50 when there were five dose groups (excluding the control group) (Nguyen, 2022). In this experiment, the first mode tPOD algorithm identifies the first mode in the multimodal

distribution of each chemical experiment despite the Glyphosate experiment having only 56 genes passing all filtering criteria (Figure 7). None of the exposure experiments had too few DEGs to calculate gene level tPODs.

However, the minimum density threshold parameter of the mode-finding algorithm is sometimes sensitive to very low-dose minute gene expression effects, hence the error bars extending to the lower doses for the Prochloraz and 1-Decanol chemical experiments in Figure 10. As mentioned in the previous paragraph, the first-mode finding algorithm does not perform well when the overall BMD count is low, hence the error bars extending to the lower doses for the Glyphosate and Trenbolone experiments in Figure 10. Thus, the first mode shows instability despite having a meaningful biological interpretation. On the other hand, the 20th gene tPOD showed stability in Figure 10. However, the biological meaning of the 20^{th} gene tPOD is muddy. Finally, the 10^{th} percentile tPOD is the last tPOD considered for this experiment. The 10^{th} percentile was stable in this experiment. However, percentile-based tPODs are flawed because they drift as the number of doses in the experiment increases (drift to higher concentrations with more doses) and decreases (drift to lower concentrations with fewer doses).

Recently, a novel gene level tPOD derivation method, the Lowest Consistent Response Dose (LCRD), was developed. The LCRD is the lowest BMD of a consistent response group of BMDs (Reardon *et al.*, 2023; Crizer *et al.*, 2021). The consistent response group of BMDs is a rank-ordered list of BMDs where all subsequent ratio values from adjacently ranked BMCs are within 1.66 (Reardon *et al.*, 2023). The consistent response group of BMDs is similar to the concept of distinct modes in distributions of BMDs (Figure 7). Conceptually, the first mode tPOD is the peak, or centre, of the first consistent response group of BMDs, whereas the LCRD is the inflection point on the left side of the first mode. However, unlike the first mode tPOD,

BMDs are rank-ordered (similar to the accumulation plots in Figure 8 and Figure 9). The LCRD is not dependent on the multi-modal distribution and appears to be a sensitive, stable, reproducible tPOD. I am excited to incorporate the LCRD in the future.

Despite minor shortcomings in the data (discussed in the "Importance of quality control and recommendations for the future" section), the results show that tPODs are consistently the most sensitive experimental PODs (Figure 11). The findings indicate that this thesis's transcriptomic dose-response analysis framework is robust (Figure 5). Thus, supplementing the ZFET test with a transcript-level gene expression endpoint increases the sensitivity of the ZFET test and provides an animal-alternative method of deriving tPODs to assess the potential hazards of chemical exposure.

Differential gene expression is the most informative endpoint

Compared to the standard endpoint of the original ZFET test, overt toxicity, the transcript-level gene expression endpoint is far more informative. Whereas the overt toxicity endpoint can reveal information about the toxic mechanisms of action (e.g., perturbs somite formation, perturbs essential organ and system development such as the circulatory system, nervous system or muscles, etc.), transcript-level gene expression data can reveal intricate details about mechanistic perturbations in sensitive pathways and biological functions that have no observable effects after acute exposure but may be observable after chronic exposure (Pagé-Larivière *et al.*, 2019).

Consider the EE2 exposure experiment and refer to the supplementary information files 'R-ODAF_DEG_Report.html' and 'All_Chemical_Reactome_Data.csv' and 'All_Chemical_GO_Data.csv' for information about enriched gene sets and associated BMDs. Regardless of the chosen analysis method (BMD analysis or R-ODAF DEG analysis), the same

sensitive REACTOME pathways were enriched at the gene set level. The consistently enriched REACTOME pathways were visual transduction (10.6% enriched), sensory perception (7.0% enriched) and the canonical retinoid cycle in rods - twilight vision (38.5% enriched). The three pathways were enriched at approximately the same median BMD level (~2.4×10^{-5}mg/L). These gene set level effects were largely driven by 3 DEGs at the gene level (retinoid isomerohydrolase a - rpe65a, retinol dehydrogenase 5 - rdh5 and retinaldehyde binding protein 1b - rlbp1b). The gene BMDs are located on the left tail of the first mode peak in Supplementary Figure 42, showing the distribution of the differentially expressed gene BMDs. EE2 is known to affect the development of the zebrafish nervous system and the sensory system, as well as larval swimming behaviour (Nasri *et al.*, 2021). These results indicate a potential mechanism of action of EE2 exposure to zebrafish larvae: perturbation of the retinoid (visual) cycle (Kiser *et al.*, 2014). This mechanism of action is also active in other endocrine-disrupting chemicals, such as BPA (Martínez *et al*., 2018). Additionally, enrichment of GO biological function terms revealed that the lowest median BMD corresponded to the retinoid metabolic process. These results are corroborated by the swimming behaviour assay performed in Chapter 2, specifically in the dark cycles (Supplementary Figure 24 & Figure 2). Exposure to EE2 resulted in statistically significant increases in the total swimming distance of zebrafish larvae after being exposed for five days during development. The fish in the highest dose group (0.001mg/L) swim further distances than the control group, especially in the dark phase. However, in Chapter 2, the statistical LOECs were too spurious to be reliable, and the BMR was too high to detect biologically significant changes in response. The transcript-level gene expression results suggest that the swimming behaviour effects were biologically significant. Thus, these results confirm that the analysis from Chapter 2 was not sensitive or robust enough to detect these significant

changes in response. The transcript-level gene expression endpoint provides more valuable information than the other endpoints in the supplemented ZFET test. This includes the overt toxicity endpoint, the only endpoint in the standard ZFET test.

An in-depth analysis of enriched gene sets and differential gene expression is possible for every chemical in the data set, excluding TGSH (n = 9). The representative in-depth enriched gene set analysis was conducted for EE2 specifically because there was sufficient evidence to facilitate the analysis, and the experimental results were corroborated by both Chapter 2 and previously published research (Pagé-Larivière *et al.*, 2019; Nasri *et al.*, 2021).

For my biological interpretation of the EE2 exposure experiment, I cross-referenced all DEGs and enriched gene sets from the R-ODAF DEG and BMD analysis methods. The 3 DEGs I identified (rpe65a, rdh5 and rlbp1b) were differentially expressed no matter the analysis method used. Thus I felt confident interpreting the results. However, common DEGs and gene sets enriched across both analysis methods were sparse, especially for the enriched REACTOME pathways (Commonly_Enriched_GO_Terms_Across_Both_Analysis_Methods_NTPBMD_AND_RODAFDEG.csv & Commonly_Enriched_REACTOME_Pathways_Across_Both_Analysis_Methods_NTPBMD_AND_RODAFDEG.csv). Thus, researchers should exercise caution when interpreting these types of biological data because so much information is sensitive to analysis methods, input sample quality, dose range, dose-response relationship shape and many other factors related to RNAseq. These additional factors were mentioned previously, such as rRNA contamination, genomic DNA (gDNA) contamination, inter-species contamination, errors in library preparation, outlier samples, and poor sequencing depth.

It is reasonable to conclude that transcript-level gene expression data is potentially the greatest source of information for chemical hazard and risk assessment. However, it should also be noted that biological interpretations of chemical mechanisms of action based on DEGs are prone to over-interpretation because of false-positive results due to the analysis methods and factors that I mentioned in the previous section of the discussion. Another important point to consider is that as scientists continue to improve their understanding of the relationship between gene expression and biological outcomes, our interpretations of biological findings should also improve over time. This ongoing process requires constant attention and updates as new research is published.

Experimental transcriptomic PODs protect literature-based chronic overt toxicity PODs from juvenile and adult fish

Finally, the chronic and acute overt toxicity PODs from fish experiments in the literature were compared to the experimental PODs. On average, tPODs were 10,000X more sensitive than acute overt toxicity PODs from fish experiments available in the literature. Transcriptomic PODs were protective of acute overt toxicity PODs from fish experiments available in the literature 100% of the time when a literature-based POD was available (excluding Fadrozole and Trenbolone). On average, tPODs were 1X to 10X more sensitive than chronic overt toxicity PODs from fish experiments in the literature (excluding the EE2 and Trenbolone experiments because the dose range was not low enough to capture the chronic overt toxicity fish PODs). Transcriptomic PODs were also protective of chronic overt toxicity PODs from fish experiments in the literature 6 times out of 9. However, the dose ranges of the EE2 and Trenbolone experiments were too high to capture the chronic overt toxicity PODs from the literature.

Excluding EE2 and Trenbolone, experimentally derived tPODs were protective of chronic overt toxicity PODs from fish experiments in the literature 6 times out of 7. For the EE2 experiment, capturing the full range of effects is difficult because the chemical is so potent. Effects of EE2 range from ~1×10^{-7}mg/L to 1mg/L (Figure 11). Therefore at a minimum, exposure experiments (especially EE2 experiments) should include 7 to 8 dose groups on a $\log_{10}$ scale. Human errors were made when selecting the dose range for Trenbolone. In the future, instead of conducting a literature review to select an appropriate dose range, experiments will be conducted to determine the appropriate dose range. Preliminary range-finding experiments in more recent studies improved dose selection (unpublished).

Despite these limitations in the experimental design, the results support the hypothesis that tPODs from the supplemented ZFET test are protective of relevant PODs from the literature related to overt toxicity. Specifically, experimental tPODs were protective of the acute overt toxicity PODs and the chronic overt toxicity PODs from juvenile and adult fish exposure experiments in the literature.

Importance of quality control and recommendations for the future

Diligent quality control measures are necessary for regulatory acceptance. Chapter 3 included many quality control measures and thresholds at each step in the analysis process (Figure 5), as well as before RNA sequencing (e.g., sample libraries <3nM were not sequenced).

Due to poor RINs (perhaps from many freeze-thaws), high rRNA contamination levels and poor mapping results, a transcript-level gene expression POD was unavailable for the TGSH exposure experiment. The samples were removed during the 70% mapping filtering criteria (as per R-ODAF). Although the tPOD is always the most sensitive experimental POD, diligent

quality control measures should ensure high-quality data and reliable results since the gene expression endpoint is prone to common RNA sequencing errors at multiple steps.

Of course, it is possible that RNA sequencing libraries can be contaminated with rRNA, gDNA, or interspecies RNA during library preparation (Supplementary Figures 27 and 28). To mitigate rRNA contamination, RNAseq experiments must undergo rRNA depletion or PolyA enrichment of mRNAs when appropriate for the experimental design (this experiment did include a PolyA capture protocol in the library preparation step). It is also possible that biological sample quality can degrade over time, especially after multiple freeze-thaw cycles (Supplementary Figure 25). Appropriate accommodations have been made in the O'Brien and Mennigen labs, including using RNAlater for long-term storage of biological samples in ultra freezers. Factors such as these (i.e., rRNA contamination and low RINs) can contribute to low-quality reads mapping to the transcriptome (Ji and Sadreyev, 2018). It has also been shown that PolyA enrichment of mRNAs does not perform well with partially degraded mRNAs (Zhao *et al.*, 2014). This occurred with the TGSH samples (n = 18) and the Aniline control dose 0mg/L replicate C sample.

RNA extraction methods are crucially important for high-integrity RNA samples. To mitigate the risk of degrading RNA samples before and during extraction, care should be taken during the extraction protocol to keep samples frozen for as long as possible (we used a frozen metal block) before the RLT buffer is added to the samples before homogenization. Randomized and semi-automated (via the Qiagen QiaCube RNA extraction instrument) RNA extractions of the Fadrozole, Glyphosate, and Aniline samples (labelled 'Round 1') began in September of 2021. The Qiagen QiaCube instrument experienced a power failure early in September 2021 and was serviced in December 2021, after which RNA extractions continued slowly. RIN scores

were poor until May 2022, when I used a frozen metal block (stored in the ultra freezer) to hold sample tubes and keep samples frozen before RLT buffer was added. Immediately after incorporating the frozen metal block into the RNA extraction protocol, there was a plastic crash in the Qiagen Qiacube. The remainder of the 'Round 1' samples (6/54), specifically, Fadrozole 0mg/L A, Fadrozole 0.00001mg/L B, Fadrozole 0.1mg/L C, Glyphosate 0.0001mg/L A, Glyphosate 0.01mg/L A and Aniline 10mg/L C, were extracted by hand with the frozen block. The RNA extractions of the Fadrozole, Glyphosate and Aniline samples concluded in July 2022. None of these samples from 'Round 1' were outliers. The 1-Decanol, Fenitrothion, Malathion, Prochloraz and Trenbolone ('Round 2') samples were all extracted by hand using the frozen metal block. RNA extractions for the 'Round 2' samples began in late June 2022 and concluded in July 2022. The TGSH and EE2 samples were extracted separately (by hand) in a different laboratory (Mennigen lab at uOttawa) as part of Tyler Nguyen's Master's thesis project in March 2021. It should also be noted that the remainder of the 19 other chemicals included in Chapter 2 were extracted along with the 10 chemicals included in Chapter 3.

After quantifying the RNAseq reads, a principal component analysis was conducted to identify outliers. The analysis revealed two distinct clusters of samples (Supplementary Figure 32). It is possible that the two clusters were driven mainly by the effect of sample RIN scores (Supplementary Figure 25) and possibly exacerbated by environmental laboratory conditions, different reagents, or different laboratory technicians on different days (Jan 6th and Jan 11th) during library preparation (Supplementary Figure 44). I think it is likely, however, that the two main clusters were driven by the 'Round 1' and 'Round 2' extractions, more specifically, the effect of the frozen block on the RINs of the extracted samples in the 'Round 2' cluster (Supplementary Figure 43). The effect of using a frozen metal block during RNA extractions can

also be seen in Supplementary Figure 37. These sample clusters align with the RINs in Supplementary Figure 25. For the dose-response analysis, individual chemicals are analyzed, so all samples within a chemical are clustered together. Thus, the effects driving the distinct clusters did not conflict with the dose-response analysis. However, caution should be exercised when attempting to interpret effects across chemicals (this may be a temptation when comparing similar chemicals or chemicals within the same class).

Another issue to discuss is removing outlier samples and the influence of additional factors on the DEG analysis. When samples are removed for reasons such as poor mapping results, errors in library preparation or not clustering with other replicates in the same group after quantification, the smaller sample size affects the results of the DEG analyses. For example, the Aniline, Glyphosate, and Fadrozole experiment analyses are affected by removing outliers. This is reflected by the difference in the number of DEGs identified by William's trend test and fold-change filter of 1.5 times (Figure 10) and the number of DEGs in specific dose groups determined by the R-ODAF DEG pipeline (Figure 6). In the case of the Fadrozole experiment, removing replicate A from the highest dose group (1mg/L) is correlated with a large number of DEGs identified by William's trend test and fold-change filter of 1.5 times in the BMD analysis (Figure 10). Factors influencing or removing the outlier sample could have also contributed to the U-shaped dose-response relationship observed in DEG analysis from the R-ODAF pipeline (Figure 6). Generally, when a non-monotonic dose-response relationship was observed from Figure 6, the number of DEGs identified by William's trend test with a fold-change filter of 1.5 times was affected, either by a reduced number of DEGs in comparison (as was the case with fenitrothion, glyphosate and Aniline) or an exaggerated number of DEGs in comparison (as was the case with Fadrozole) (Figure 10). However, the effects of outlier samples on the DEG

analysis are most apparent when the outlier samples are present in the control dose group. This is the case for the Aniline exposure experiment. The effects of removing one of the control group replicates are most apparent in the R-ODAF DEG analysis, where only three unique genes were differentially expressed (Figure 6). This is likely due to the crucial aspect of comparing treated samples (n = 3 in most cases) to the control samples (n = 2). William's trend test and a fold change filter of 1.5 times may be less sensitive to the effects of removing outlier samples since the number of DEGs in Figure 10 was much different from the number of DEGs in Figure 6 (R-ODAF_DEG_Report.html).

In the future, I recommend that a dose range finding experiment be included for each chemical, and a minimum of 7 dose groups should be included (as well as an 8th internal solvent control group). In addition to these two changes, future experiments should include 4 replicates instead of 3. In an experimental design with 3 replicates, none of the sample replicates can fail the minimum 1 CPM gene filter (Verheijen *et al.*, 2022). Including 4 replicates allows 1 sample to fail the R-ODAF relevant gene filter, which has a cutoff of 75% (Figure 5).

3.6. Conclusion

Overall, my analysis of exposed early-stage zebrafish transcriptomes: 1) increased the sensitivity of the internationally standardized ZFET test (OECD TG 236); the potencies of the experimental tPODs were lower than the potencies of the experimental overt toxicity PODs, 2) revealed DEGs, biological processes and pathways that inform potential chemical mechanisms of action and overall enriched the information from the ZFET test, and 3) derived tPODs that were protective of known chronic overt toxicity effects (death, deformity and reproduction) in juvenile

and adult fish from the literature; the potencies of the experimental tPODs were lower than the potencies of the literature-based chronic overt toxicity PODs.

A comparison of the most sensitive (lowest potency) PODs from each experimental endpoint from both Chapter 2 and Chapter 3 (overt toxicity, swimming behaviour, oxidative metabolism-dependent energy expenditure and transcriptomic/gene expression) revealed that significant changes in gene expression were generally observable at lower potencies than overt toxicity, swimming behaviour and oxidative metabolism-dependent energy expenditure effects. Although tPODs were typically derived concurrently with overt toxicity, swimming behaviour and oxidative metabolism-dependent energy expenditure PODs, the certainty of the overt toxicity, swimming behaviour, and oxidative metabolism-dependent energy expenditure PODs were below 50%. The lack of certainty in PODs from Chapter 2 can be attributed to variability in the data stemming from various sources of variability such as individual differences between fishes, breeding batch effects, environmental acclimation time, technical sources of variability from recording software and tools, analysis filtering methods, etc. However, the Chapter 2 endpoint effects can serve as a valuable source to validate the observed significant changes in gene expression.

By integrating multiple endpoints that measure diverse effects, my NAM provides various endpoints to potentially ground-truth observed transcriptomic effects. Chapter 4 will further explore the overall integration of the various endpoints for assessing the potential hazards of a diverse suite of chemicals for environmental and health risk assessment

Chapter 4 - Summary and Conclusions

4.0. Background

There is a growing concern in environmental toxicology over the presence and persistence of environmental contaminants. High-risk substances are known for their ability to persist in the environment, bioaccumulate, and cause various health issues, such as carcinogenicity, mutagenicity, and reproductive toxicity. The continuous introduction of new substances into the global market and environment challenges toxicologists and chemical risk assessors to assess hazards and risks. NAMs such as TDRM are becoming increasingly crucial in toxicology research as they provide reliable alternatives to traditional animal testing methods and help better assess the potential hazards associated with human and environmental exposure to substances. The recent enactment of legislation like the royal ascent of Bill S-5, known as the Strengthening Environmental Protection for a Healthier Canada Act in June (Bill S-5, 2023), underscores the increasing importance of incorporating NAMs in ecotoxicology research and environmental and health risk assessment to better protect the environment and well-being of Canadians.

Environmental and health risk assessors are particularly interested in chronic toxicity effects such as endocrine disruption and cancer. A notable challenge in risk assessment is the reliance on toxicity data from acute animal exposure studies due to the resource-intensive nature of chronic animal exposure studies. The literature unequivocally shows that tPODs from acute exposures derived from TDRM are protective of estimates of chronic toxicity, thereby significantly reducing the need for traditional chronic animal exposure studies (Pagé-Larivière *et al.*, 2019; Thomas *et al.*, 2013, 2011; Farmahin *et al.*, 2017; Moffat *et al.*, 2015; Jackson *et al.*,

2014; Dong *et al.*, 2016; Webster *et al.*, 2015). However, the majority of transcriptomic evidence comes from adult rodents. The sparsity of transcriptomic evidence in zebrafish embryos is a knowledge gap that currently hinders the use of zebrafish embryos for 21st-century toxicity testing. Nevertheless, Environment and Climate Change Canada is investigating the use of TDRM in the zebrafish embryo/larval model as an ecologically relevant alternative model for testing in adult fish. Zebrafish embryos and larvae, particularly those not yet independently feeding (i.e. < 6dpf), have emerged as valuable NAM, gaining recognition in the European Union, United States of America and Canada (Cassar *et al.*, 2020; Sipes *et al.*, 2011; Howe *et al.*, 2013; Sobanska *et al.*, 2018; OECD, 2013). The zebrafish embryo model holds numerous advantages: it is a cost-effective option, exhibits transparency during the well-characterized initial developmental stages (Kimmel *et al.*, 1995), and shares genetic similarities with other aquatic vertebrates and terrestrial mammals (Dai *et al.*, 2014), including humans (Howe *et al.*, 2013; Cassar *et al.*, 2020), and is capable of whole organism-level tissue interactions. Furthermore, it is a more ethical model compared to utilizing juvenile or adult fish or mammals.

There is growing interest in whether short-term changes in gene expression observed in zebrafish embryos can serve as predictive or protective indicators of chronic toxicity effects. Researchers are exploring integrating higher-order organismal responses, such as swimming behaviour and oxidative metabolism-dependent energy expenditure, to evaluate chemical hazards comprehensively. However, a fundamental uncertainty remains concerning the reliability of short-term molecular or swimming behaviour changes observed in animal alternative models, especially in early-life stages, for predicting and protecting against chronic toxicity effects. Furthermore, there is a lack of standardization in the field. Further discussion on standardization

and how my thesis is beginning to address standardization is provided in section ‘4.6. Standards for regulatory uptake’.

4.1. Hypothesis, objectives and predictions

I aimed to demonstrate that using the zebrafish embryo model in overt toxicity, swimming behaviour, oxidative metabolism-dependent energy expenditure, and transcriptomic assays can be a sensitive, informative, and protective animal alternative screening method for chemical hazard assessments of a diverse suite of chemicals. Overall, a more integrated multiple endpoint approach to the internationally standardized ZFET test (OECD TG236) can help elucidate potential chemical mechanisms of action and health hazards and establish confidence in TDRM for risk assessments and hazardous chemical screenings. I hypothesized that including swimming behaviour, oxidative metabolism-dependent energy expenditure, and gene expression endpoints in the standard ZFET test will make the ZFET more sensitive, informative and protective than when only standard mortality and deformity endpoints are considered. I predicted that: 1) PODs from swimming behaviour, oxidative metabolism-dependent energy expenditure, and gene expression endpoints would be more sensitive (lower POD) than experimental PODs from the overt toxicity (mortality and deformity) endpoint; 2) the various endpoints would also provide critical information on the potential mechanisms of action consistent with the known mechanism for well-characterized toxicants; and 3) the experimental tPOD will be protective of literature-based acute and chronic toxicity PODs (based on death, deformity and reproductive effects) from traditional juvenile and adult fish toxicity tests. By using swimming behaviour, energy expenditure and transcriptomic endpoints, this supplemented ZFET tests could provide valuable insights into potential neurotoxic and endocrine-disrupting mechanisms of action.

4.2. Swimming behaviour, oxidative metabolism-dependent energy expenditure, and gene expression endpoints increase the sensitivity of the ZFET test

The standard ZFET test (OECD TG 236) is designed to determine the acute toxicity of chemicals in embryonic stages of zebrafish over 96h (OECD, 2013). Four standard endpoints are recorded as indicators of lethality over the exposure period: 1) coagulation of fertilized eggs, 2) lack of somite formation, 3) lack of detachment of the tail-bud from the yolk sac and 4) lack of heartbeat. However, lethality is often not the most sensitive indicator of toxicity. Other subtle indicators, such as developmental deformities, swimming behaviour changes, energy expenditure perturbations, and differential gene expression, can be observed at lower doses than those that cause lethality. Therefore, it is important to consider multiple endpoints when assessing toxicity. I aimed to improve the sensitivity of the standard ZFET test by incorporating the additional endpoints mentioned above (deformities, swimming behaviour, energy expenditure and gene expression).

In Chapter 2, I incorporated the swimming behaviour and oxidative metabolism-dependent energy expenditure endpoints into the ZFET test. Figure 4 in Chapter 2 summarizes the BMDs and LOECs of all measured endpoints for all 29 tested chemicals. In the four instances where overt toxicity effects were observed (2,4-DNP, 3,4-DCA, 4-TPP, and DES), the PODs of the swimming behaviour and oxidative metabolism-dependent energy expenditure endpoints (including spurious swimming behaviour LOECs) were equally or more sensitive (lower POD) than the overt toxicity endpoint PODs (4/4 times). On average, the most sensitive swimming behaviour or oxidative metabolism-dependent energy expenditure POD (including

LOECs and excluding inactive BMDs) was ~32X lower than the overt toxicity POD. Swimming behaviour LOECs were included because active BMDs were unavailable for 2,4-DNP and 3,4-DCA to compare PODs quantitatively. For 2,4-DNP, the swimming behaviour endpoint (LOEC - 0.01mg/L) was 120X more sensitive than the lowest overt toxicity endpoint (BMD - 1.2mg/L). For 3,4-DCA, the swimming behaviour endpoint (LOEC - 0.1mg/L) was 1.6X more sensitive than the lowest overt toxicityendpoint (BMD - 0.16mg/L). For 4-TPP, the oxidative metabolism-dependent energy expenditure endpoint (BMD - 0.025mg/L) was 5.6X more sensitive than the overt toxicity endpoint (BMD - 0.14mg/L). Finally, for DES, the oxidative metabolism-dependent energy expenditure endpoint (BMD - 0.0006mg/L) was 1.7X more sensitive than the lowest overt toxicityendpoint (BMD - 0.001mg/L).

In many instances, significant swimming behaviour and oxidative metabolism-dependent effects were observed even when significant overt toxicity effects were not observed (20 times), providing further evidence that these endpoints enhanced the sensitivity of the ZFET. Out of the 29 chemicals tested, 24 chemicals elicited significant effects on zebrafish embryos after 5 days of exposure. No significant effects were observed in zebrafish embryos after exposure to 2,4-DMP, BPA, BPAF, SDS and TEG. The swimming behaviour or oxidative metabolism-dependent energy expenditure endpoint was more sensitive than the overt toxicity endpoint 24/24 times. Thus, by including swimming behaviour and oxidative metabolism-dependent energy expenditure endpoints, the sensitivity of the ZFET test was improved. However, if spurious swimming behaviour LOECs are excluded, PODs from the swimming behaviour and oxidative metabolism-dependent energy expenditure endpoints had lower PODs than the classical death and deformity endpoint for only 7 chemicals (17b-estradiol, 4-TPP, 4-TOP, Clofibric acid, DES, Fluoxetine and Flutamide). Despite this, the sensitivity of the ZFET test was still improved.

In Chapter 3, I incorporated the gene expression (transcriptomic) endpoint in the ZFET test. Figure 11 in Chapter 3 summarizes the most sensitive tPODs and overt toxicity PODs for 9 chemicals (excluding TGSH from the original 10 chemicals). Statistical LOECs were excluded to simplify POD comparisons. In all 9 instances, the certainty of the overt toxicity, swimming behaviour and oxidative metabolism-dependent energy expenditure PODs was below 50%. If only these endpoints were evaluated, the compounds would be considered non-active. However, a tPOD could be detected in all of these instances, suggesting potential chronic toxicity. Because the compound did not induce overt, behavioural or metabolic effects, a quantitative comparison to the tPODs is impossible. Nevertheless, these results suggest that the transcriptomic endpoint was the most sensitive in the integrated ZFET.

Overall, Chapter 2 and Chapter 3 showed that the ZFET test is more sensitive when supplemented with metabolic, behavioural and transcriptomic endpoints. The PODs of these additional endpoints were generally lower than the overt toxicity PODs. Often, the overt toxicity endpoint was not sensitive enough to derive PODs. The inclusion of LC_{50}s or EC_{50}s in the dose range is recommended for future experiments to facilitate comparisons of PODs. In contrast, the swimming behaviour, oxidative metabolism-dependent energy expenditure and transcriptomic endpoints were sensitive enough to derive PODs in most cases.

4.3. Information from overt toxicity, swimming behaviour, oxidative metabolism-dependent energy expenditure, and gene expression facilitates a more holistic understanding of potential chemical mechanisms of action and hazard

One of the primary objectives of using NAMs is to provide biological information of equivalent or better scientific quality and relevance than traditional animal toxicity tests that will support regulatory decisions (US EPA, 2018). By integrating many endpoints, I aimed to establish a NAM that provides high-quality and relevant biological information, specifically about potential toxic mechanisms of action and effects.

In Chapter 2, I gathered information about potential neurotoxic and metabolism-disrupting effects by utilizing the swimming behaviour and alamarBlue® assay. Figure 4 summarizes the derived BMDs and LOECs from each chemical exposure experiment. Overall, 17b-estradiol, 4-TPP, 4-TOP, Clofibric Acid, DES, Fluoxetine and Flutamide elicited significant oxidative metabolism-dependent energy expenditure effects, and 4-TPP and Fluoxetine elicited significant effects on the swimming behaviour of zebrafish embryos suggesting potential hormone-related metabolism-disrupting or neurotoxic/neurodevelopmental hazards. In the Discussion section of Chapter 2, I investigated the effects of Fluoxetine and cited literature that supports the hypothesis that Fluoxetine may be eliciting effects through inhibition of serotonin transporters and cholinesterase activity (Airhart *et al.*, 2007; De Farias *et al.*, 2019; Diaz-Camal *et al*., 2022; Zindler *et al*., 2020). However, the endpoints in Chapter 2 alone were not information-rich enough to interpret potential mechanisms of action for each chemical.

To further explore the potential cause of swimming behaviour and oxidative metabolism-dependent energy expenditure effects, I analyzed the transcriptomes of the zebrafish embryos in Chapter 3.

Transcriptomic PODs were derived from DEG BMDs, and I performed a GSEA to investigate the potential mechanisms of action (Verheijen *et al.*, 2022; National Toxicology Program, 2018). GSEA of the EE2 chemical exposure experiment revealed that zebrafish larvae had significantly enriched biological pathways related to visual transduction, sensory perception and the canonical retinoid cycle in rods (linked to twilight vision). In the Chapter 3 Discussion section, I cited sources validating these results. EE2 affects the development of zebrafish nervous and sensory systems and larval swimming behaviour (Nasri *et al.*, 2021). I concluded that perturbation of the retinoid (visual) cycle is the potential mechanism of action for EE2, supported by the literature (Kiser *et al*., 2014; Martínez *et al*., 2018). The raw average total distance swimming behaviour (in mm) plots in Supplementary Figure 24 from Chapter 2 support these results in Chapter 3. It can be seen in Supplementary Figure 24 that EE2 elicits effects on larval zebrafish swimming behaviour (especially at the highest dose of EE2 tested - 0.001mg/L). The results in Chapter 3 are also supported by Figure 2 of Chapter 2, which shows statistically significant effects on zebrafish larvae's average total distance swimming behaviour at the highest dose (0.001mg/L). Thus, the integrated information from Chapter 2 and Chapter 3 was valuable for interpreting the toxicity test results of the EE2 exposure experiment. EE2 was chosen as a proof-of-concept to illustrate that the data from Chapters 2 and 3 provided rich and informative results.

The true strength of this method is how the endpoints are integrated and collected concurrently on the same population of zebrafish larvae. One major challenge for NAMs to use

transcriptomics is interpreting differential gene expression and correlating it with biologically significant effects (Johnson *et al*., 2022; National Toxicology Program, 2018). Integrating overt toxicity, swimming behaviour, oxidative metabolism-dependent energy expenditure, and gene expression data creates a holistic foundation for formulating biological interpretations of toxic effects. Thus, this integrated NAM provides more relevant biological and mechanistic information than traditional juvenile and adult animal toxicity tests that can better inform regulatory decisions.

4.4. Transcriptomic PODs from short-term zebrafish embryo toxicity tests are protective of PODs from acute and chronic juvenile and adult fish toxicity tests

Under the authority of CEPA 1999, Health Canada and Environment and Climate Change Canada are responsible for regulating industrial substances and protecting the environment and well-being of Canadians. With the recent amendment to CEPA 1999 (Bill S-5: Strengthening Environmental Protection for a Healthier Canada Act), the government of Canada is focused on assessing the hazards of new chemicals with NAMs to understand their toxic mechanisms of action better, and most importantly, protect Canadians and the environment from chemicals of concern (Bill S-5, 2023). Thus, NAMS must be capable of deriving PODs that are protective of known toxicity effects (including chronic toxicity effects). I aimed to demonstrate that my integrated NAM that utilizes several endpoints and techniques for deriving PODs, including TDRM, derives PODs from short-term toxicity tests that are protective (equal or lower) of chronic overt toxicity (death, deformity, reproduction) PODs from long-term juvenile and adult toxicity tests. Ultimately, experimentally derived PODs that are protective against chronic

toxicity PODs help regulators and risk assessors better protect the health of the environment and well-being of Canadians.

In Chapter 2, I demonstrated that early-stage zebrafish experimental PODs were protective of chronic overt toxicity effects in juvenile and adult fish for 3/29 chemical exposure experiments when excluding inactive (< 50% certainty) BMDs and swimming behaviour LOECs. Figure 4 shows that 4-TPP, Clofibric acid and Flutamide all had experimentally derived PODs from the swimming behaviour or oxidative metabolism-dependent energy expenditure endpoint that were protective of chronic overt toxicity PODs from juvenile and adult fish toxicity tests. However, including the spurious statistical LOECs in the POD comparison was more protective (i.e., PODs were protective of chronic toxicity PODs for 11/29 chemical exposure experiments). Figure 4 shows that 1-Decanol, 1-Octanol, 2,4-DNP, 4-TPP, 4-TOP, Aniline, Clofibric acid, Fenitrothion, Fluoxetine, Flutamide and Malathion all had experimental PODs protective of chronic toxicity PODs from the literature.

I concluded that on their own, swimming behaviour and oxidative metabolism-dependent energy expenditure PODs were not protective enough of chronic overt toxicity effects in juvenile and adult fish to substitute traditional juvenile and adult fish toxicity tests. I demonstrated that my NAM utilizing swimming behaviour and oxidative metabolism-dependent energy expenditure endpoints can be protective as long as the mechanism of action of chemicals acts on hormone-mediated metabolism, redox potential, energy expenditure, muscle/skeletal development or neurodevelopment/neurotransmission.

In Chapter 3, I derived transcriptomics PODs for 10 of the 29 chemicals tested in Chapter 2. I demonstrated that early-stage zebrafish experimental tPODs were protective of chronic overt toxicity effects in juvenile and adult fish 6/7 times (excluding EE2, Trenbolone and TGSH). The

dose ranges used for EE2 and Trenbolone experiments were too high to capture the chronic overt toxicity PODs from the literature, and TGSH failed the R-ODAF 70% mapping filter criteria. Thus, these compounds were omitted from the protectiveness assessment. Figure 11 shows that, on average, tPODs were 1X to 10X more sensitive (lower) than chronic overt toxicity PODs from juvenile and adult fish experiments in the literature. The tPOD from the Malathion exposure was the only tPOD explicitly unprotective of the chronic overt toxicity (death, deformity, reproduction) POD from juvenile and adult toxicity tests in the literature. However, the tPOD was protective of the acute overt toxicity (death) POD and within 10X the potency of the chronic overt toxicity POD. Unsurprisingly, tPODs were more protective than the PODs from Chapter 2.

Overall, I demonstrated that tPODs are better protectors of environmental and human health. Although PODs from swimming behaviour and oxidative metabolism-dependent energy expenditure can protect chronic overt toxicity effects, the methods could not capture all toxic effects because they only measured two endpoints specific to certain effects, such as metabolic disruption and neurotransmission (energy expenditure and swimming behaviour). PODs derived from DEGs can capture organism-wide molecular effects and thus are more protective. Despite tPODs being better protectors of environmental and human health, swimming behaviour and oxidative metabolism-dependent energy expenditure PODs are also valuable as more inexpensive means of deriving PODs for metabolism and endocrine disrupting chemicals and neurotoxins. One of the major advantages of this NAM is the ability to screen for understudied metabolism and endocrine-disrupting and neuroactive chemicals. These effects are typically elusive and can lead to chronic adversity in animals exposed over a long time. This integrated NAM is useful for chemical hazard screening and informing risk assessors and regulators so they

can make informed decisions to protect the environment's health and the well-being of Canadians.

4.5. Recommendations and cautions for future experiments

In the discussion sections of Chapters 2 and 3, I gave extensive and specific recommendations to improve the methods and identified several key issues. In Chapter 2, I found that the swimming behaviour data was highly variable (noisy), and the swimming behaviour analysis methods require improvements. The oxidative metabolism-dependent energy expenditure data was also variable but to a lesser extent. Potential sources of variability in Chapter 2 include, but are not limited to, 1) overall health and batch effects from zebrafish breeding stocks (e.g., sickness, disease, age, genetically distinct breeding 'batches'), 2) individual differences between larval fishes, 3) environmental factors (stressors such as noise, acclimation time, water quality, etc.), 4) assay reagent quality (i.e. almarBlue®), 5) technical artifacts from recording instruments, 6) analytical sources of variability (data transformations, normalization, etc.) and 7) differences in chemical mechanisms of action. In Chapter 3, I found that differences in RNA extraction methods between batches (i.e. extracting frozen samples vs. extracting samples on ice and using an automated RNA extraction tool vs. extracting by hand) significantly affected sample RINs and sequencing results.

Here, I present several recommendations for future experiments, but this is not an exhaustive list: 1) extract RNA from frozen samples by hand, 2) minimize potential batch effects from breeding pairs (e.g., use a MEPs with maximum capacity to increase egg yield and allow for as many concurrent exposure experiments as possible, 3) increase sample sizes for analysis methods in Chapter 2 (e.g., n = 16 per dose group), 4) increase the number of sample replicates

to n=4 instead of n=3 for Chapter 3, 5) include 8 dose groups instead of 5 to better capture dose-response curve shapes, 6) perform dose-range-finding experiments to identify appropriate ranges for each chemical, 7) use fresh reagents for the alamarBlue® assay, 8) always use fresh/new 96-well plates, 9) use inert glassware whenever possible to prevent plastic-binding chemicals from binding to labware, 10) do not use methylene blue media because it interferes with the alamarBlue® assay and confounds toxicity tests (Hedge, 2022), 11) further elucidate nuanced swimming behaviour effects by analyzing additional endpoints of swimming behaviour like small bursting movement durations or inactivity duration, 12) incorporate the alamarBlue® and swimming behaviour assays together to address the confound of general activity for the alamarBlue® assay, 13) acclimate zebrafish larvae to the Viewpoint Zebrabox environment for a longer duration to reduce noise in the swimming behaviour data, 14) conduct a formal systematic literature review and compare equivalent experimental PODs and literature-based PODs (e.g., experimental BMD vs literature-based BMD) and 15) incorporate the analytical method recommendations I made in Chapters 2 and 3 (e.g., for the statistical LOECs in Chapter 2, use a magnitude of effect filter rather than a stricter p-value, for the tPOD analysis use the LCRD POD, in Chapter 3 it was important to remove outlier samples, etc.).

To improve the methods of Chapter 3, it would be wise to flash-freeze and sequence samples at different times (e.g., 24hpf, 48hpf, 72hpf and 96hpf) to elucidate the molecular cause of effects observed at 120hpf and characterize the dose-time-response relationship. Although transcriptomics is a powerful tool for toxicity tests, it is difficult to discern adverse effects from adaptive effects and determine the initiating molecular cause of effects. GSEA of dose-response data is also unreliable and needs standards and guidelines. There is yet to be a best practice for conducting GSEAs on dose-response toxicity data. The standard regulatory-accepted methods

available now are the R-ODAF (for comparing treated to control groups) and the standard NTP 'omics guidelines (for deriving gene set level and gene level tPODs) (Verheijen *et al.*, 2022; National Toxicology Program, 2018). Additionally, the NTP guidelines for deriving tPODs are imperfect, and work needs to be done to improve the sensitivity vs. specificity. The current standard method of deriving tPODs was established in 2018 and is prone to false positives. In unpublished work from the O'Brien, Mennigen, and Yauk labs, it was discovered that the NTP-recommended analysis method for deriving tPODs can result in high false positive rates. Thus, sufficient perturbations must be present and rigorous filtering must be applied to minimize false discoveries. Additionally, there are concerns about being overly conservative with tPODs. Since gene expression changes are the first cellular responses to toxicity, the doses at which responses are detected may be extremely low. On the one hand, this can be viewed as a good thing - being protective of adverse health effects - but the regulated community may have concerns about high false positive rates, and regulators may encounter challenges with risk management when tPODs are at such low doses. However, foundational works have shown that tPODs are generally conservative but not overly conservative and provide reasonable estimates of apical PODs (Thomas *et al.*, 2011, 2013; Farmahin *et al.*, 2017; Moffat *et al.*, 2015; Jackson *et al.*, 2014; Webster *et al.*, 2015).

Although integrating different techniques provides more valuable information about the potential mechanisms of action, the swimming behaviour and oxidative metabolism-dependent energy expenditure effects can only detect and not provide direct evidence for the mechanism behind those effects. Transcriptomics can detect sensitive molecular changes directly related to molecular mechanisms of action but only provides a snapshot of the transcriptome. Thus, it is warranted to investigate the time-dependent effects of chemical exposure on gene expression,

which may better inform risk assessors about the molecular mechanism of chemical hazards observed at later time points.

4.6. Standards for regulatory uptake

Standardized protocols and guidelines are crucial to ensure reliable and reproducible laboratory experiments. Standards increase regulatory acceptance for NAMs, allowing for better result comparison and POD evaluations. One widely recognized NAM model is the zebrafish embryo. Many studies have shown that the zebrafish embryo model is a strong model for evaluating behavioural toxicity/developmental neurotoxicity (Von Hellfeld, 2022; Atzei *et al.*, 2021; Klüver *et al.*, 2015; Lanzarin *et al.*, 2020; De Farias *et al.*, 2019; Chackal *et al.*, 2022; Al Shuraiqi *et al.*, 2021; Achenbach *et al.*, 2020; Muriana, 2022; Morash *et al.*, 2022; Hsieh *et al.*, 2019). Standard protocols like the ZFET test (OECD TG236) provide a solid foundation for this NAM to build upon. I explored integrating swimming behaviour, oxidative metabolism-dependent energy expenditure and gene expression endpoints in zebrafish embryo toxicity testing. However, standard protocols and procedures from the OECD still need to be developed for assessing these additional endpoints.

One important challenge with standardizing the swim behaviour assay is that there are many different ways to analyze the resulting data. The ‘Curvep’ BMD analysis method developed by Jui-Hua Hsieh *et al.* from the US NTP is an attractive approach since it generates BMDs that are familiar to regulators, which helps to move this approach toward regulatory acceptance (Hsieh *et al.*, 2019; Muriana *et al.*, 2022). However, as discussed in Chapter 2, the sensitivity of the Curvep approach can be negatively affected by highly variable data, which must be addressed before regulatory acceptance. Similarly, although it has been shown that the

alamarBlue® assay is a reliable method for measuring metabolic effects in zebrafish (Renquist *et al.*, 2013; Williams and Renquist, 2016; Mittal *et al.*, 2021; Chackal *et al.*, 2022), there are no standardized methods available. There is still work to be done with this method for the possibility of regulatory acceptance. Finally, the transcriptomic endpoint is the closest of all the endpoints to regulatory acceptance. The R-ODAF provides a standardized framework for preliminary DEG analysis of transcriptomic data, an important step toward regulatory acceptance (Verheijen *et al.*, 2022; National Toxicology Program, 2018). However, the R-ODAF does not guide TDRM analysis. Although the US NTP has published guidance for TDRM (National Toxicology Program, 2018), this approach may require further refinement before international regulatory acceptance. I incorporated both the R-ODAF and the NTP guidance for TDRM to maximize the chances of regulatory acceptance and reproducibility of my results. Although specific protocols have yet to be standardized, the OECD provides a Transcriptomic Reporting Framework to ensure high-quality data, biological interpretations, and reproducibility of the results (Harrill *et al.*, 2021). This reporting framework was not used in this thesis, but future publications stemming from this work will follow reporting frameworks. The Transcriptomic Reporting Framework guides the reporting of information and methods to facilitate transparency and reproducibility and allows for methods to develop further. For this NAM to be accepted in the future, standard reporting must be followed as best possible.

Careful curation of gene lists relevant to neurotoxic and metabolic effects will increase confidence in integrated approaches utilizing transcriptomics, swimming behaviour and oxidative metabolism-dependent energy expenditure endpoints. By creating curated gene lists relevant to metabolism disruption, endocrine disruption, redox potential, general activity, energy expenditure, neurotransmission and neurodevelopment, this integrated NAM could be an

extremely effective tool for screening metabolism, endocrine-disrupting and neuroactive chemicals. Anchoring transcriptomic effects with early-life stage organismal effects such as swimming behaviour or energy expenditure is an effective solution for increasing confidence in TDRM and detecting elusive chemical hazards. As long as researchers transparently report their findings and results (e.g., via a standard reporting framework), regulators will likely begin to accept NAMs.

4.7. Major findings and contributions

This thesis had three major experimental findings: 1) integrating swimming behaviour, oxidative metabolism-dependent energy expenditure and gene expression endpoints increased the sensitivity of the ZFET test, 2) integrating information from overt toxicity, swimming behaviour, oxidative metabolism-dependent energy expenditure and gene expression facilitated a more holistic understanding of potential chemical mechanisms of action and 3) tPODs from short-term early-life-stage zebrafish toxicity tests were protective of PODs from chronic and acute juvenile and adult fish toxicity tests reported in the literature. With the legislative support of Bill S-5, it is important to phase out adult animal testing, and these results establish confidence in the zebrafish embryo model as an alternative.

This research enhanced the field of toxicology by developing an integrated New Approach Methodology (NAM) aligned with established standardized and regulatory-accepted guidelines, approaches, and frameworks. I contributed to the field of toxicology a large set of transcriptomic, swimming behaviour and oxidative metabolism-dependent energy expenditure data that can be used for further analysis. Risk assessors carefully selected the 29 chemicals included in the study, ensuring a diverse range of chemicals with varying mechanisms of action.

I demonstrated that my integrated NAM can be used to screen potentially hazardous chemicals and inform risk assessors and regulators so that they can effectively fulfil their role of protecting the environment and the well-being of Canadians. The methods were most effective at screening hazardous neuroactive and metabolism-disrupting chemicals. The results from Chapter 2 helped to support the gene expression results from Chapter 3. Together, the integrated results were more informative than they were on their own and helped to establish confidence in biological interpretations. In the case of Fluoxetine and EE2, these interpretations were concurrent with their previously known chemical mechanisms of action (Table 1). Chapter 2 also provided certainty measures for BMDs, which helped build confidence in the results. Using zebrafish embryos to screen for swimming behaviour, oxidative metabolism-dependent energy expenditure, and gene expression effects is a medium-throughput method (it takes ~ 6 days to conduct) for screening potential metabolism and endocrine-disrupting, neuroactive chemicals. The applicability of the methods for screening for potential metabolism and endocrine-disrupting and neuroactive chemicals is significant because behavioural and oxidative metabolism-dependent energy expenditure effects are elusive and can lead to chronic adversity if animals are exposed for a long time. In the future, I or other researchers will analyze this rich data set and make further discoveries. The data and methods are publicly available on my GitHub repository (https://github.com/JoryC/MolToxLab). My methods and findings will also help inform other researchers on best practices and practices to avoid. My work has shown that this integrated NAM is potentially a sensitive, informative, and protective chemical hazard screening tool that can protect the health of the environment and humans alike.

Bibliography

Achenbach,J.C. *et al.* (2020) Comparison of the Zebrafish Embryo Toxicity Assay and the General and Behavioral Embryo Toxicity Assay as New Approach Methods for Chemical Screening. *Toxics*, **8**, 126.

Airhart,M.J. et al. (2007) Movement disorders and neurochemical changes in zebrafish larvae after bath exposure to fluoxetine (PROZAC). *Neurotoxicology and Teratology*, **29**, 652–664.

Al Shuraiqi,A. et al. (2021) Time-, dose- and transgenerational effects of fluoxetine on the behavioural responses of zebrafish to a conspecific alarm substance. *Environmental Pollution*, **270**, 116164.

Anders,S. and Huber,W. (2010) Differential expression analysis for sequence count data. *Genome biology*., **11**, r106

Ankley,G.T. *et al.* (2010) Adverse outcome pathways: A conceptual framework to support ecotoxicology research and risk assessment. *Environ. Toxicol. Chem.*, **29**, 730–741.

Arnold,C. (2015) Toxcast^{TM} wants you recommendations for engaging the broader scientific community. *Environ. Health Perspect.*, **123**, A20.

Atzei,A. *et al.* (2021) Developmental Neurotoxicity of Environmentally Relevant Pharmaceuticals and Mixtures Thereof in a Zebrafish Embryo Behavioural Test. *Int. J. Environ. Res. Public. Health*, **18**, 6717.

Bambino,K. and Chu,J. (2017) Zebrafish in Toxicology and Environmental Health. *Current Topics in Developmental Biology*, ***124***, 331-367.

Basu,N. *et al.* (2019) EcoToxChip: A next-generation toxicogenomics tool for chemical prioritization and environmental management. *Environ. Toxicol. Chem.*, **38**, 279–288.

Bernhardt,E.S. *et al.* (2017) Synthetic chemicals as agents of global change. *Front. Ecol. Environ.*, **15**, 84–90.

Bill S-5 (2023), An Act to amend the Canadian Environmental Protection Act, 1999, to make related amendments to the Food and Drugs Act and to repeal the Perfluorooctane Sulfonate Virtual Elimination Act, 1st Session, 44th Parliament, Canada, 2023 (assented to 13 June 2023), *CEPA 1999*, c 33.

Bridi,D. *et al.* (2017) Glyphosate and Roundup® alter morphology and behavior in zebrafish. *Toxicology*, **392**, 32–39.

Canadian Environmental Protection Act (1999) SC 1999, c. 33.-https://laws-lois.justice.gc.ca/PDF/C-15.31.pdf

Cassar,S. *et al.* (2020) Use of Zebrafish in Drug Discovery Toxicology. *Chem. Res. Toxicol.*, **33**, 95–118.

Chackal,R. *et al.* (2022) Metabolic Consequences of Developmental Exposure to Polystyrene Nanoplastics, the Flame Retardant BDE-47 and Their Combination in Zebrafish. *Front. Pharmacol.*, **13**, 822111.

Choi,J.K. (2010) Systems biology and epigenetic gene regulation. *IET Systems Biology*, **4**, 289–295.

Crizer,D.M. *et al.* (2021) Benchmark Concentrations for Untargeted Metabolomics Versus Transcriptomics for Liver Injury Compounds in *In Vitro* Liver Models. *Toxicol. Sci.*, **181**, 175–186.

Dai,Y.-J. *et al.* (2014) Zebrafish as a model system to study toxicology: Zebrafish toxicology monitoring. *Environ. Toxicol. Chem.*, **33**, 11–17.

De Farias,N.O. *et al.* (2019) Exposure to low concentration of fluoxetine affects development, behaviour and acetylcholinesterase activity of zebrafish embryos. *Comp. Biochem. Physiol. Part C Toxicol. Pharmacol.*, **215**, 1–8.

Diaz-Camal,N. *et al.* (2022) Consumption and ocurrence of antidepressants (SSRIs) in pre- and post-COVID-19 pandemic, their environmental impact and innovative removal methods: A review. *Sci. Total Environ.*, **829**, 154656.

Dong,H. *et al.* (2016) Toxicogenomic assessment of liver responses following subchronic exposure to furan in Fischer F344 rats. *Arch. Toxicol.*, **90**, 1351–1367.

EFSA Scientific Committee *et al.* (2022) Guidance on the use of the benchmark dose approach in risk assessment. *EFSA J.*, **20**.

Environment and Climate Change Canada (2020) Canadian Environmental Sustainability Indicators: Bisphenol A in water and sediment.-https://www.canada.ca/en/environment-climate-change/services/environmental-indicators/bisphenol-a-water-sediment.html

Ewels,P.A. *et al.* (2020) The nf-core framework for community-curated bioinformatics pipelines. *Nat. Biotechnol.*, **38**, 276–278.

Farmahin,R. *et al.* (2019) Hepatic transcriptional dose-response analysis of male and female Fischer rats exposed to hexabromocyclododecane. *Food Chem. Toxicol.*, **133**, 110262.

Glasauer,S.M.K. and Neuhauss,S.C.F. (2014) Whole-genome duplication in teleost fishes and its evolutionary consequences. *Mol. Genet. Genomics*, **289**, 1045–1060.

Government of Canada (2019) Chemicals Management Plan Progress Report. Issue 11, Fall 2018. https://www.canada.ca/en/environment-climate-change/services/evaluating-existing-substances/chemicals-management-plan-progress-report/fall-2018.html

Government of Canada and Government of the United States of America (1972) Great Lakes Water Quality Agreement. Compendium of Canada's Engagement in International

Environmental Agreements and Instruments. Compendium Edition 2022. https://www.canada.ca/en/environment-climate-change/corporate/international-affairs/partnerships-countries-regions/north-america/great-lakes-water-quality-agreement.html

Grammarly. "Grammarly GO." Grammarly, https://www.grammarly.com/go. I used Grammarly and the Grammarly GO AI assistance tool to help with writing grammar and rewording complex sentences for clarity.

Harrill,J.A., Everett,L.J., *et al.* (2021) High-Throughput Transcriptomics Platform for Screening Environmental Chemicals. *Toxicol. Sci.*, **181**, 68–89.

Harrill,J.A., Viant,M.R., *et al.* (2021) Progress towards an OECD reporting framework for transcriptomics and metabolomics in regulatory toxicology. *Regul. Toxicol. Pharmacol.*, **125**, 105020.

Hedge,J. (2022) Methylene blue and developing zebrafish--Effects on mortality, morphology, and behavior. Zebrafish husbandry workshop by Aquaneering Inc. within Aquaculture 2022 annual meeting presented by the World Aquaculture Society. The United States Environmental Protection Agency Office of Research and Development's Center for Computational Toxicology and Exposure. Presentation.

von Hellfeld,R. *et al.* (2020) Adverse effects in the fish embryo acute toxicity (FET) test: a catalogue of unspecific morphological changes versus more specific effects in zebrafish (Danio rerio) embryos. *Environ. Sci. Eur.*, **32**, 122.

von Hellfeld,R. *et al.* (2022) Zebrafish embryo neonicotinoid developmental neurotoxicity in the FET test and behavioral assays. *ALTEX*. **39**, 267-387

Hogan,K. *et al.* (2012) US EPA Benchmark Dose Technical Guidance. https://www.epa.gov/risk/benchmark-dose-technical-guidance

Howe,K. *et al.* (2013) The zebrafish reference genome sequence and its relationship to the human genome. *Nature*, **496**, 498–503.

Hsieh,J.H. *et al.* (2019) Application of benchmark concentration (BMC) analysis on zebrafish data: A new perspective for quantifying toxicity in alternative animal models. *Toxicol. Sci.*, **167**, 282–292.

Izadi,H. *et al.* (2012) Evaluation of the Benchmark Dose for Point of Departure Determination for a Variety of Chemical Classes in Applied Regulatory Settings. *Risk Anal.*, **32**, 830–835.

Jackson,A.F. *et al.* (2014) Case study on the utility of hepatic global gene expression profiling in the risk assessment of the carcinogen furan. *Toxicol. Appl. Pharmacol.*, **274**, 63–77.

Ji,F. and Sadreyev,R.I. (2018) RNA-seq: Basic Bioinformatics Analysis. *Curr. Protoc. Mol. Biol.*, **124**, e68.

Johnson,K.J. *et al.* (2022) A Transformative Vision for an Omics-Based Regulatory Chemical

Testing Paradigm. *Toxicol. Sci.*, **190**, 127–132.

Kais,B. *et al.* (2013) DMSO modifies the permeability of the zebrafish (Danio rerio) chorion-Implications for the fish embryo test (FET). *Aquat. Toxicol.*, **140–141**, 229–238.

Kassotis,C. *et al.* (2022) Nonylphenol Polyethoxylates Enhance Adipose Deposition in Developmentally Exposed Zebrafish. *Toxics*, **10**, 99.

Kimmel,C.B. *et al.* (1995) Stages of embryonic development of the zebrafish. *Dev. Dyn.*, **203**, 253–310.

Kinch,C.D. *et al.* (2015) Low-dose exposure to bisphenol A and replacement bisphenol S induces precocious hypothalamic neurogenesis in embryonic zebrafish. *Proc. Natl. Acad. Sci.*, **112**, 1475–1480.

Kiser,P.D. *et al.* (2014) Chemistry of the Retinoid (Visual) Cycle. *Chem. Rev.*, **114**, 194–232.

Klüver,N. *et al.* (2015) Fish Embryo Toxicity Test: Identification of Compounds with Weak Toxicity and Analysis of Behavioral Effects To Improve Prediction of Acute Toxicity for Neurotoxic Compounds. *Environ. Sci. Technol.*, **49**, 7002–7011.

Knapen,D. *et al.* (2018) Adverse outcome pathway networks I: Development and applications. *Environ. Toxicol. Chem.*, **37**, 1723–1733.

Knecht,A.L. *et al.* (2017) Developmental benzo[a]pyrene (B[a]P) exposure impacts larval behavior and impairs adult learning in zebrafish. *Neurotoxicol. Teratol.*, **59**, 27–34.

Kopylova,E. *et al.* (2012) SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics*, **28**, 3211–3217.

Kozol,R.A. *et al.* (2016) Function Over Form: Modeling Groups of Inherited Neurological Conditions in Zebrafish. *Front. Mol. Neurosci.*, **9**.

Kreuger,F. (2023) Trim Galore. https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/

Labib,S. *et al.* (2015) Nano-risk Science: application of toxicogenomics in an adverse outcome pathway framework for risk assessment of multi-walled carbon nanotubes. *Part. Fibre Toxicol.*, **13**, 15.

Lanzarin,G.A.B. *et al.* (2020) Behavioural toxicity of environmental relevant concentrations of a glyphosate commercial formulation - RoundUp® UltraMax - During zebrafish embryogenesis. *Chemosphere*, **253**, 126636.

Liu,J. *et al.* (2021) Occurrence, toxicity and ecological risk of Bisphenol A analogues in aquatic environment – A review. *Ecotoxicol. Environ. Saf.*, **208**, 111481.

Love,M.I. *et al.* (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.*, **15**, 550.

Malkiewicz,K. *et al.* (2009) Assessment factors for extrapolation from short-time to chronic

exposure—Are the REACH guidelines adequate? *Toxicol. Lett.*, **190**, 16–22.

Martínez,R. *et al.* (2018) Dose-dependent transcriptomic responses of zebrafish eleutheroembryos to Bisphenol A. *Environ. Pollut.*, **243**, 988–997.

Mittal,K. *et al.* (2021) Characterizing the effects of titanium dioxide and silver nanoparticles released from painted surfaces due to weathering on zebrafish (*Danio rerio*). *Nanotoxicology*, **15**, 527–541.

Mittal,K. et al. (2022) Resource requirements for ecotoxicity testing: A comparison of traditional and new approach methods (Preprint). *Pharmacology and Toxicology*.

Moffat,I. *et al.* (2015) Comparison of toxicogenomics and traditional approaches to inform mode of action and points of departure in human health risk assessment of benzo[*a*]pyrene in drinking water. *Crit. Rev. Toxicol.*, **45**, 1–43.

Mount,D.I. and Stephan,C.E. (1967) A Method for Establishing Acceptable Toxicant Limits for Fish—Malathion and the Butoxyethanol Ester of 2,4-D. *Trans. Am. Fish. Soc.*, **96**, 185–193.

Morash,M.G. *et al.* (2022) Assessing the Morphological and Behavioral Toxicity of Catechol Using Larval Zebrafish. *Int. J. Mol. Sci.*, **23**, 7985.

Muriana,A. *et al.* (2022) An Inter-laboratory Case Study to Harmonize Zebrafish Light-Dark Transition Test to Predict Developmental Neurotoxicity (WC11). The United States Environmental Protection Agency's Center for Computational Toxicology and Exposure. Presentation.

National Research Council (2009) Science and Decisions: Advancing Risk Assessment. *Natl. Acad.* https://doi.org/10.17226/12209

National Toxicology Program (2018) NTP Research Report on National Toxicology Program Approach to Genomic Dose-Response Modeling. *National Toxicology Program,* **5**, 1-44

North American Agreement on Environmental Cooperation (NAAEC) (2012) In, *Encyclopedia of U.S.-Latin American Relations*. CQ Press, 2300 N Street, NW, Suite 800, Washington DC 20037 United States. https://ustr.gov/sites/default/files/naaec.pdf

Nasri,A. *et al.* (2021) Ethinylestradiol (EE2) residues from birth control pills impair nervous system development and swimming behavior of zebrafish larvae. *Sci. Total Environ.*, **770**, 145272.

Nguyen,P. (2022) Transcriptomic Dose-Response Analysis in Zebrafish Embryos to Estimate Aquatic Toxicity of Plasticizers and Plastic Monomers (Master's Thesis).

OECD (2013) Test No. 236: Fish Embryo Acute Toxicity (FET) test, OECD guidelines for the testing of chemicals, Section 2, OECD Publishing, Paris, https://doi.org/10.1787/9789264203709-en.

OECD (2007) Test No. 426: Developmental Neurotoxicity Study, OECD guidelines for testing

of chemicals, Section 4, OECD Publishing, Paris, https://doi.org/10.1787/9789264067394-en.

Page,B. *et al.* (1993) A New Fluorometric Assay For Cytotoxicity Measurements In-Vitro. *Int. J. Oncol.*, **3**, 473-476

Pagé-Larivière,F. *et al.* (2019) Transcriptomic points-of-departure from short-term exposure studies are protective of chronic effects for fish exposed to estrogenic chemicals. *Toxicol. Appl. Pharmacol.*, **378**, 114634.

Papalou,O. et al. (2019) Endocrine Disrupting Chemicals: An Occult Mediator of Metabolic *Disease. Front. Endocrinol.*, **10**, 112.

Patel,H. (2023) nf-core/rnaseq: nf-core/rnaseq v3.12.0 - Osmium Octoppus. https://zenodo.org/record/7998767

Patro,R. *et al.* (2017) Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods*, **14**, 417–419.

Phillips,J. (2022) BMDExpress-3. https://github.com/auerbachs/BMDExpress-3

Price,P.S. *et al.* (2022) A Framework that Considers the Impacts of Time, Cost, and Uncertainty in the Determination of the Cost Effectiveness of Toxicity-Testing Methodologies. *Risk Anal.*, **42**, 707–729.

Pu,X. *et al.* (2022) Adverse Outcome Pathway Network-Based Chemical Risk Assessment Using High-Throughput Transcriptomics. *Adv. Toxicol. Risk Assess. Nanomater. Emerg. Contam.*, 307–324.

Reardon,A.J.F. *et al.* (2023) From vision toward best practices: Evaluating in vitro transcriptomic points of departure for application in risk assessment using a uniform workflow. *Front. Toxicol.*, **5**, 1194895.

Reid,R.M. *et al.* (2018) The validation of a sensitive, non-toxic in vivo metabolic assay applicable across zebrafish life stages. *Comp. Biochem. Physiol. Part C Toxicol. Pharmacol.*, **208**, 29–37.

Renquist,B.J. *et al.* (2013) Development of an assay for high-throughput energy expenditure monitoring in the zebrafish. *Zebrafish*, **10**, 343–352.

Roy,N.M. *et al.* (2016) Glyphosate induces neurotoxicity in zebrafish. *Environ. Toxicol. Pharmacol.*, **42**, 45–54.

Sedykh,A. (2016) CurveP Method for Rendering High-Throughput Screening Dose-Response Data into Digital Fingerprints. In, Zhu,H. and Xia,M. (eds), *High-Throughput Screening Assays in Toxicology*, Methods in Molecular Biology. Springer New York, New York, NY, **1473**, 135–141.

Sipes,N.S. *et al.* (2011) Zebrafish-As an integrative model for twenty-first century toxicity

testing. *Birth Defects Res. Part C - Embryo Today Rev.*, **93**, 256–267.

Sobanska,M. *et al.* (2018) Applicability of the fish embryo acute toxicity (FET) test (OECD 236) in the regulatory context of Registration, Evaluation, Authorisation, and Restriction of Chemicals (REACH): Fish embryo acute toxicity (FET) test and REACH regulation. *Environ. Toxicol. Chem.*, **37**, 657–670.

Soneson,C. *et al.* (2015) Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Research*, **4**, 1521.

Song,Y. *et al.* (2023) Multiomics Point of Departure (moPOD) Modeling Supports an Adverse Outcome Pathway Network for Ionizing Radiation. *Environ. Sci. Technol.*, acs.est.2c04917.

Srivastava,A. *et al.* (2019) Alignment and mapping methodology influence transcript abundance estimation. *Genome biology*, **21**, 239

Sun,J. et al. (2022) A review of environmental metabolism disrupting chemicals and effect biomarkers associating disease risks: Where exposomics meets metabolomics. *Environment International*, **158**, 106941.

Tal,T. *et al.* (2023) Building confidence in larval zebrafish behavior assays: From phenotypes to mechanisms (2023 ICCVAM Communities of Practice Webinar). Helmholtz Centre for Environmental Research - UFZ. Presentation

Thermo Fisher Scientific (2022) Ribosomal RNA Depletion | rRNA Depletion. https://www.thermofisher.com/ca/en/home/life-science/dna-rna-purification-analysis/rna-extraction/rna-applications/ribosomal-rna-depletion.html

Thomas,R.R. *et al.* (2013) Temporal concordance between apical and transcriptional points of departure for chemical risk assessment. *Toxicol. Sci.*, **134**, 180–194.

Thomas,R.S. *et al.* (2011) Application of transcriptional benchmark dose values in quantitative cancer and noncancer risk assessment. *Toxicol. Sci.*, **120**, 194–205.

Tsuji,R. and Crofton,K.M. (2012) Developmental neurotoxicity guideline study: Issues with methodology, evaluation and regulation. *Congenit. Anom.*, **52**, 122–128.

Tu,W. *et al.* (2019) Bioconcentration and Metabolic Effects of Emerging PFOS Alternatives in Developing Zebrafish. *Environ. Sci. Technol.*, **53**, 13427–13439.

US Environmental Protection Agency (2019) Statistics for the new chemicals review program under TSCA.-https://www.epa.gov/reviewing-new-chemicals-under-toxic-substances-control-act-tsca/statistics-new-chemicals-review

US EPA (2019) Administrator memo prioritizing efforts to reduce animal testing, September 10, 2019. *U. S. Environ. Prot. Agency*.-https://www.epa.gov/sites/default/files/2019-09/documents/image2019-09-09-231249.pdf

US EPA (2018) Strategic Plan to Promote the Development and Implementation of Alternative

Test Methods Within the TSCA Program. *HttpsWwwEpaGovSitesProductionFiles2018-06DocumentsEpaAltStratPlan6-20-18CleanFinalPdf*, 1–39.

US EPA (1994) ECO Update: Using Toxicity Tests in Ecological Risk Assessment. **2**. https://www.epa.gov/risk/using-toxicity-tests-ecological-risk-assessment

Verheijen,M.Ct. *et al.* (2022) R-ODAF: Omics data analysis framework for regulatory application. *Regul. Toxicol. Pharmacol.*, **131**, 105143.

Villeneuve,D.L. *et al.* (2014) Adverse Outcome Pathway Development II: Best Practices. *Toxicol. Sci.*, **142**, 321–330.

Webster,A.F. *et al.* (2015) Impact of genomics platform and statistical filtering on transcriptional benchmark doses (bmd) and multiple approaches for selection of chemical point of departure (PoD). *PLoS ONE*, **10**, 1–19.

Watford,S. *et al.* (2019) ToxRefDB version 2.0: Improved utility for predictive and retrospective toxicology analyses. *Reprod. Toxicol.*, **89**, 145–158.

Webster,A.F. *et al.* (2015) Impact of genomics platform and statistical filtering on transcriptional benchmark doses (bmd) and multiple approaches for selection of chemical point of departure (PoD). *PLoS ONE*, **10**, 1–19.

Williams,S.Y. and Renquist,B.J. (2016) High Throughput Danio Rerio Energy Expenditure Assay. *J. Vis. Exp.*, 53297.

Woodruff,T.J. *et al.* (2023) A science-based agenda for health-protective chemical assessments and decisions: overview and consensus statement. *Environ. Health*, **21**, 132.

Zhao,W. et al. (2014) Comparison of RNA-Seq by poly (A) capture, ribosomal RNA depletion, and DNA microarray for expression profiling. BMC Genomics, **15**, 419.

Zimmerman,D.W. (2004) A note on preliminary tests of equality of variances. *Br. J. Math. Stat. Psychol.*, **57**, 173–181.

Zindler,F. et al. (2020) Do environmentally relevant concentrations of fluoxetine and citalopram impair stress-related behavior in zebrafish (Danio rerio) embryos? *Chemosphere*, **261**, 127753.

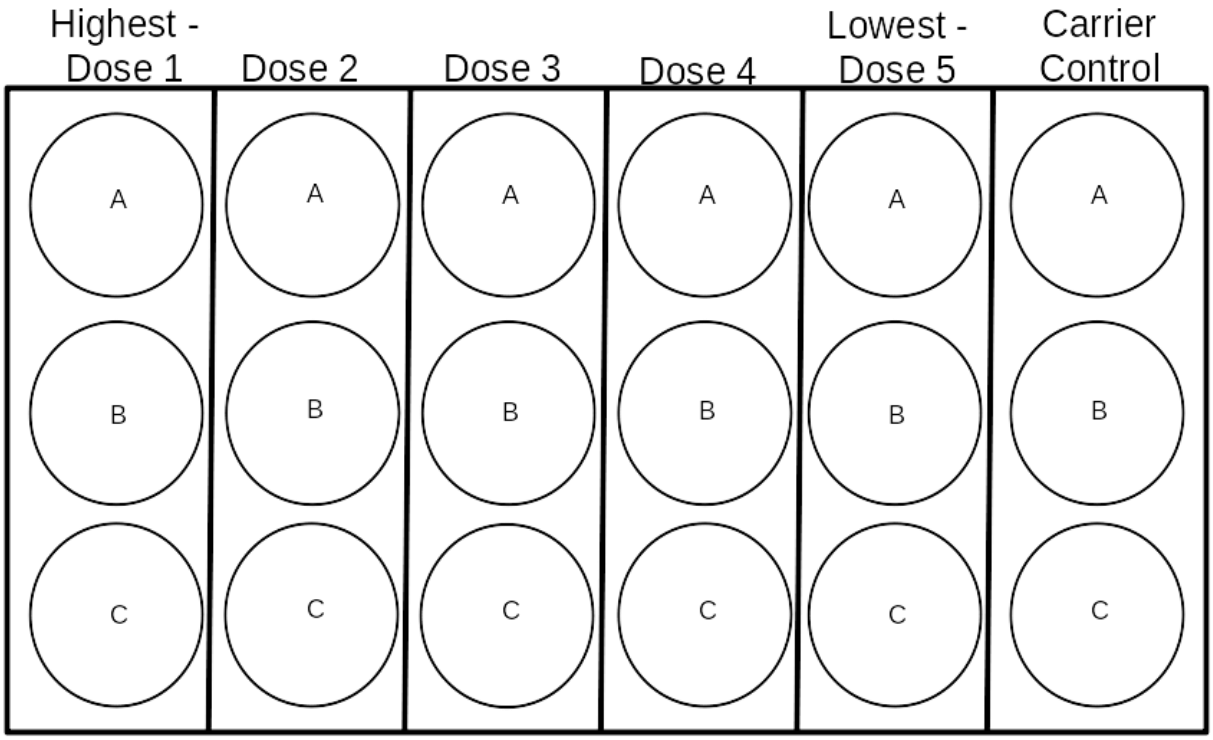
Appendix

Supplementary Table 1: Atlas of developmental toxicity endpoints scored during exposure experiments. Lethal endpoints described in the OECD FET test guideline 236 are described in the Lehtal endpoint column, additional sub-lethal endpoints relating to deformities from Hellfeld *et al.* are described in the Sub-lethal endpoint column and major key developmental events that could be the potential source of perturbation are described in the final column.

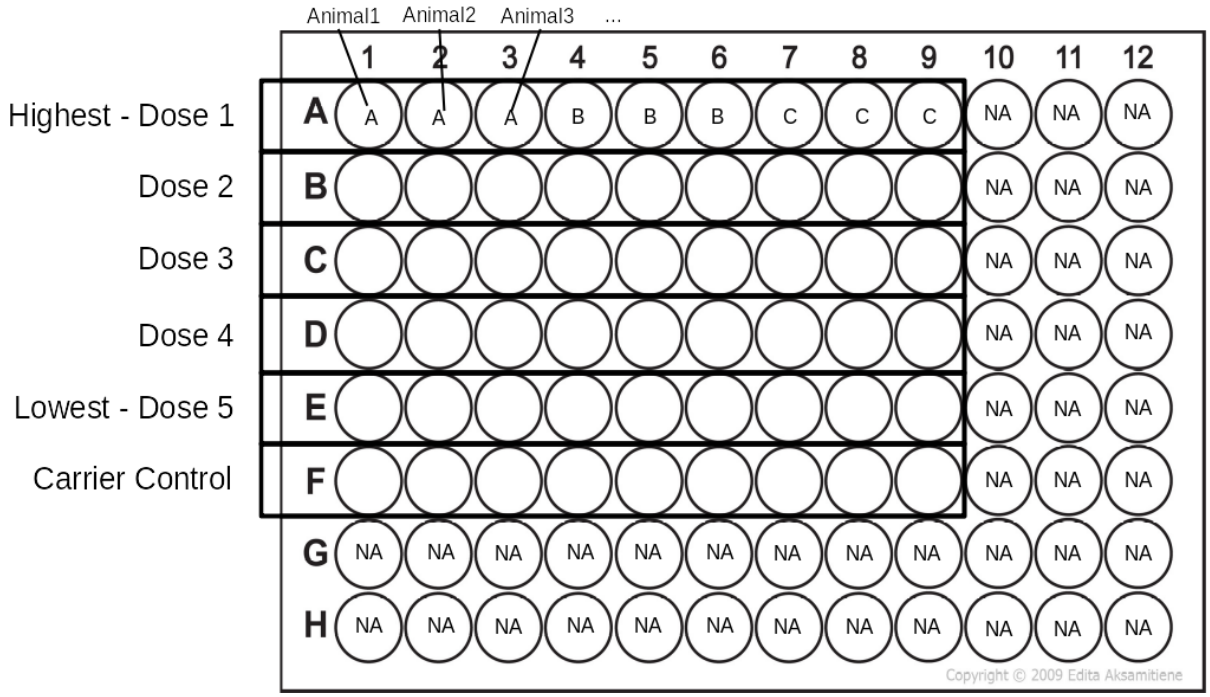
Time (hpf)	Developmental stage	Lethal endpoint (OECD TG 236)	Sub-lethal endpoint (Hellfeld *et al.*, 2020)	Events (Kimmel *et al.*, 1995)
1.5	partial cleavage 16-cell stage			blastodisc cleaves into blastomeres
6	gastrulation period embryonic shield stage	coagulation of the embryo ('stone') or gastrulation arrest		transcription of the zygotic genome begins formation of the dorsal-ventral and anterior-posterior axis
24	segmentation complete (26 somites visible) pharyngula period beings primordium-5 stage	coagulation of the embryo ('stone') lack of somite formation non-detachment of the tail	spinal cord malformation	embryos begin muscle twitches and movements the tail extends and detaches from the yolk sac five lobes of the brain have developed: telencephalon, diencephalon, epiphysis, mesencephalon, and the cerebellum retina and skin begin early pigmentation (melanogenesis) somites form early organs and tissues

48	pharyngula period complete hatching period begins long pectoral fin stage	coagulation of the embryo (‘stunted’) lack of somite formation non-detachment of the tail lack of heartbeat	spinal cord malformation edema (swelling) malformed heart yolk deformation/reduced absorption eye malformation	the retina is distinctly pigmented and competent ears fully develop embryos begin to hatch from their chorion the circulatory system develops, and the heart is beating pectoral and caudal fins develop behaviour emerges as rhythmic bouts of swimming and swim bursts in response to tactile stimuli liver and swim bladder develop
72	hatching period ends early larva stage protruding mouth stage	coagulation of the larvae (‘stunted’) non-detachment of the tail lack of heartbeat	spinal cord malformation edema (swelling) malformed heart yolk deformation/reduced absorption eye malformation lack of pigmentation	heart effectively circulates blood across the circulatory system spine straightens fins develop further, and larvae start becoming more active

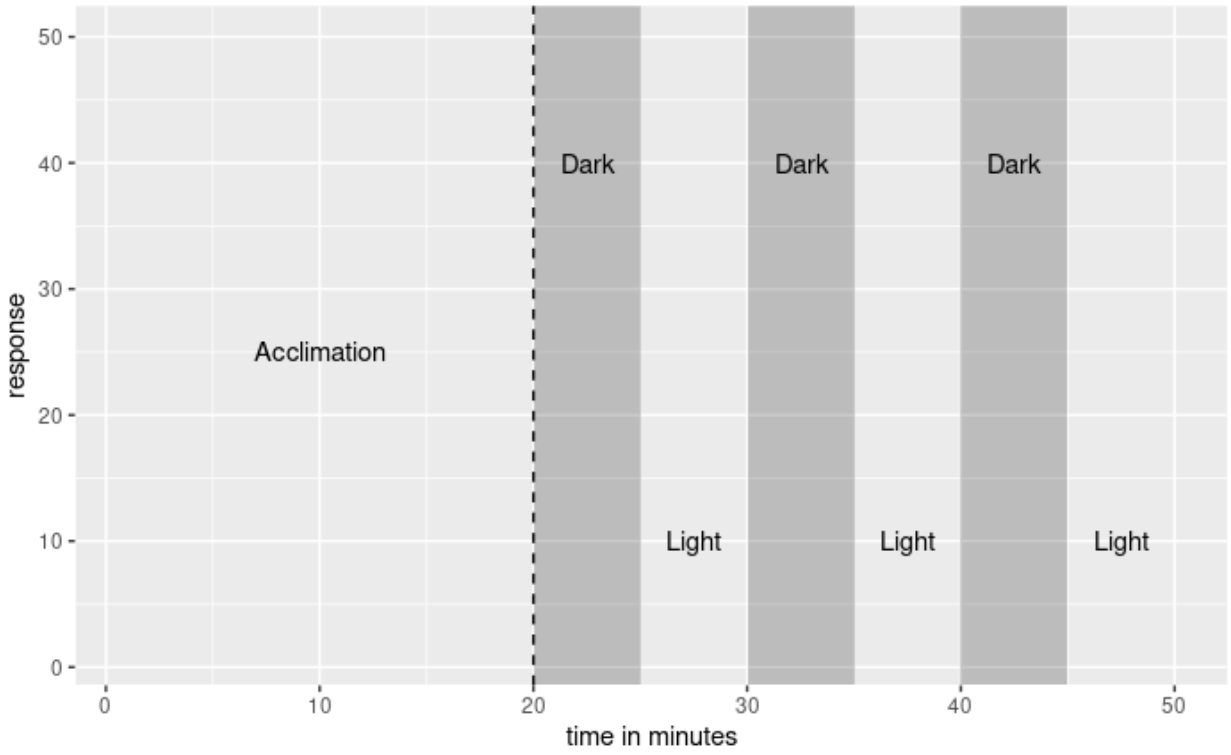
96	larval stage (day 4)	coagulation of the larvae (‘stunted’) non-detachment of the tail lack of heartbeat unhatched larvae	spinal cord malformation edema (swelling) malformed heart yolk deformation/reduced absorption eye malformation lack of pigmentation	inflation of the swim bladder continuation of craniofacial development general activity increases (prey-seeking behaviour), and a swift escape response emerges the gut is fully developed larvae gain biotransformation capabilities
120	larval stage (day 5)	coagulation of the larvae (‘stunted’) non-detachment of the tail lack of heartbeat unhatched larvae	spinal cord malformation edema (swelling) malformed heart yolk deformation/reduced absorption eye malformation lack of pigmentation	yolk almost fully absorbed full craniofacial development allows for the active capture of prey



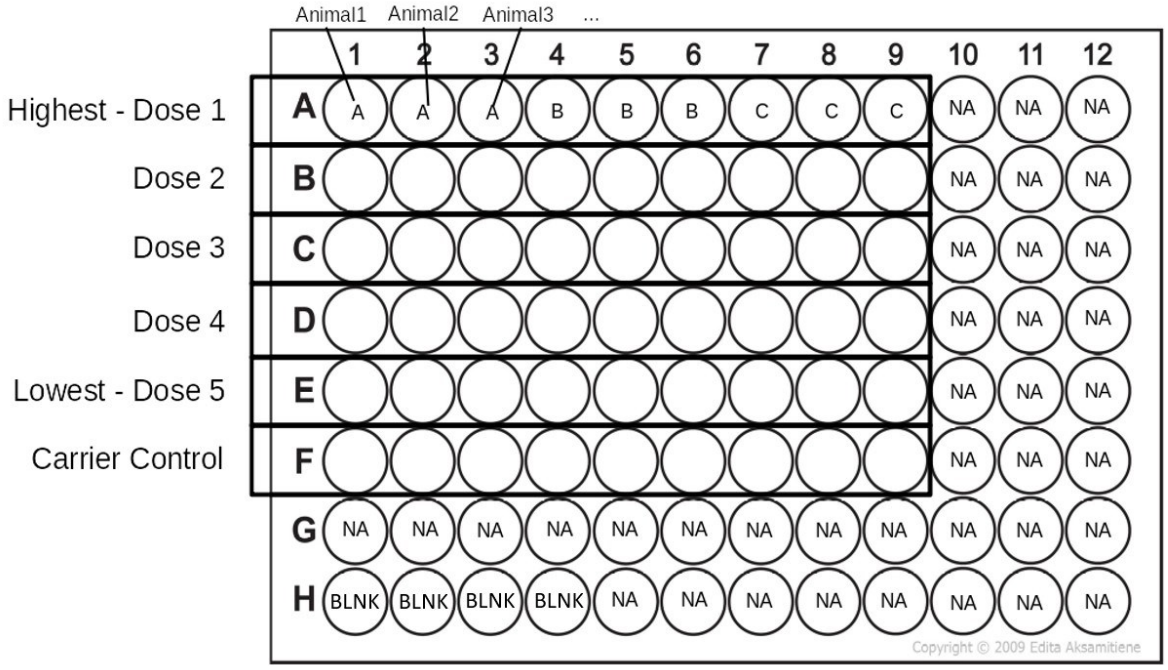
Supplementary Figure 1: Chemical exposure experiment layout. Dose groups contained three sub-groups with the same concentration (A, B and C). Each Petri dish contained 20 eggs/embryos/larvae per dish and had a final volume of 16mL.



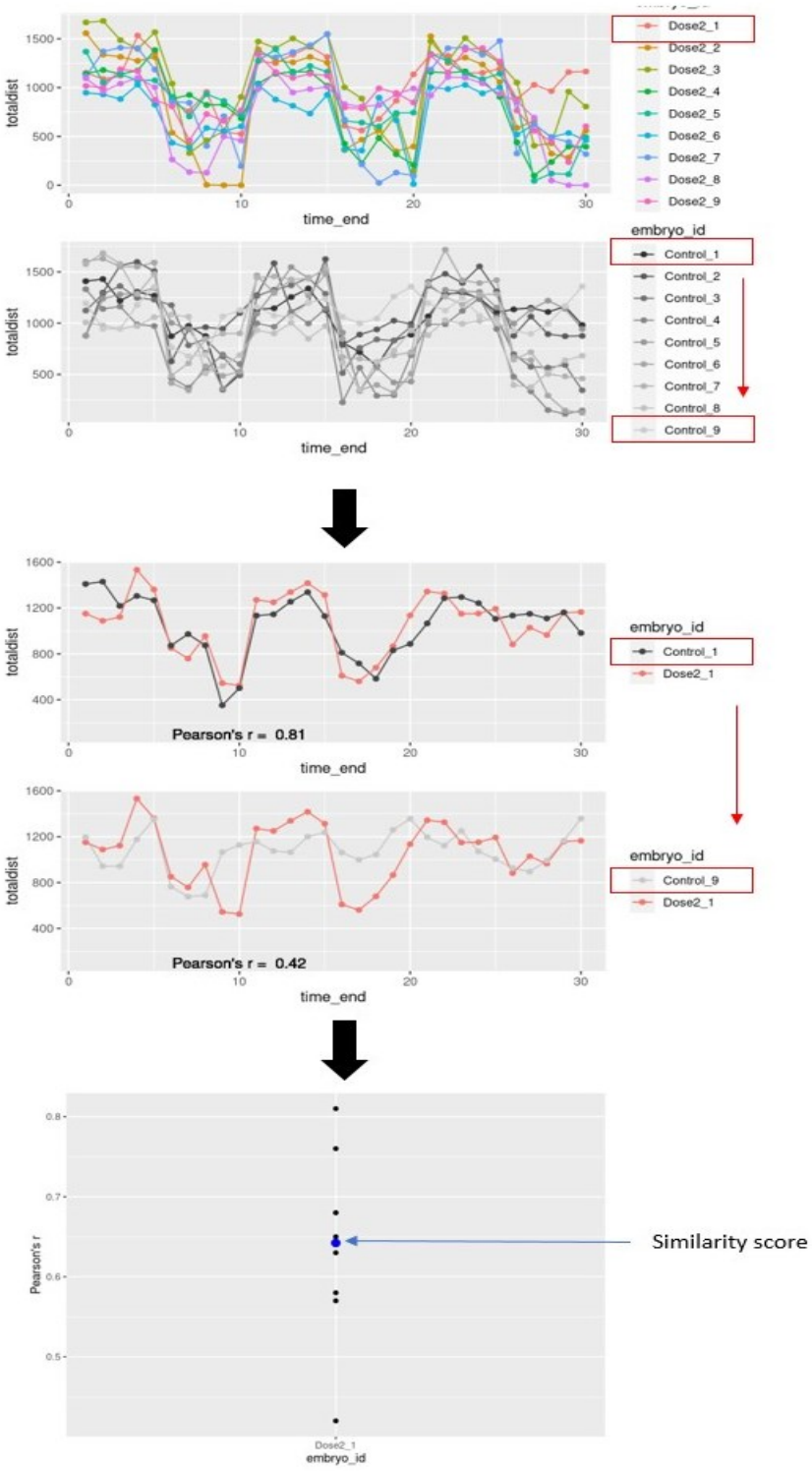
Supplementary Figure 2: 96-well plate layout for the behavioural assay.



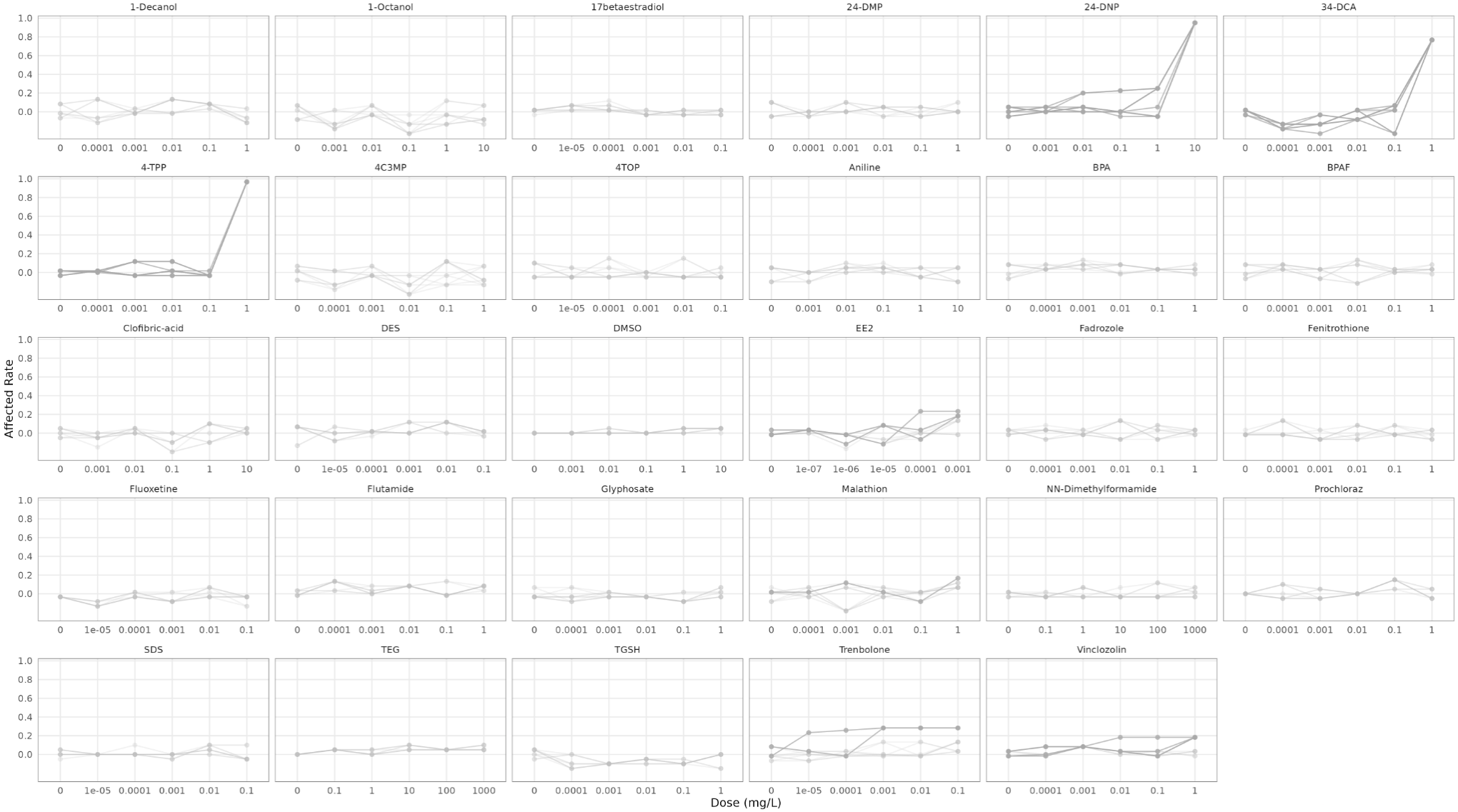
Supplementary Figure 3: A visual description of the behavioural assay protocol. There is a 20-minute acclimation period, then alternating 5-minute light and dark cycles. Dark cycles induce an increase in response from larvae, while returning to the light condition inhibits the increase in response.



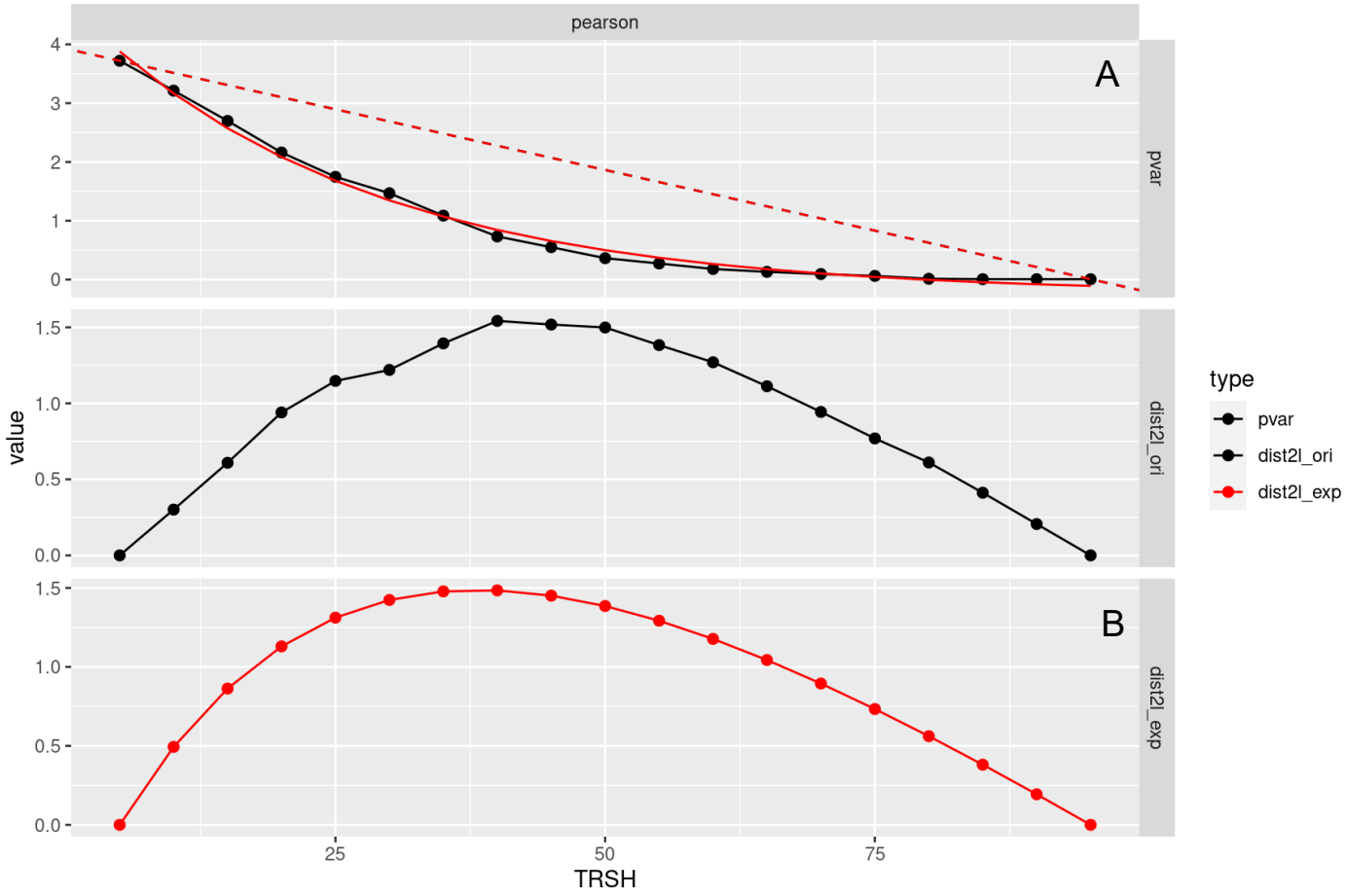
Supplementary Figure 4: 96-well plate layout for the oxidative metabolism-dependent energy expenditure (alamarBlue® assay).



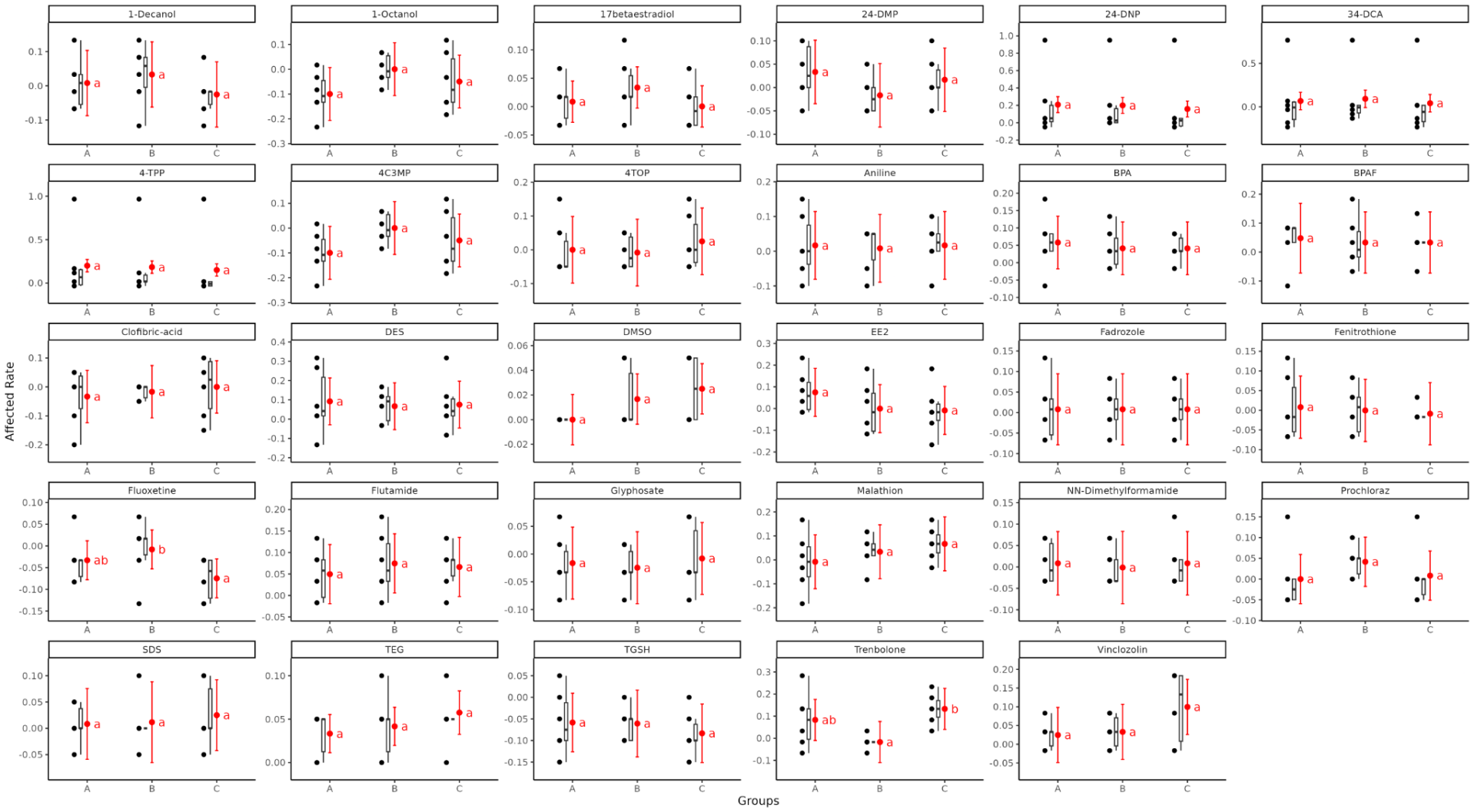
Supplementary Figure 5: A visual description of the similarity score transformation performed on the behavioural data. Each treated fish is compared to the 9 internal solvent control group fishes. The median similarity score from the 9 comparisons is the representative similarity score for that treated fish.



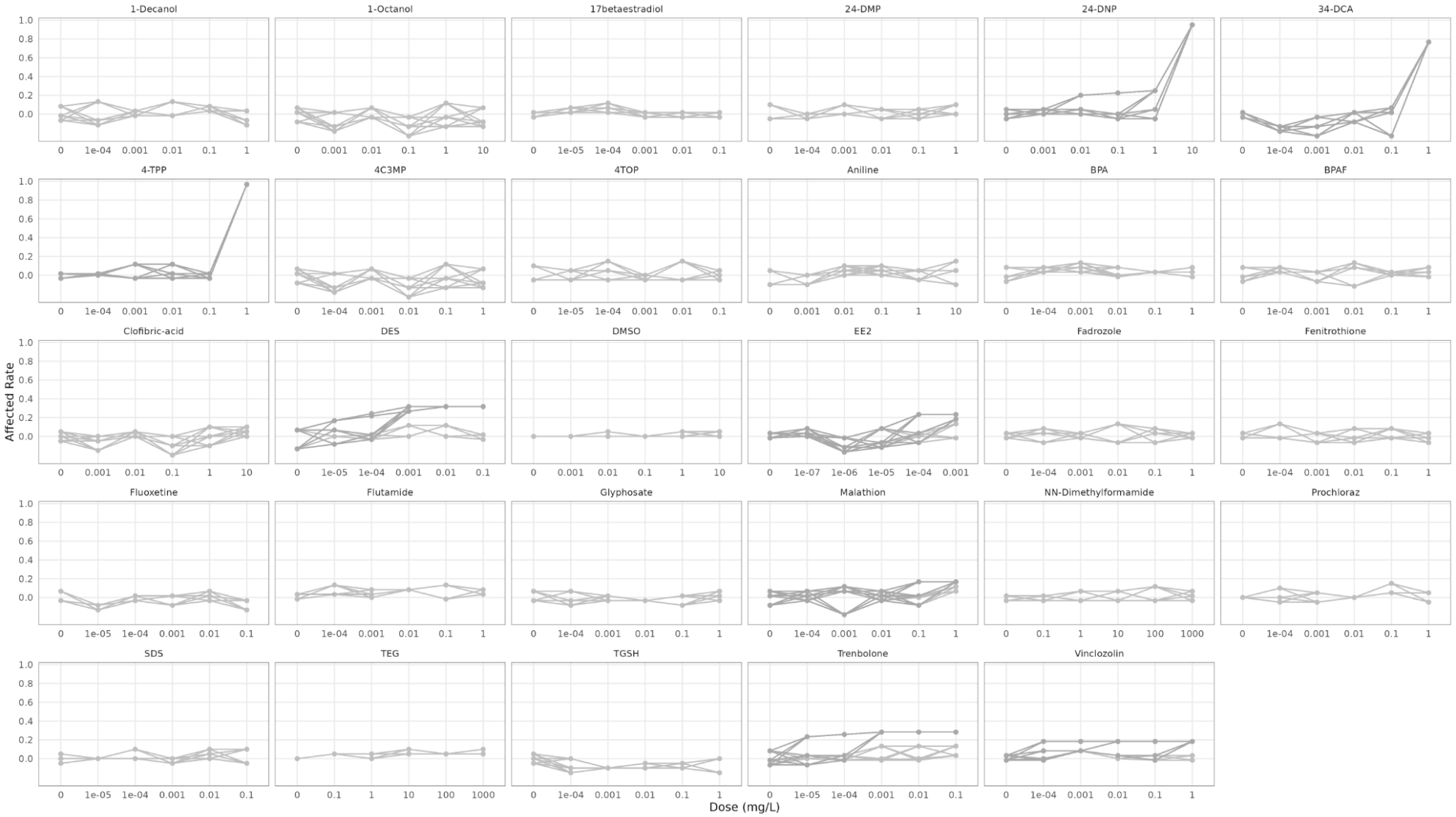
Supplementary Figure 6: An example of simulated dose-response curves with 10 samples.



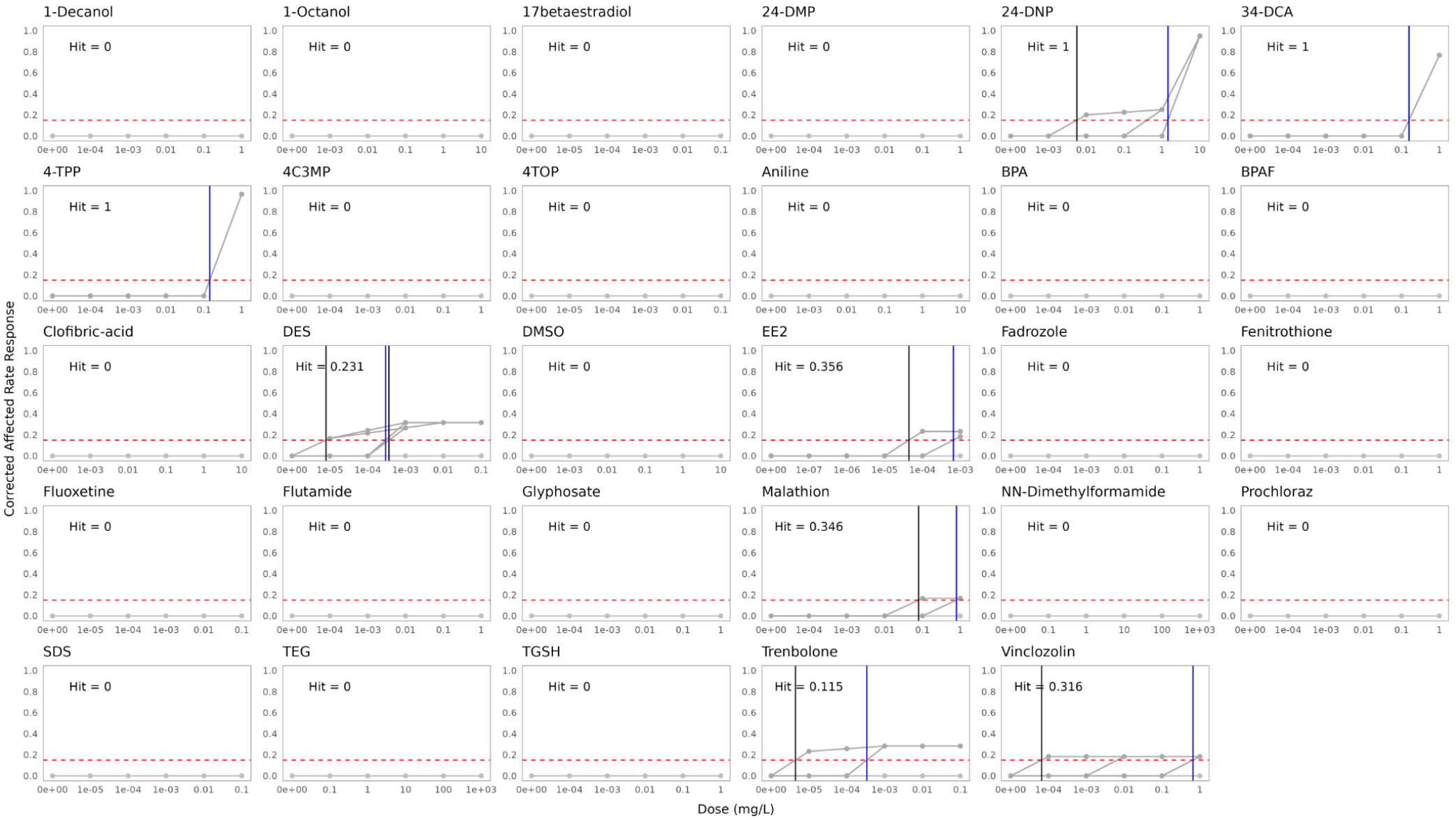
Supplementary Figure 7: The approximate benchmark threshold level was calculated using the 'Curvep' BMR calculation bootstrap. This figure shows an example BMR calculation from Pearson's similarity score behavioural endpoint. A line from the two points corresponding to the lowest to the highest noise threshold level was calculated (red dotted line in panel A). The distance from the red dotted line in panel A to the solid red exponential-like curve in panel A was calculated for each noise threshold (shown in panel B). The knee-point of the exponential-like curve was the noise threshold level with the largest distance to the line. The knee-point for the example figure is at a % noise threshold of 40%. Therefore, the approximate BMR for the data set (Pearson's r similarity score behavioural data) included in this figure is 40%.



Supplementary Figure 8: The subgrouping effect was analyzed using one-way ANOVAs, and Tukey's HSD post hoc test. This figure shows the results after an FDR adjustment. The results revealed no significant subgrouping effects in each experiment as measured by the overt toxicity/affected rate (percent affected) endpoint.

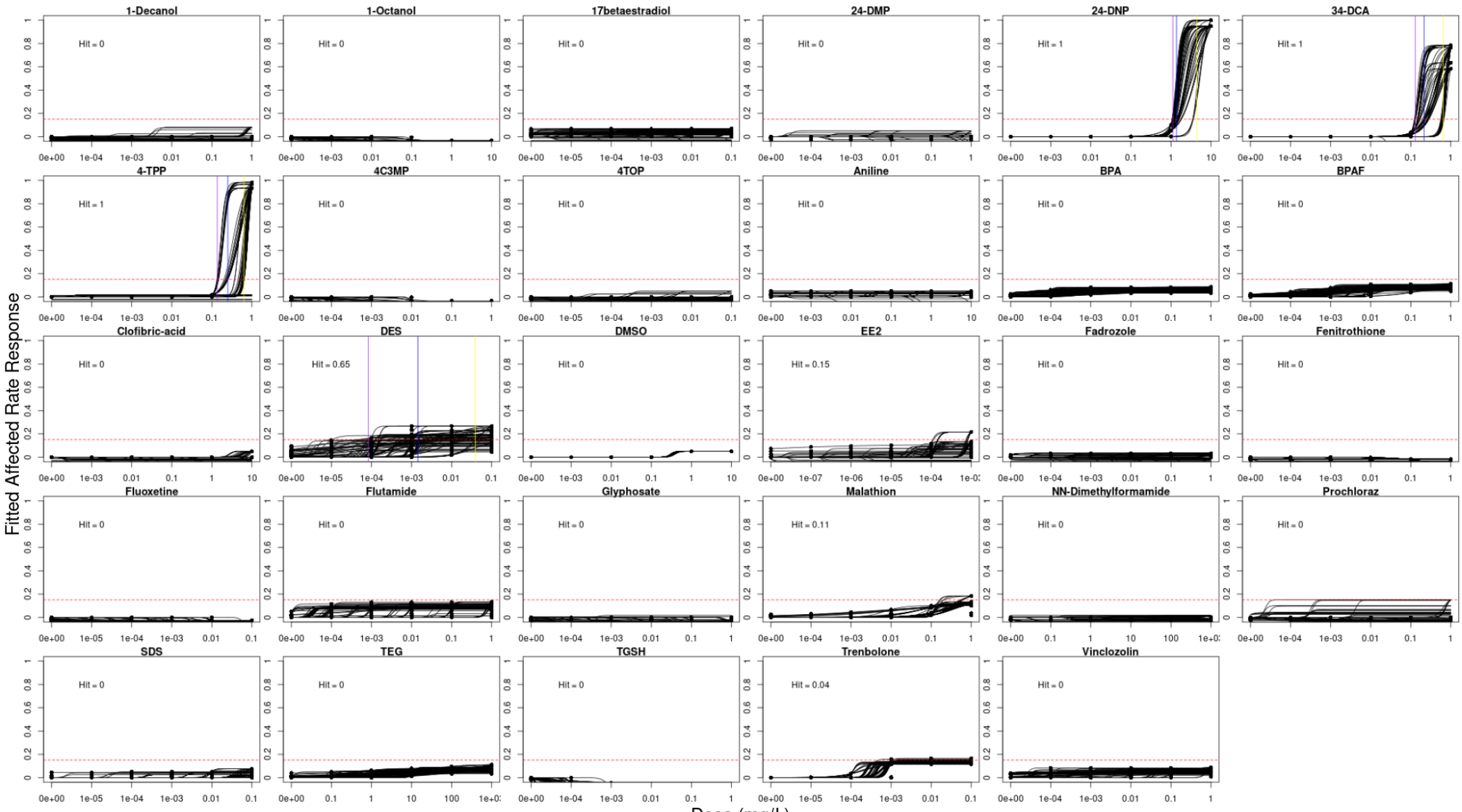


Supplementary Figure 9: 1000 dose-response curves of the percentage affected/affected rate response were simulated by sampling the overt toxicity data set with replacement. No curve processing or model fitting is conducted on these simulated dose-response curves.



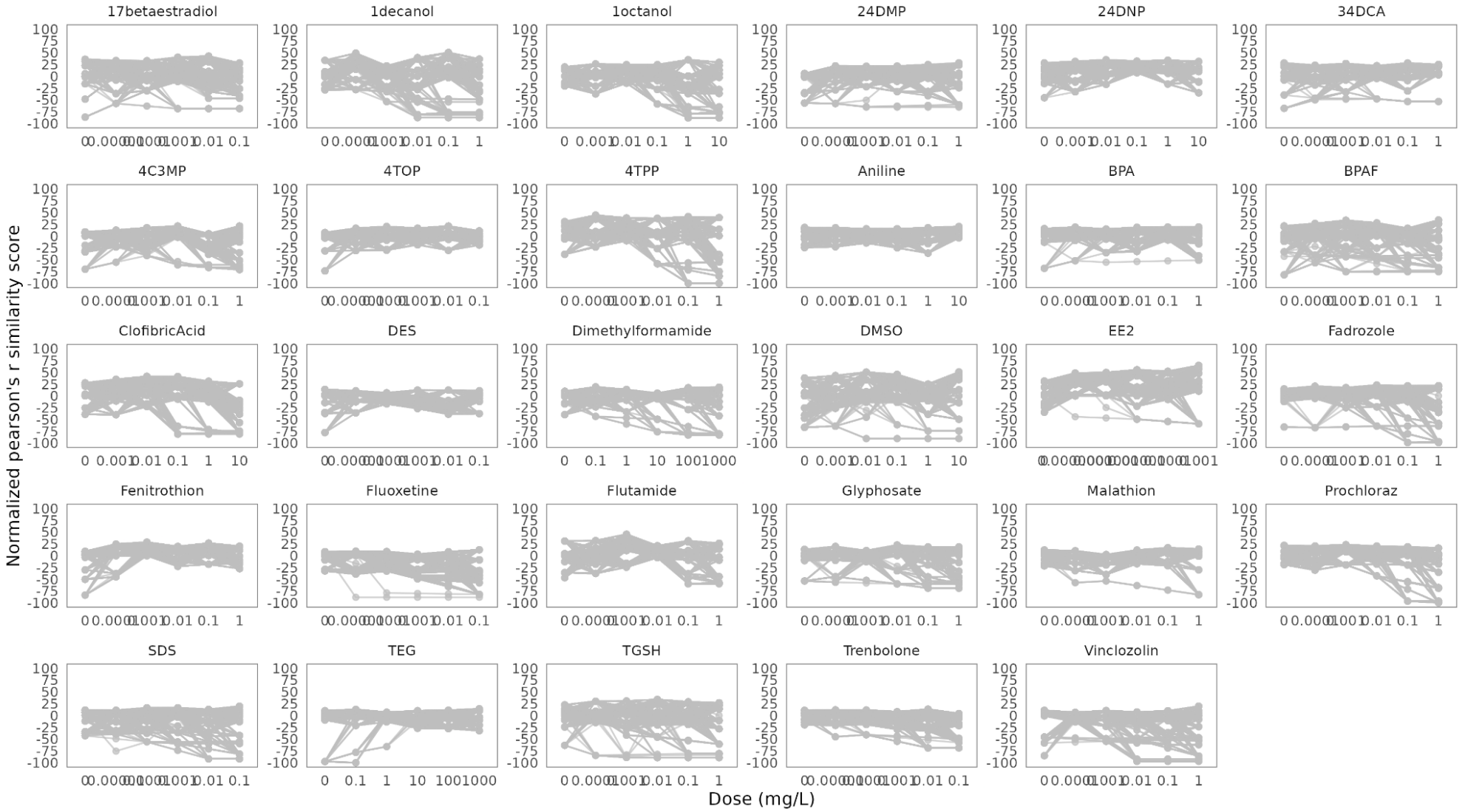
(title on the next page)

Supplementary Figure 10: The 1000 simulated dose-response curves from the overt toxicity data set (percent affected endpoint) in Supplementary Figure 9 were processed using the 'Curvep' algorithm. The 'Hit' rate defines the amount of processed dose-response curves that passed the BMR threshold level. The BMR threshold level is shown with a red dotted horizontal line. The BMR of the overt toxicity data set was 0.15 (15%). The blue vertical line highlights the BMD calculated by 'Curvep' in bootstrap. The BMD is the median of the dose-response curves that cross the BMR threshold. The gray vertical line to the left of the BMD is the BMD_L, the lower 95% confidence interval of the BMD. The gray vertical line to the right of the BMD is the BMD_U, the higher 95% confidence interval of the BMD. If the BMD_L and BMD_U are not visible in the figure, they are overlapped by the BMD, or there was no BMD.

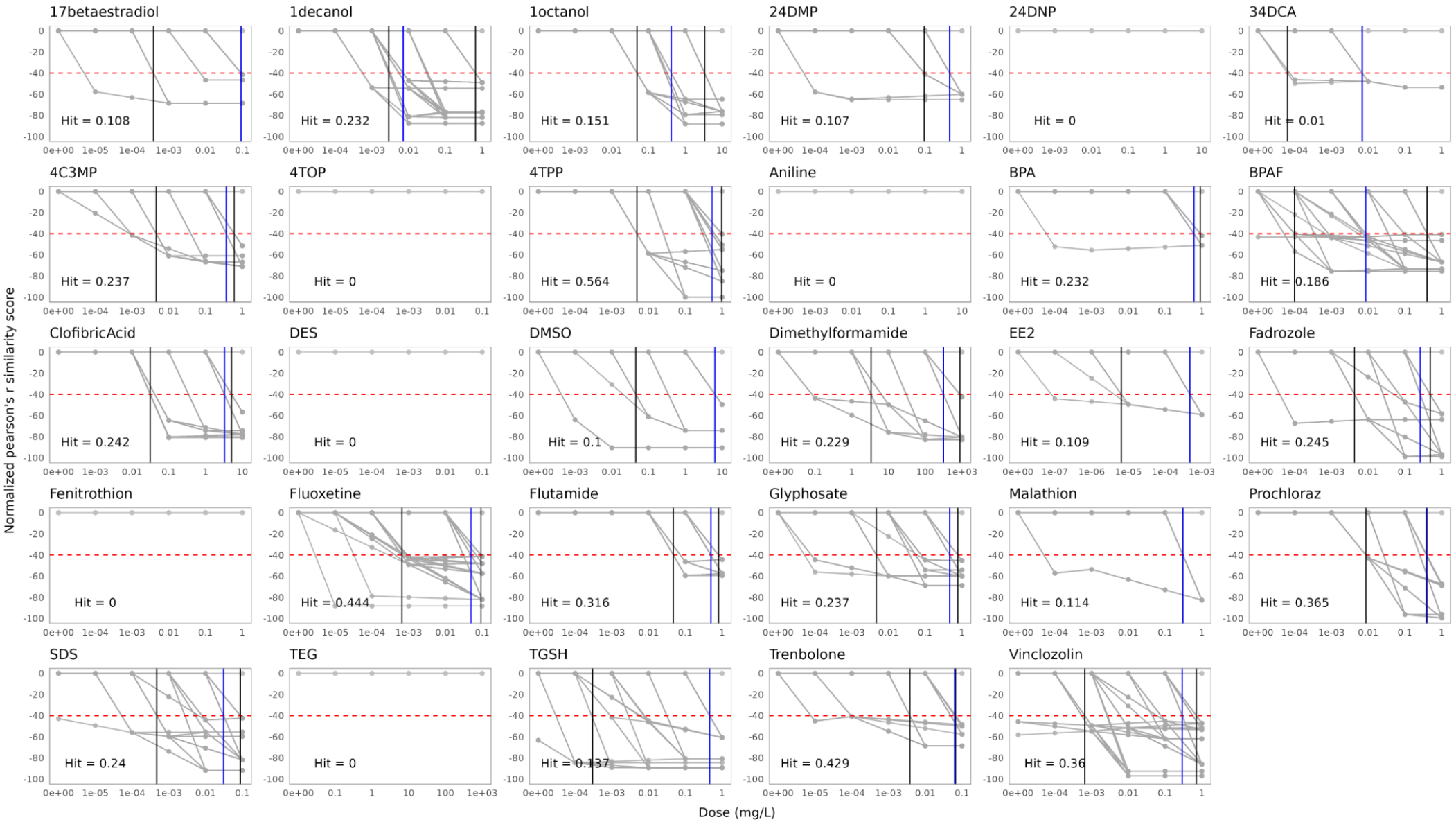


(title on next page)

Supplementary Figure 11: The 1000 simulated dose-response curves from the overt toxicity data set (percent affected endpoint) in Supplementary Figure 9 were fitted to the Hill equation model. The 'Hit' rate defines the amount of fitted dose-response curves that passed the BMR threshold level. The BMR threshold level is shown with the red dotted horizontal line The BMR of the overt toxicity data set was 0.15 (15%). The blue vertical line highlights the BMD calculated after fitting the simulated dose-response curves to the Hill equation model in bootstrap. The BMD is the median of the dose-response curves that cross the BMR threshold. The purple vertical line to the left of the BMD is BMD_L, the lower 95% confidence interval of the BMD. The vertical yellow line to the right of the BMD is the BMD_U, the higher 95% confidence interval of the BMD. If the BMD_L and BMD_U are not visible in the figure, they are overlapped by the BMD, or there was no BMD.

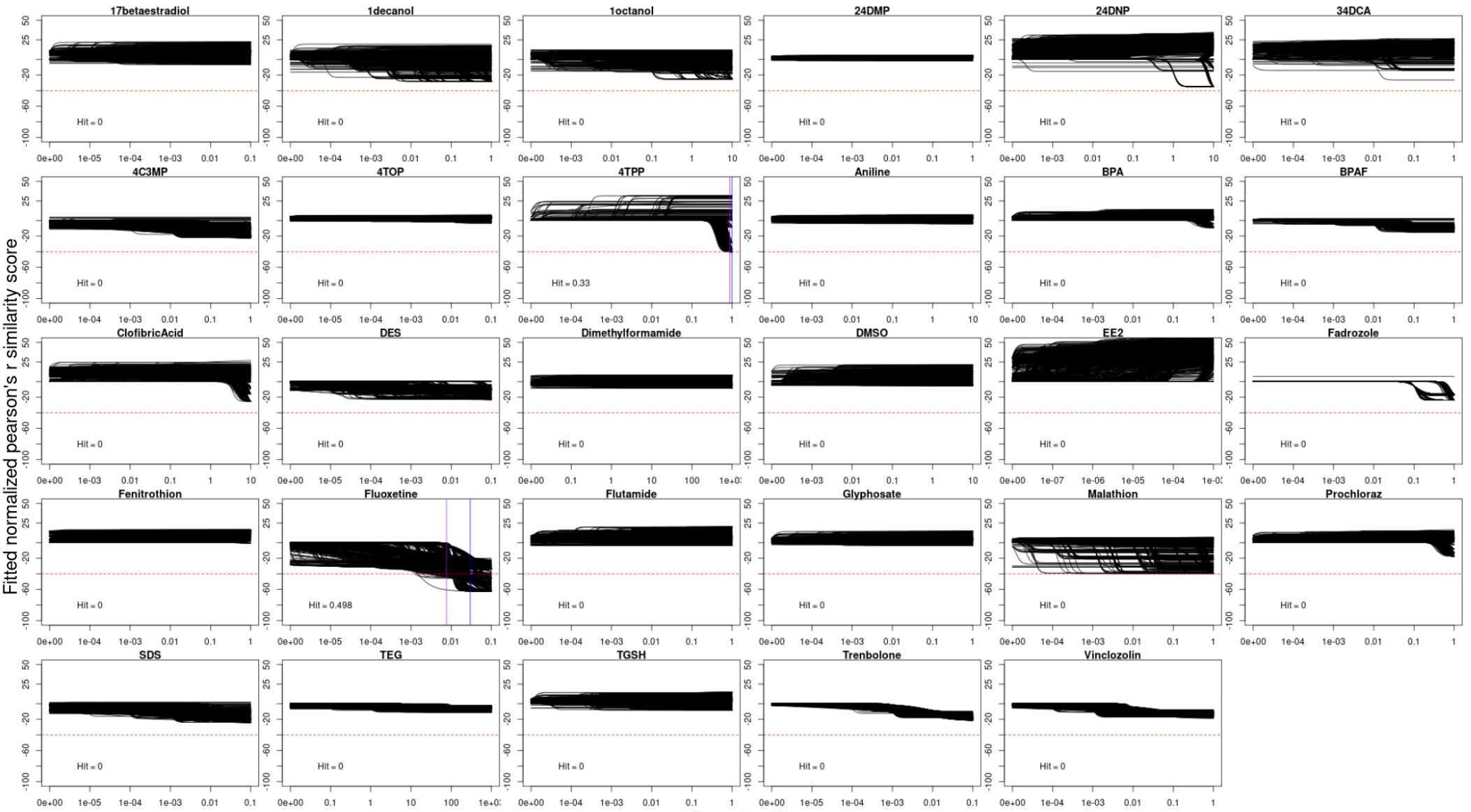


Supplementary Figure 12: 1000 dose-response curves of the transformed Pearsons's r similarity score light-dark swimming behaviour assay endpoint were simulated by sampling the transformed Pearson's r similarity score light-dark swimming behaviour data set with replacement. No curve processing or model fitting is conducted on these simulated dose-response curves.



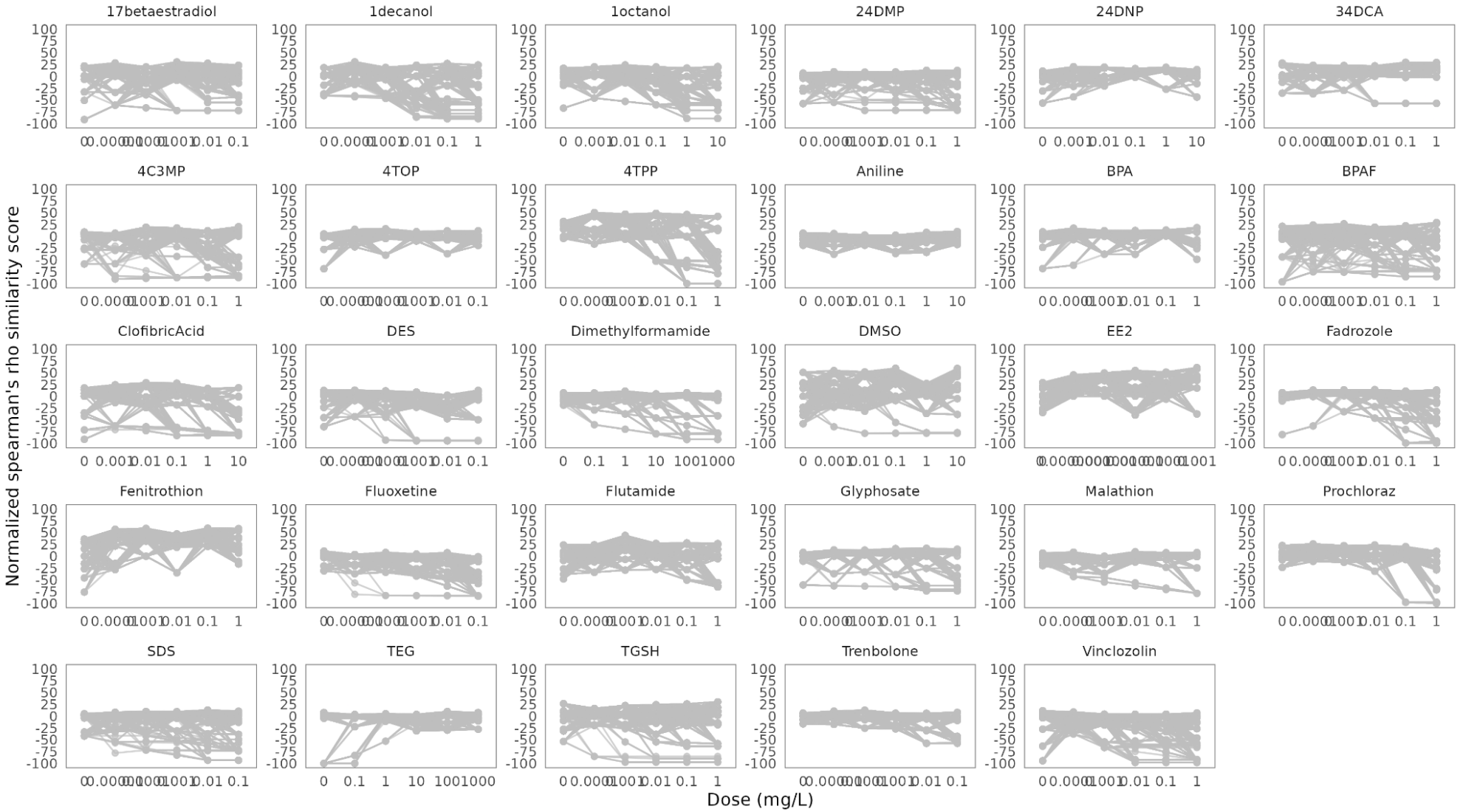
(title on next page)

Supplementary Figure 13: The 1000 simulated dose-response curves from the transformed Pearsons's r similarity score light-dark swimming behaviour endpoint data in Supplementary Figure 12 were processed using the 'Curvep' algorithm. The 'Hit' rate defines the amount of processed dose-response curves that passed the BMR threshold level. The BMR threshold level is shown with a red dotted horizontal line. The BMR of the overt toxicity data set was -40%. The blue vertical line highlights the BMD calculated by 'Curvep' in bootstrap. The BMD is the median of the dose-response curves that cross the BMR threshold. The gray vertical line to the left of the BMD is the $BMD_{L,}$ the lower 95% confidence interval of the BMD. The gray vertical line to the right of the BMD is the BMD_U, the higher 95% confidence interval of the BMD. If the BMD_L and BMD_U are not visible in the figure, they are overlapped by the BMD, or there was no BMD.

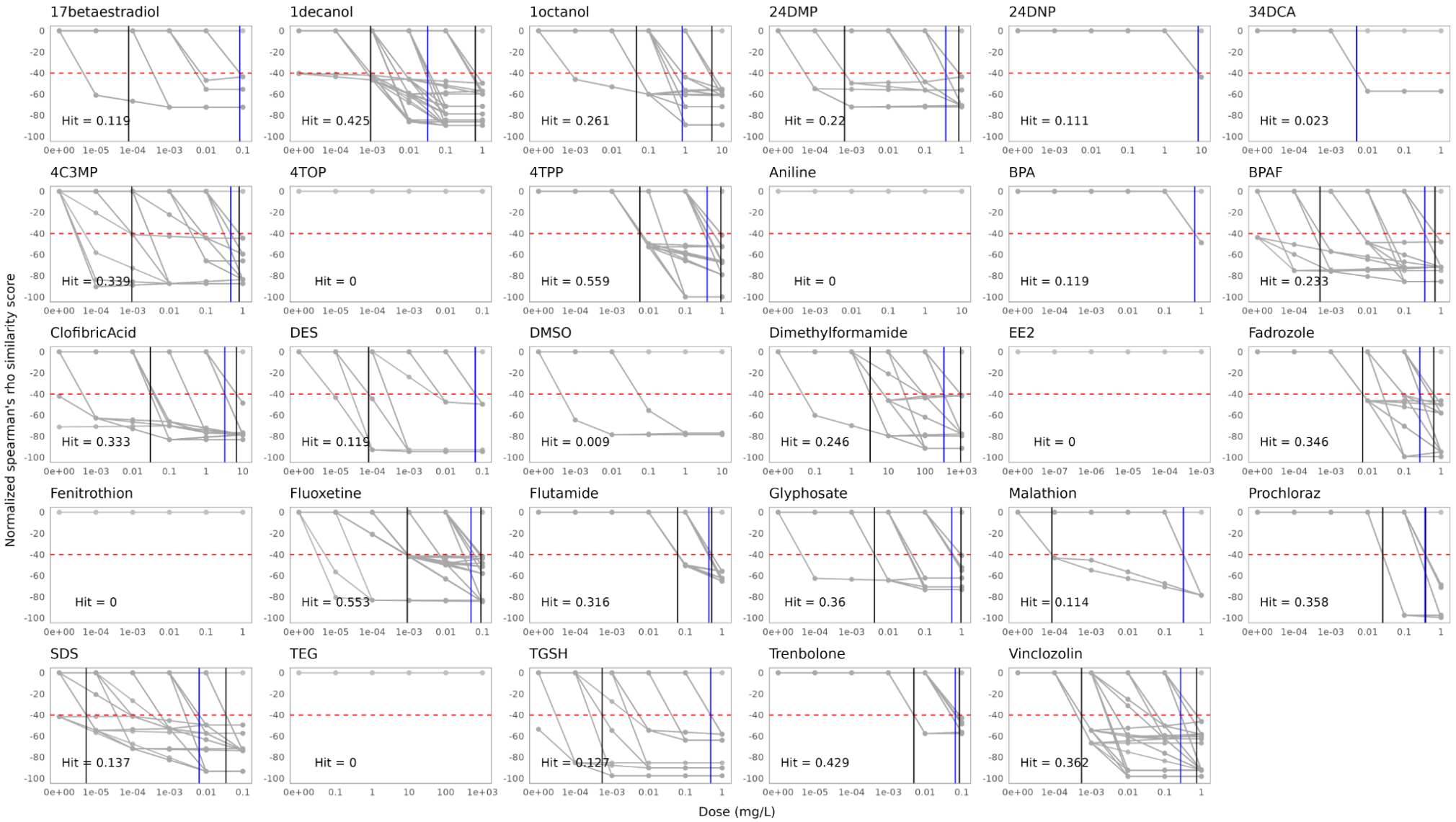


(title on next page)

Supplementary Figure 14: The 1000 simulated dose-response curves from the transformed Pearsons's r similarity score light-dark swimming behaviour endpoint data in Supplementary Figure 12 were fitted to the Hill equation model. The 'Hit' rate defines the amount of fitted dose-response curves that passed the BMR threshold level. The BMR threshold level is shown with the red dotted horizontal line The BMR of the overt toxicity data set was -40%. The blue vertical line highlights the BMD calculated after fitting the simulated dose-response curves to the Hill equation model in bootstrap. The BMD is the median of the dose-response curves that cross the BMR threshold. The purple vertical line to the left of the BMD is BMD_L, the lower 95% confidence interval of the BMD. The vertical yellow line to the right of the BMD is the BMD_U, the higher 95% confidence interval of the BMD. If the BMD_L and BMD_U are not visible in the figure, they are overlapped by the BMD, or there was no BMD.

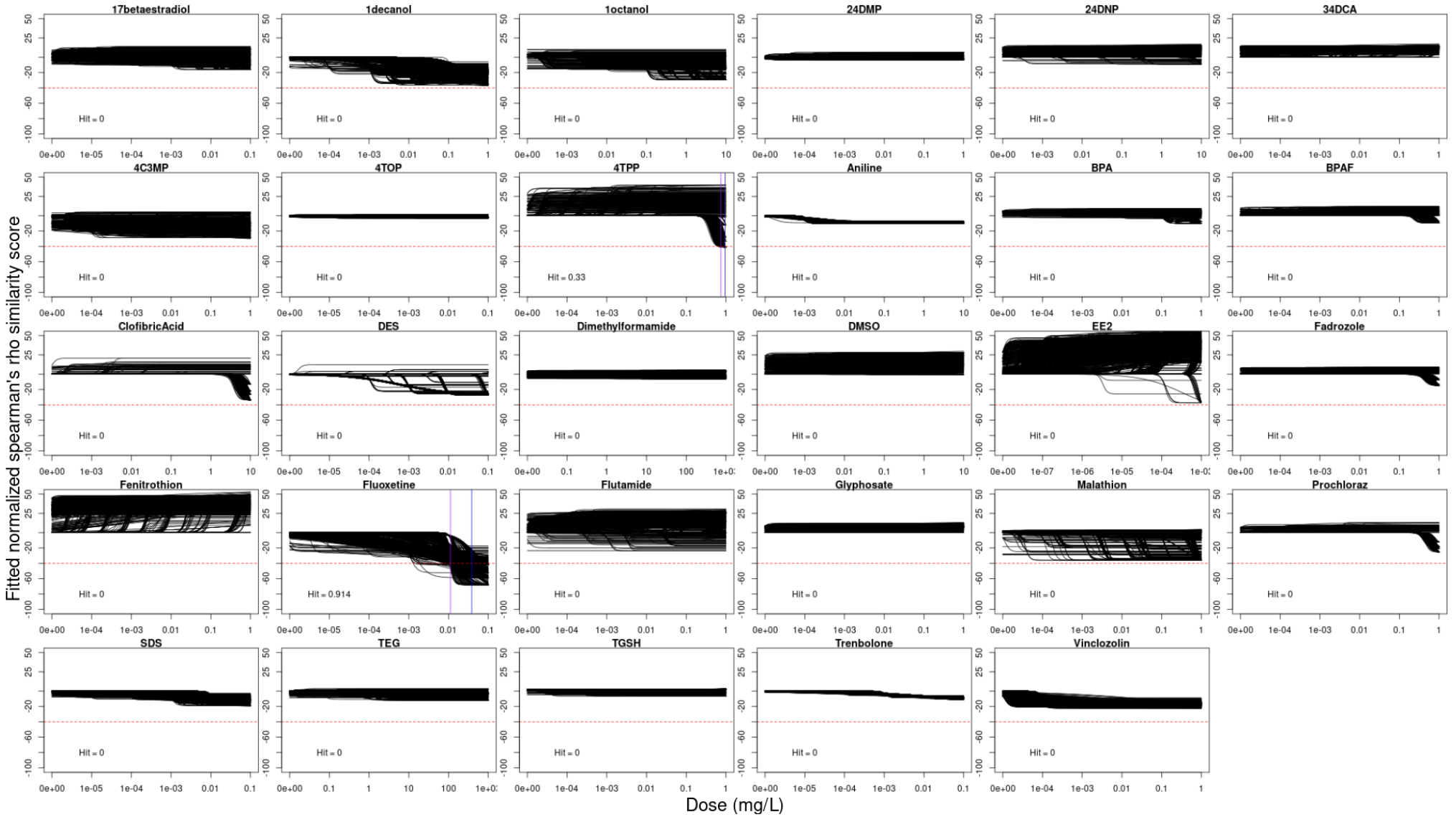


Supplementary Figure 15: 1000 dose-response curves of the transformed Spearman's rho similarity score light-dark swimming behaviour endpoint data were simulated by sampling the transformed Spearman's rho score light-dark swimming behaviour data set with replacement. No curve processing or model fitting is conducted on these simulated dose-response curves.



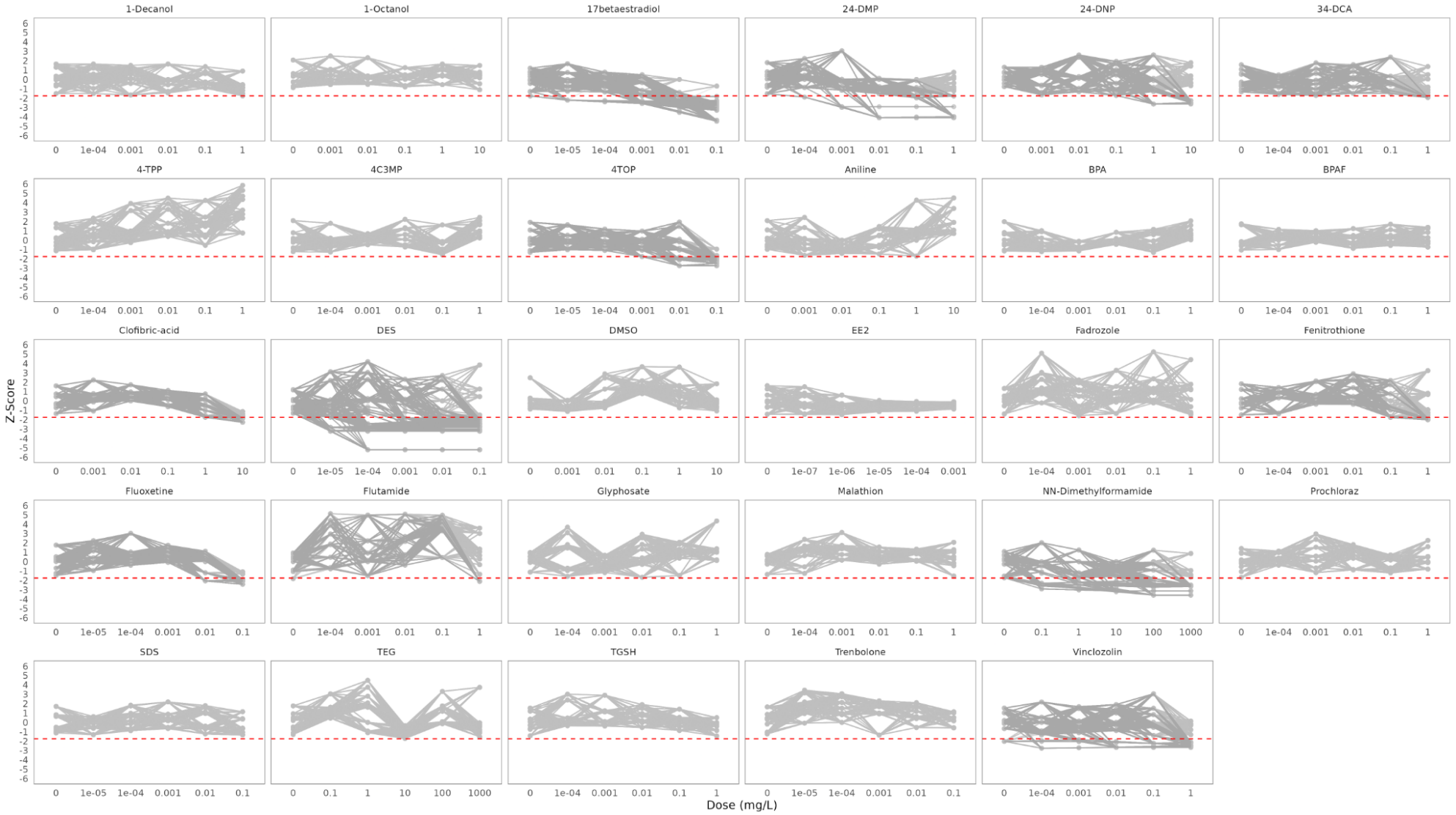
(title on the next page)

Supplementary Figure 16: The 1000 simulated dose-response curves from the transformed Spearman's rho similarity score light-dark swimming behaviour endpoint data in Supplementary Figure 15 were processed using the 'Curvep' algorithm. The 'Hit' rate defines the amount of processed dose-response curves that passed the BMR threshold level. The BMR threshold level is shown with a red dotted horizontal line. The BMR of the overt toxicity data set was -40%. The blue vertical line highlights the BMD calculated by 'Curvep' in bootstrap. The BMD is the median of the dose-response curves that cross the BMR threshold. The gray vertical line to the left of the BMD is the $BMD_{L,}$ the lower 95% confidence interval of the BMD. The gray vertical line to the right of the BMD is the BMD_U, the higher 95% confidence interval of the BMD. If the BMD_L and BMD_U are not visible in the figure, they are overlapped by the BMD, or there was no BMD.

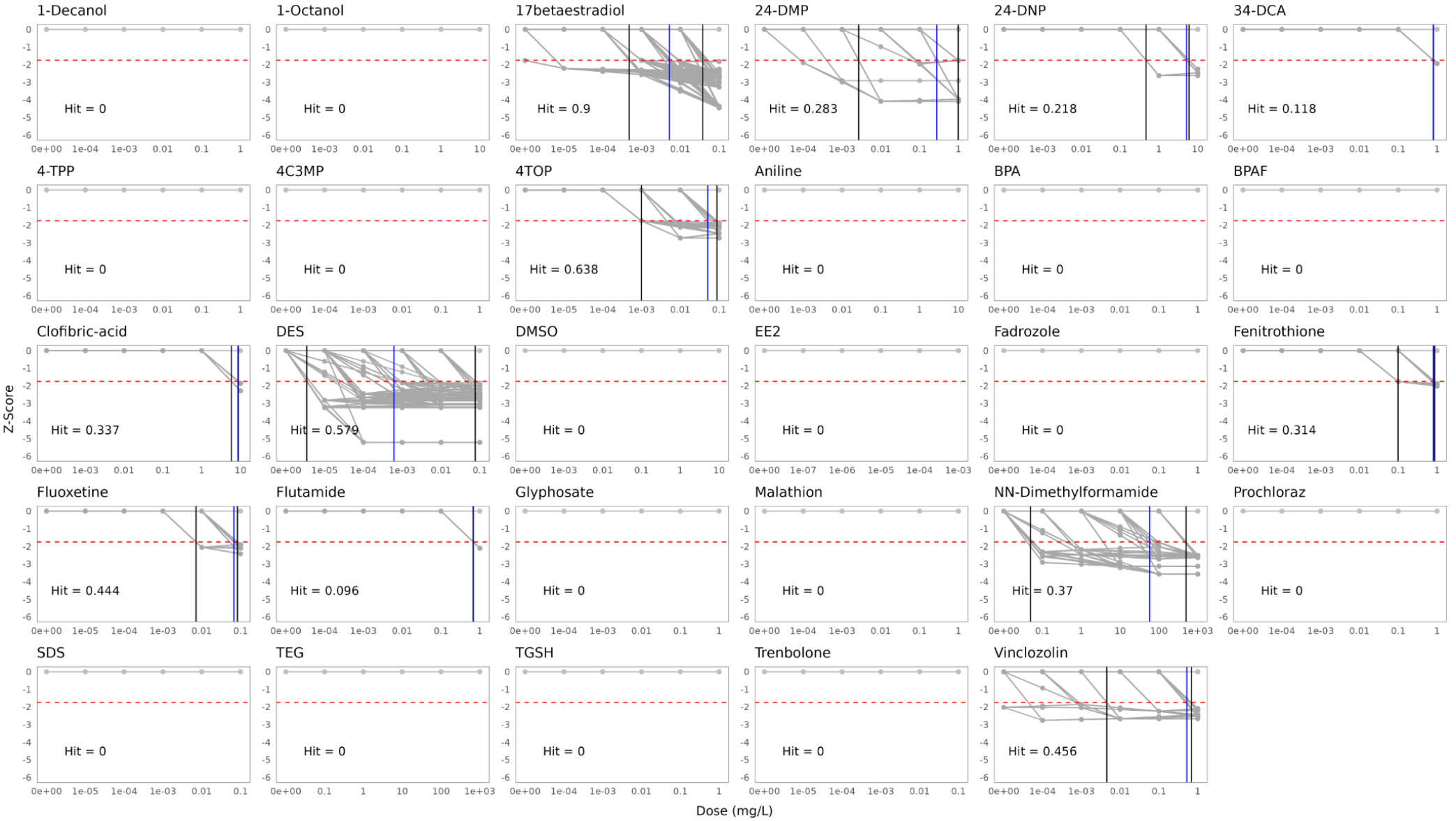


(title on the next page)

Supplementary Figure 17: The 1000 simulated dose-response curves from the transformed Spearman's rho similarity score light-dark swimming behaviour endpoint data in Supplementary Figure 15 were fitted to the Hill equation model. The 'Hit' rate defines the amount of fitted dose-response curves that passed the BMR threshold level. The BMR threshold level is shown with the red dotted horizontal line The BMR of the overt toxicity data set was -40%. The blue vertical line highlights the BMD calculated after fitting the simulated dose-response curves to the Hill equation model in bootstrap. The BMD is the median of the dose-response curves that cross the BMR threshold. The purple vertical line to the left of the BMD is BMD_L, the lower 95% confidence interval of the BMD. The vertical yellow line to the right of the BMD is the BMD_U, the higher 95% confidence interval of the BMD. If the BMD_L and BMD_U are not visible in the figure, they are overlapped by the BMD, or there was no BMD.

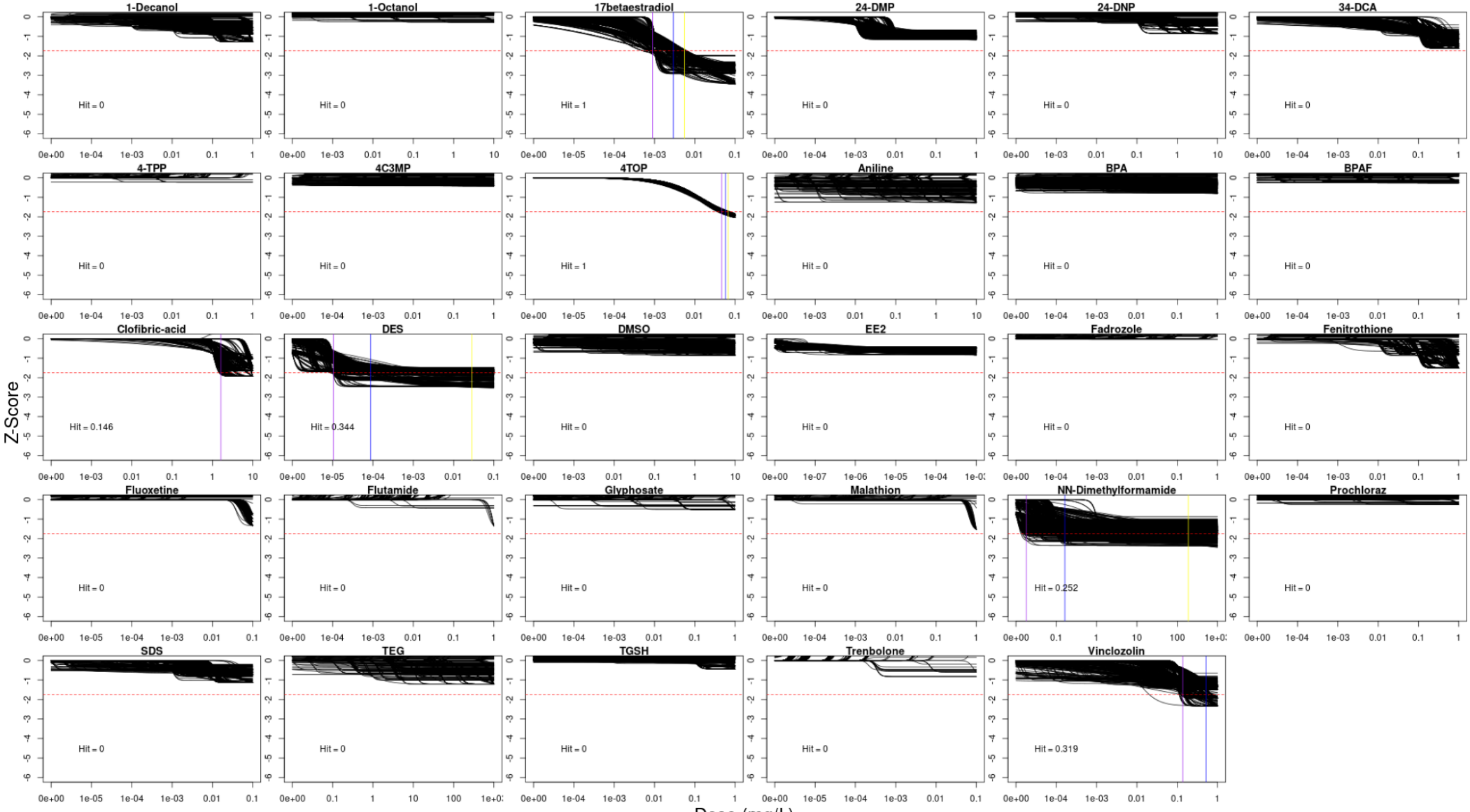


Supplementary Figure 18: 1000 dose-response curves from the oxidative metabolism-dependent energy expenditure data set were simulated by sampling oxidative metabolism-dependent energy expenditure data with replacement. No curve processing or model fitting is conducted on these simulated dose-response curves. The red dotted horizontal line is the BMR threshold of -1.75.



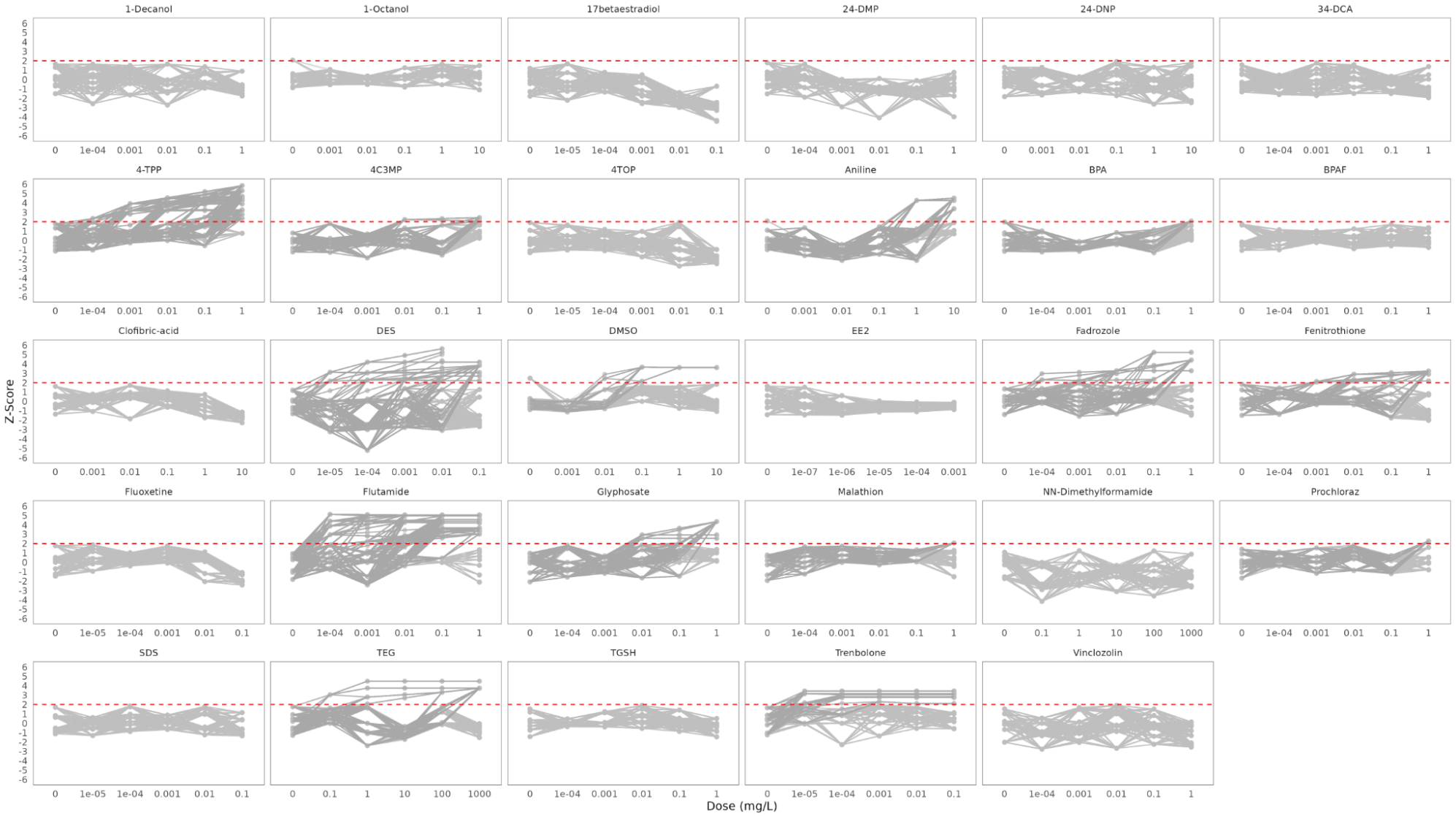
(title on the next page)

Supplementary Figure 19: The 1000 simulated dose-response curves from the oxidative metabolism-dependent energy expenditure data in Supplementary Figure 18 were processed using the 'Curvep' algorithm. The 'Hit' rate defines the amount of processed dose-response curves that passed the BMR threshold level. The BMR threshold level is shown with a red dotted horizontal line. The BMR of the overt toxicity data set was -1.75% (~-30%). The blue vertical line highlights the BMD calculated by 'Curvep' in bootstrap. The BMD is the median of the dose-response curves that cross the BMR threshold. The gray vertical line to the left of the BMD is the $BMD_{L,}$ the lower 95% confidence interval of the BMD. The gray vertical line to the right of the BMD is the BMD_U, the higher 95% confidence interval of the BMD. If the BMD_L and BMD_U are not visible in the figure, they are overlapped by the BMD, or there was no BMD.

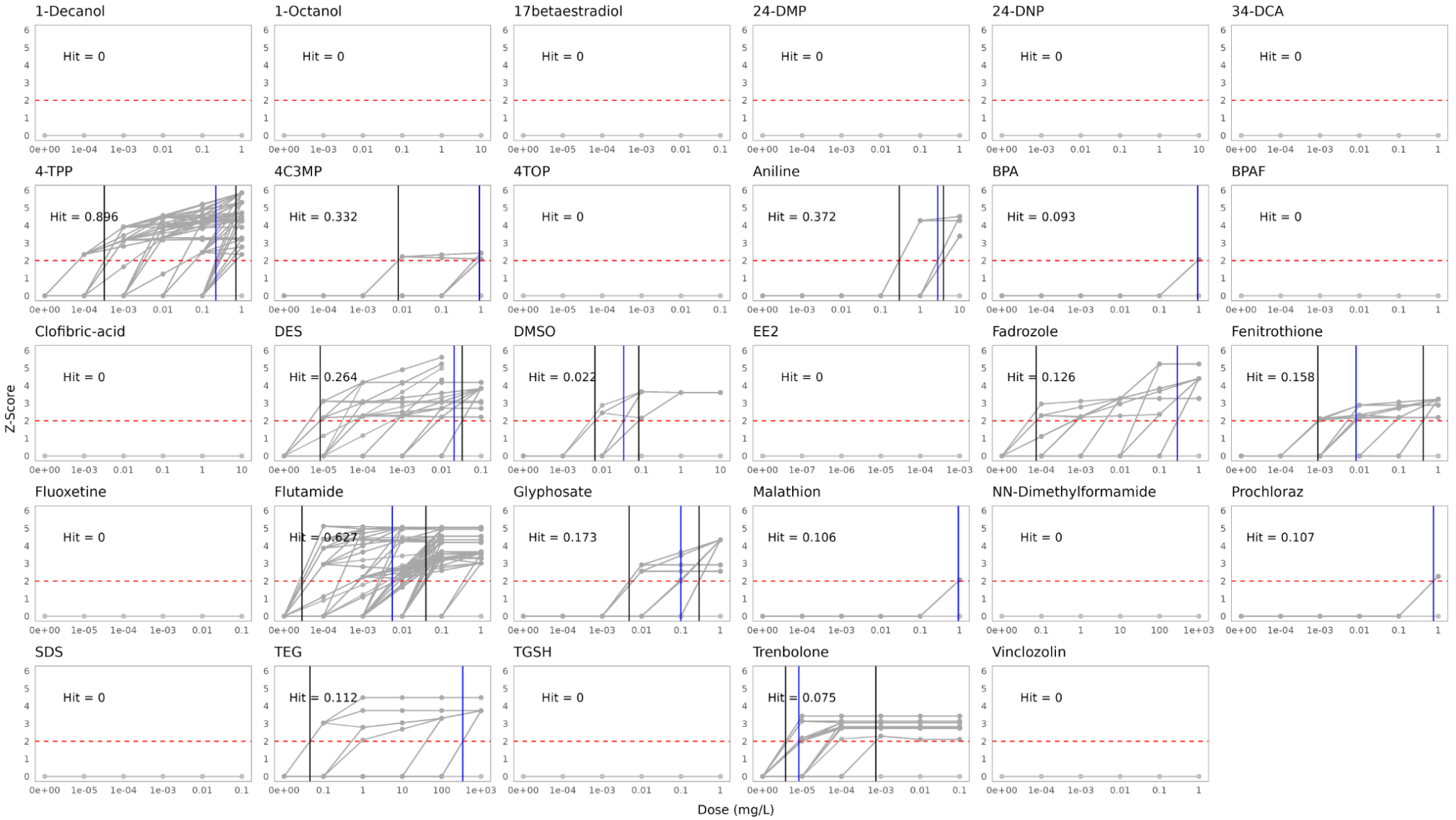


(title on the next page)

Supplementary Figure 20: The 1000 simulated dose-response curves from the oxidative metabolism-dependent energy expenditure data in Supplementary Figure 18 were fitted to the Hill equation model. The 'Hit' rate defines the amount of fitted dose-response curves that passed the BMR threshold level. The BMR threshold level is shown with the red dotted horizontal line The BMR of the overt toxicity data set was -1.75 (~-30%). The blue vertical line highlights the BMD calculated after fitting the simulated dose-response curves to the Hill equation model in bootstrap. The BMD is the median of the dose-response curves that cross the BMR threshold. The purple vertical line to the left of the BMD is BMD_L, the lower 95% confidence interval of the BMD. The vertical yellow line to the right of the BMD is the BMD_U, the higher 95% confidence interval of the BMD. If the BMD_L and BMD_U are not visible in the figure, they are overlapped by the BMD, or there was no BMD.

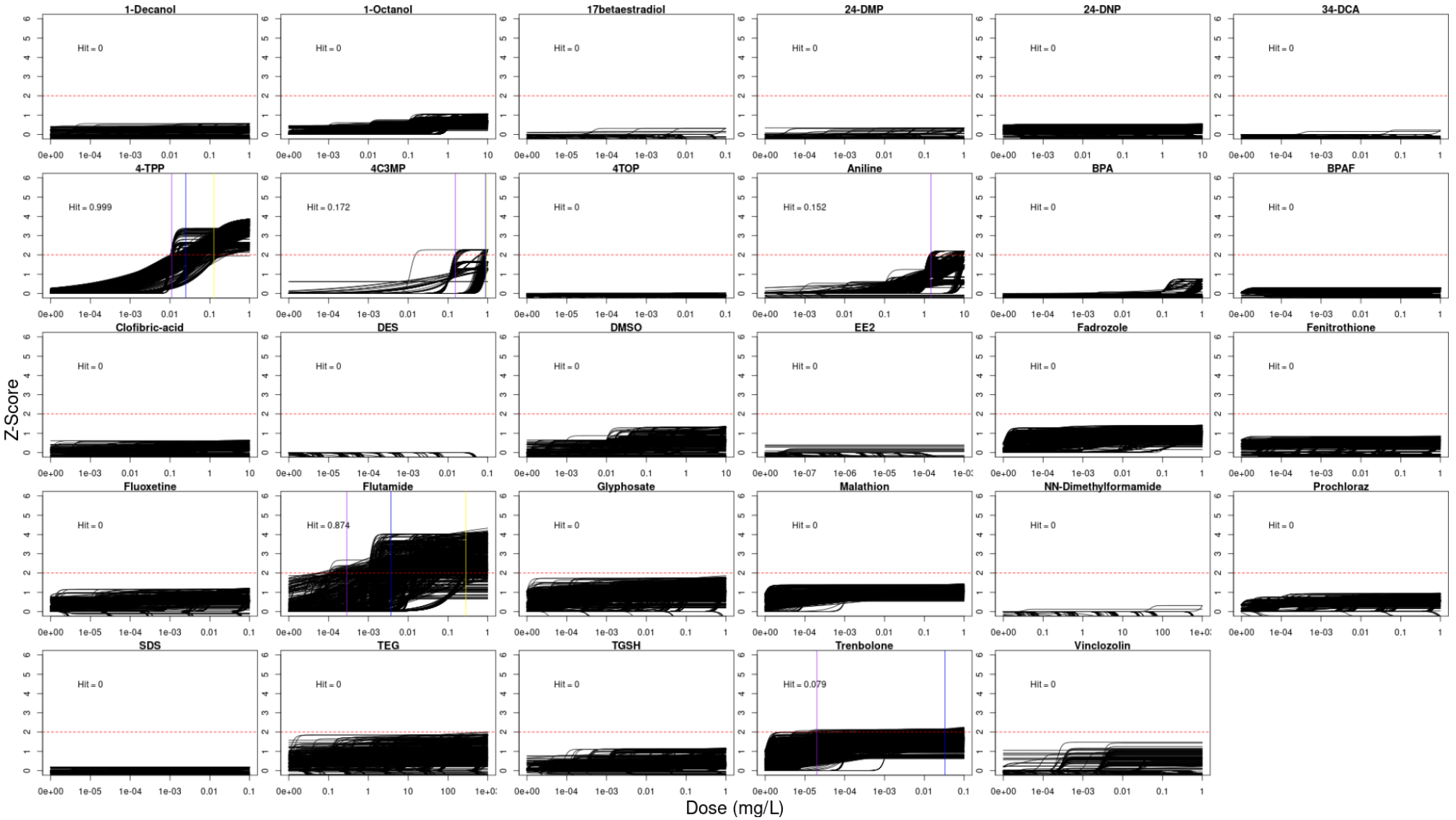


Supplementary Figure 21: 1000 dose-response curves from the oxidative metabolism-dependent energy expenditure data set were simulated by sampling oxidative metabolism-dependent energy expenditure data with replacement. No curve processing or model fitting is conducted on these simulated dose-response curves. The red dotted horizontal line is the BMR threshold of 2.



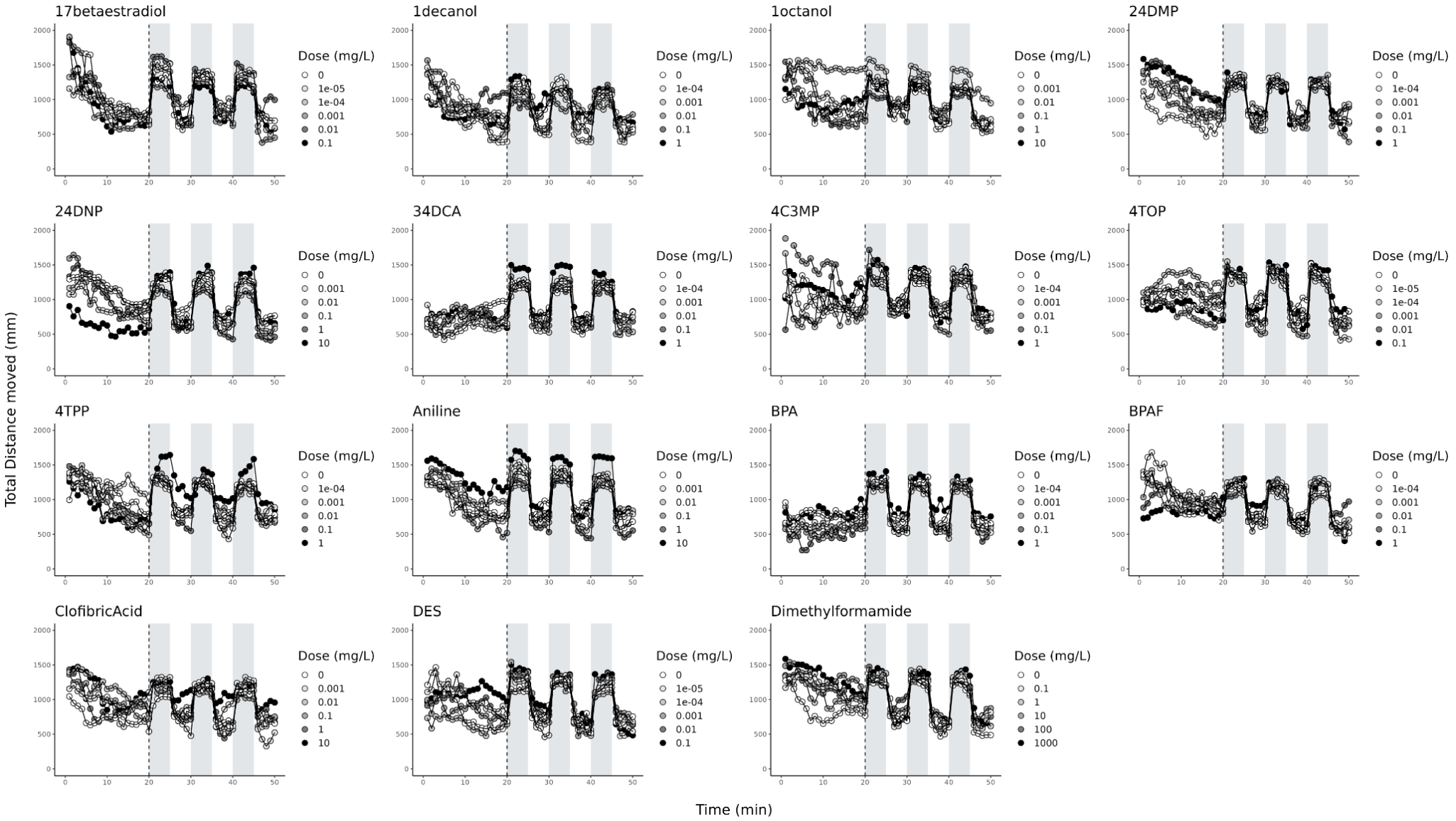
(title on the next page)

Supplementary Figure 22: The 1000 simulated dose-response curves from the oxidative metabolism-dependent energy expenditure data in Supplementary Figure 21 were processed using the 'Curvep' algorithm. The 'Hit' rate defines the amount of processed dose-response curves that passed the BMR threshold level. The BMR threshold level is shown with a red dotted horizontal line. The BMR of the overt toxicity data set was 2 (33%). The blue vertical line highlights the BMD calculated by 'Curvep' in bootstrap. The BMD is the median of the dose-response curves that cross the BMR threshold. The gray vertical line to the left of the BMD is the $BMD_{L,}$ the lower 95% confidence interval of the BMD. The gray vertical line to the right of the BMD is the BMD_U, the higher 95% confidence interval of the BMD. If the BMD_L and BMD_U are not visible in the figure, they are overlapped by the BMD, or there was no BMD.

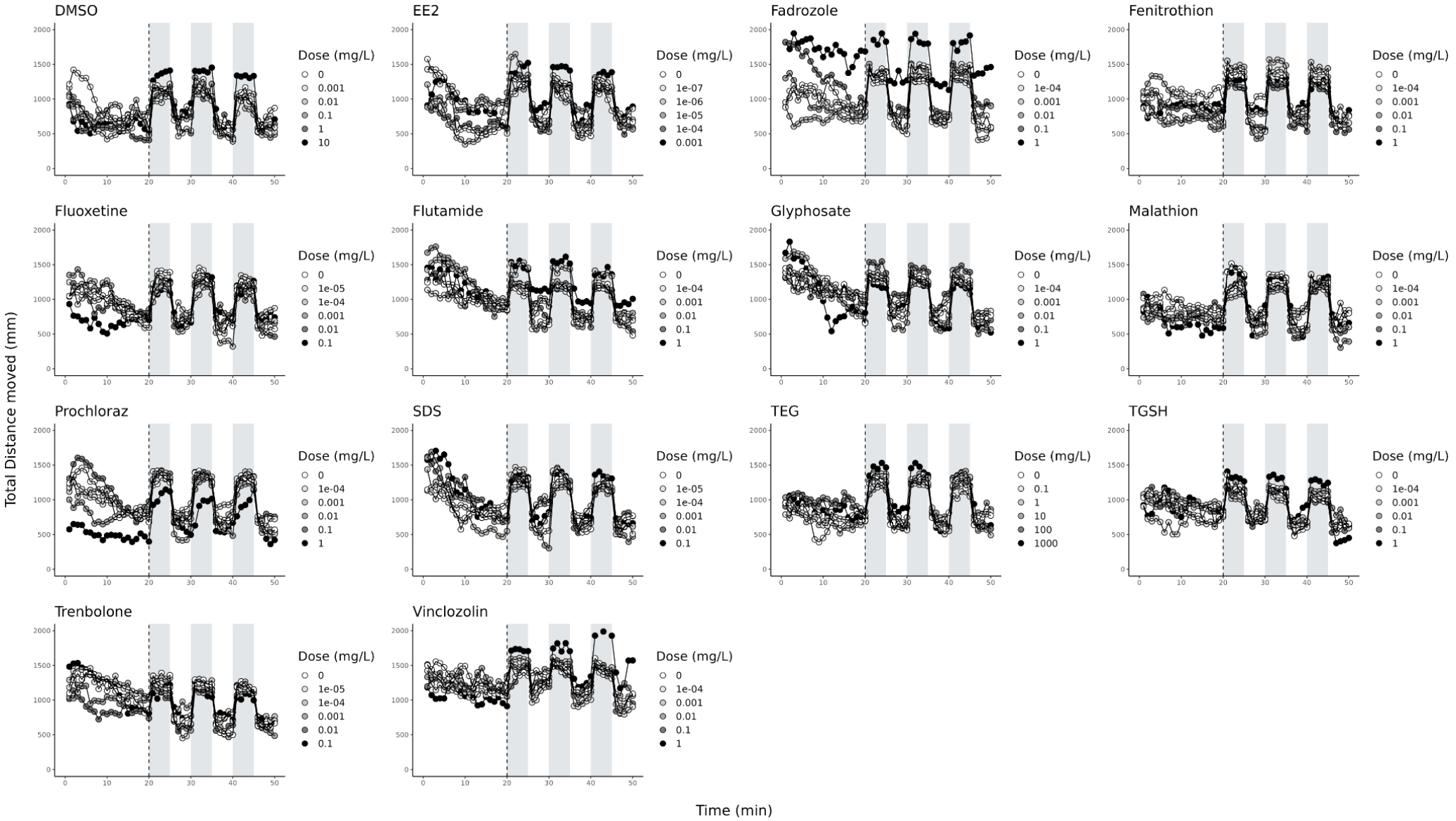


(title on the next page)

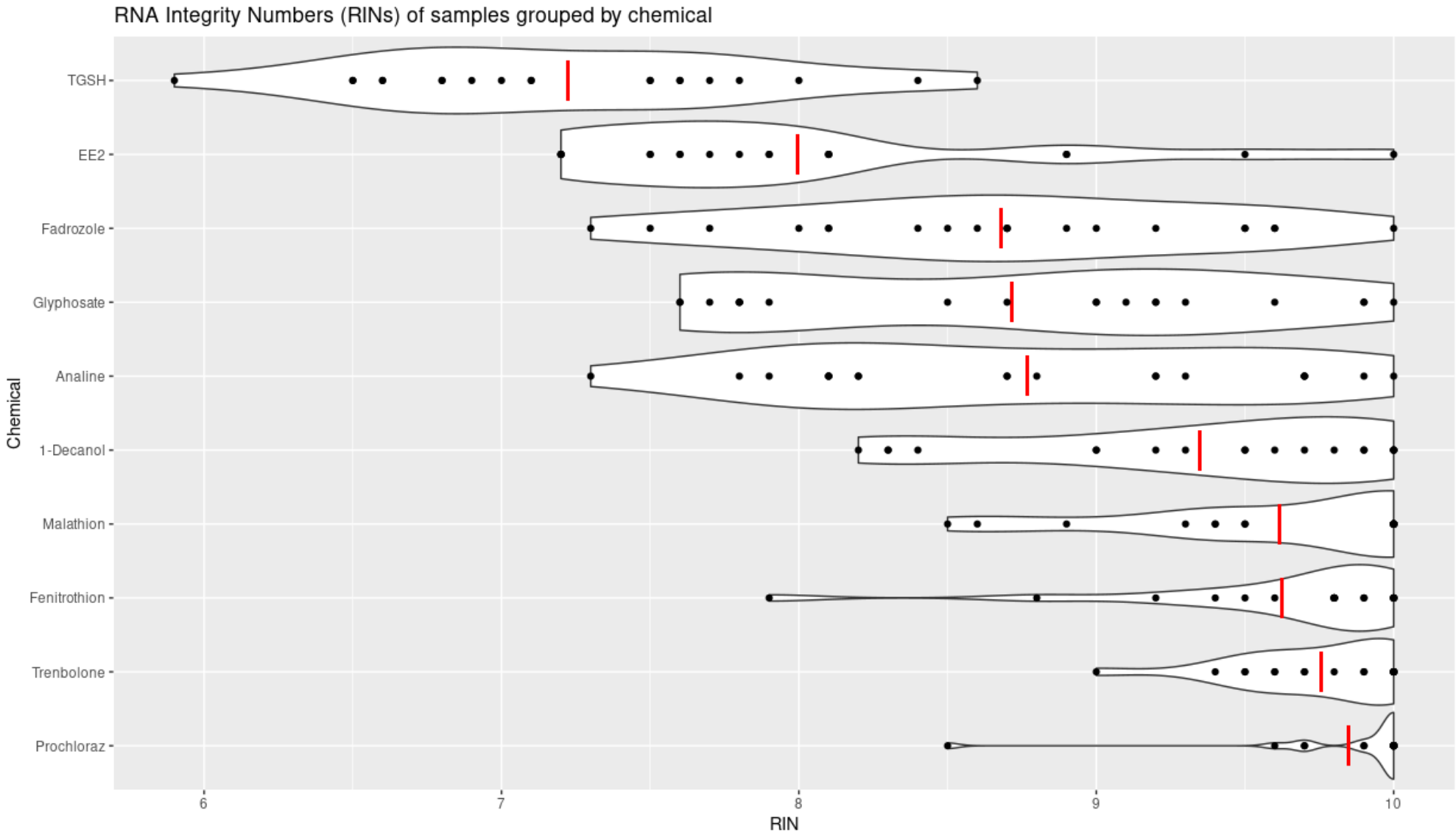
Supplementary Figure 23: The 1000 simulated dose-response curves from the oxidative metabolism-dependent energy expenditure data in Supplementary Figure 21 were fitted to the Hill equation model. The 'Hit' rate defines the amount of fitted dose-response curves that passed the BMR threshold level. The BMR threshold level is shown with the red dotted horizontal line The BMR of the overt toxicity data set was 2 (33%). The blue vertical line highlights the BMD calculated after fitting the simulated dose-response curves to the Hill equation model in bootstrap. The BMD is the median of the dose-response curves that cross the BMR threshold. The purple vertical line to the left of the BMD is BMD_L, the lower 95% confidence interval of the BMD. The vertical yellow line to the right of the BMD is the BMD_U, the higher 95% confidence interval of the BMD. If the BMD_L and BMD_U are not visible in the figure, they are overlapped by the BMD, or there was no BMD.



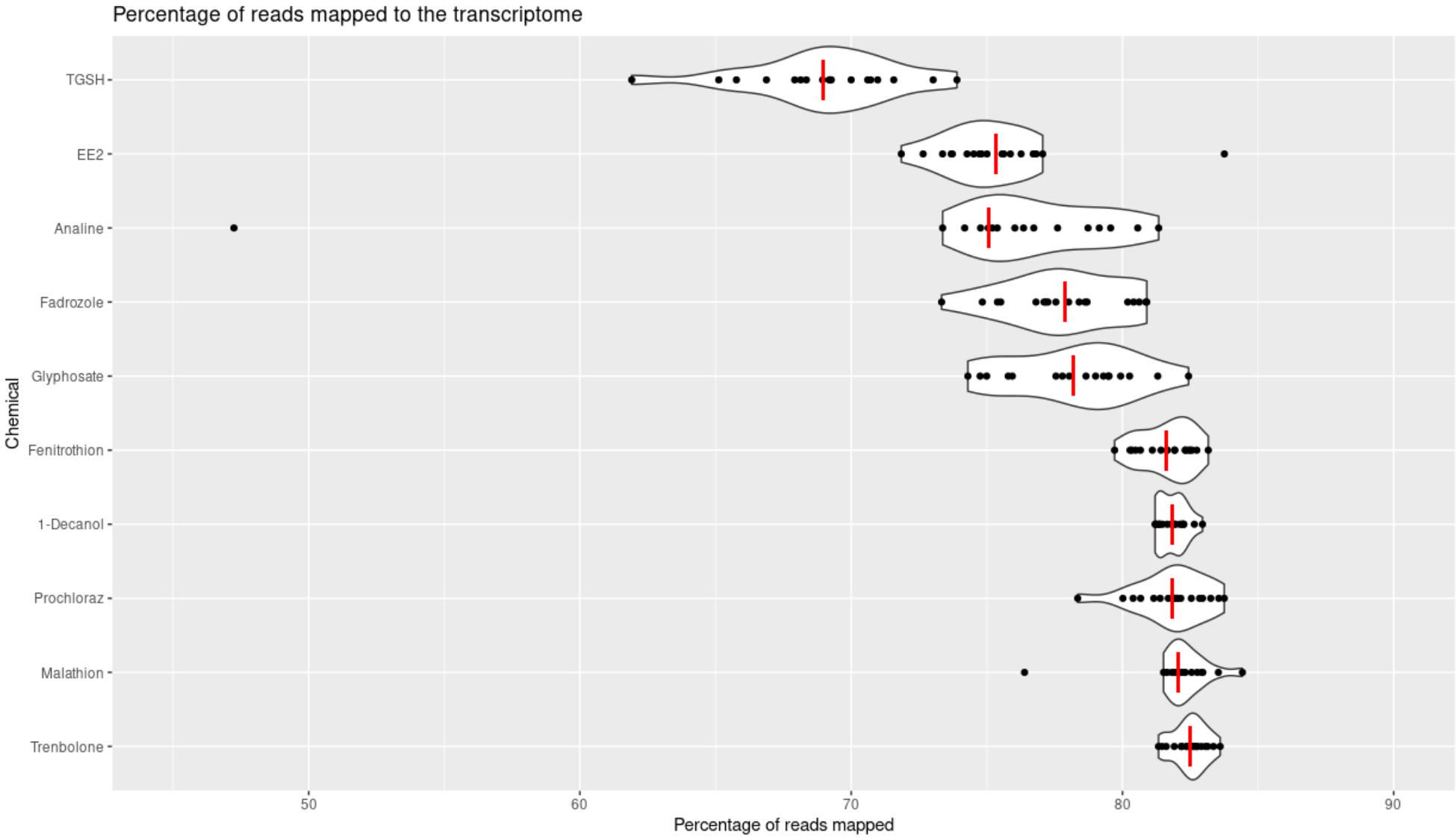
Supplementary Figure 24: The light-dark swimming behaviour assay is a 50-minute assay comprising a 20-minute acclimation period and 3 subsequent alternating 5-minute-long light and dark cycles. The total distance moved in mm endpoint from each experiment was measured and is shown in this figure.



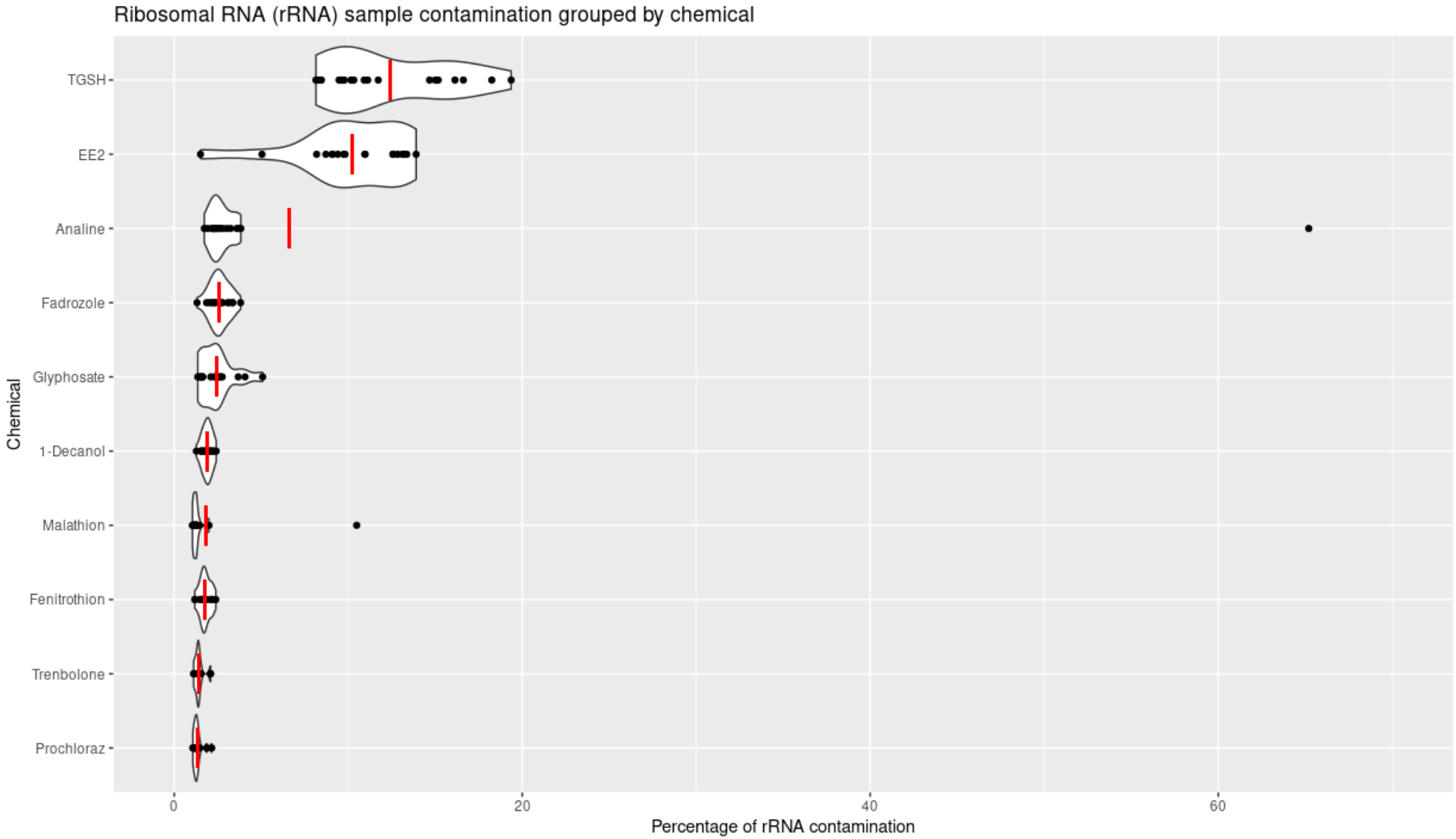
Supplementary Figure 24 (contd.): The light-dark swimming behaviour assay is a 50-minute assay comprising a 20-minute acclimation period and 3 subsequent alternating 5-minute-long light and dark cycles. The total distance moved in mm endpoint from each experiment was measured and is shown in this figure.



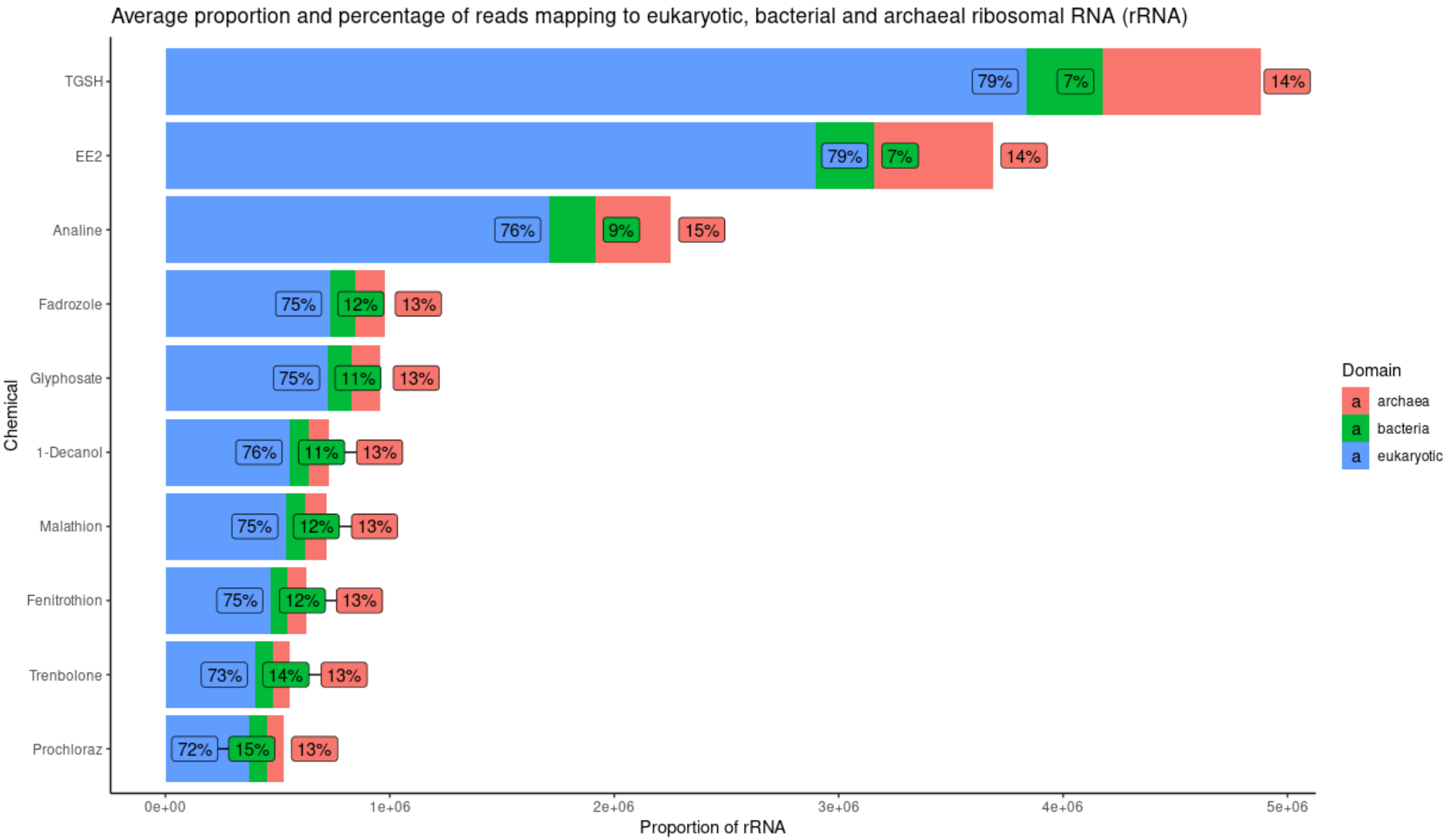
Supplementary Figure 25: The RINs of the samples used for library preparation at Genome Quebec as measured by the Tapestation system. The average RIN score of each chemical group is shown with the vertical red line. The raw data are shown as points overlapping the violin plot, which shows the distribution of sample RINs



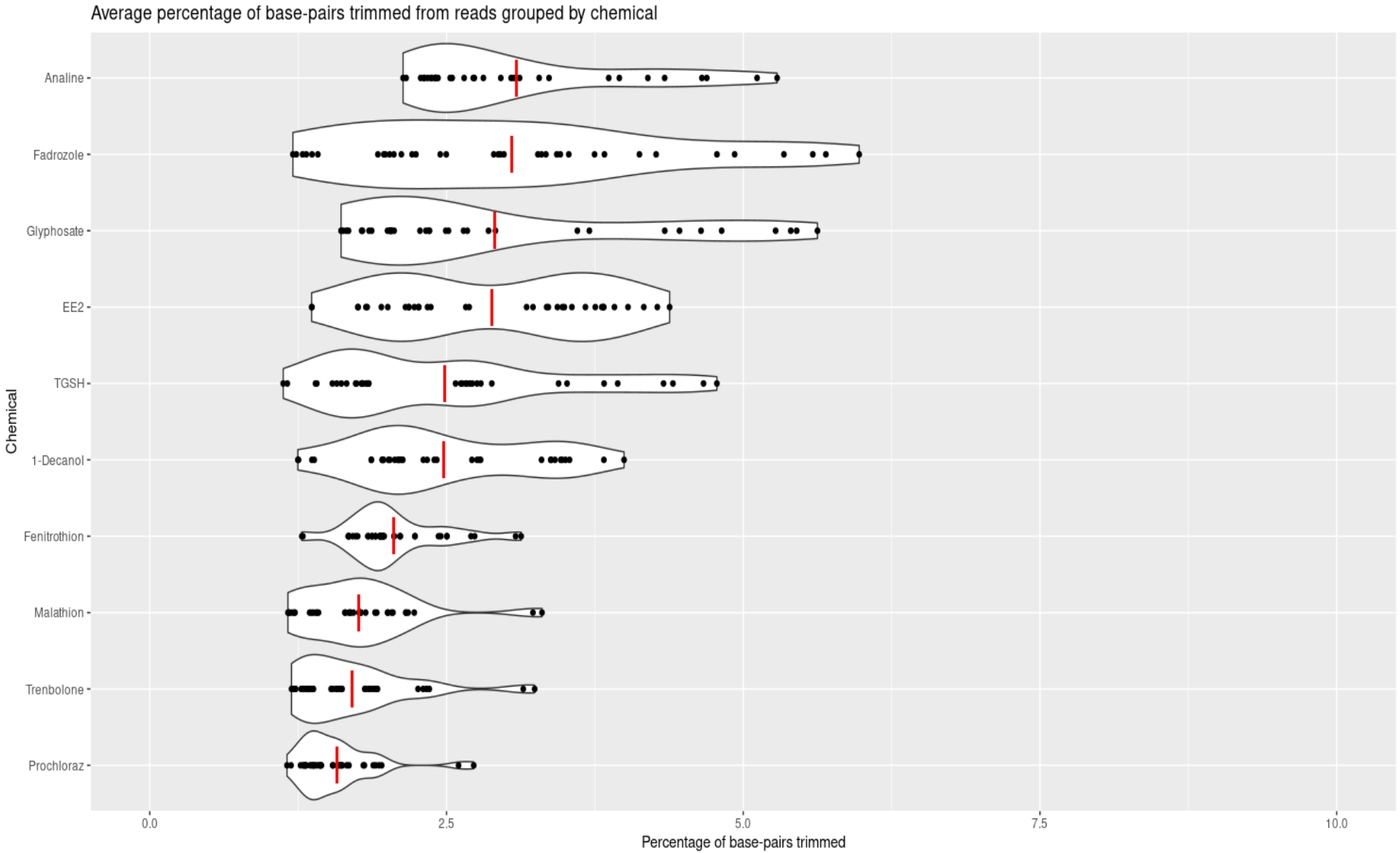
Supplementary Figure 26: The percentage of reads mapping to the transcriptome is shown for each sample. The average percentage of reads mapping to the transcriptome per chemical group is shown with the red vertical line overlapping the violin plots. The violin plots show the distribution of the samples mapping to the transcriptome. Outlier samples were omitted from the violin plots but are still shown as points on the plot. The x-axis begins at 45% mapping percentage and extends to the 90% level.



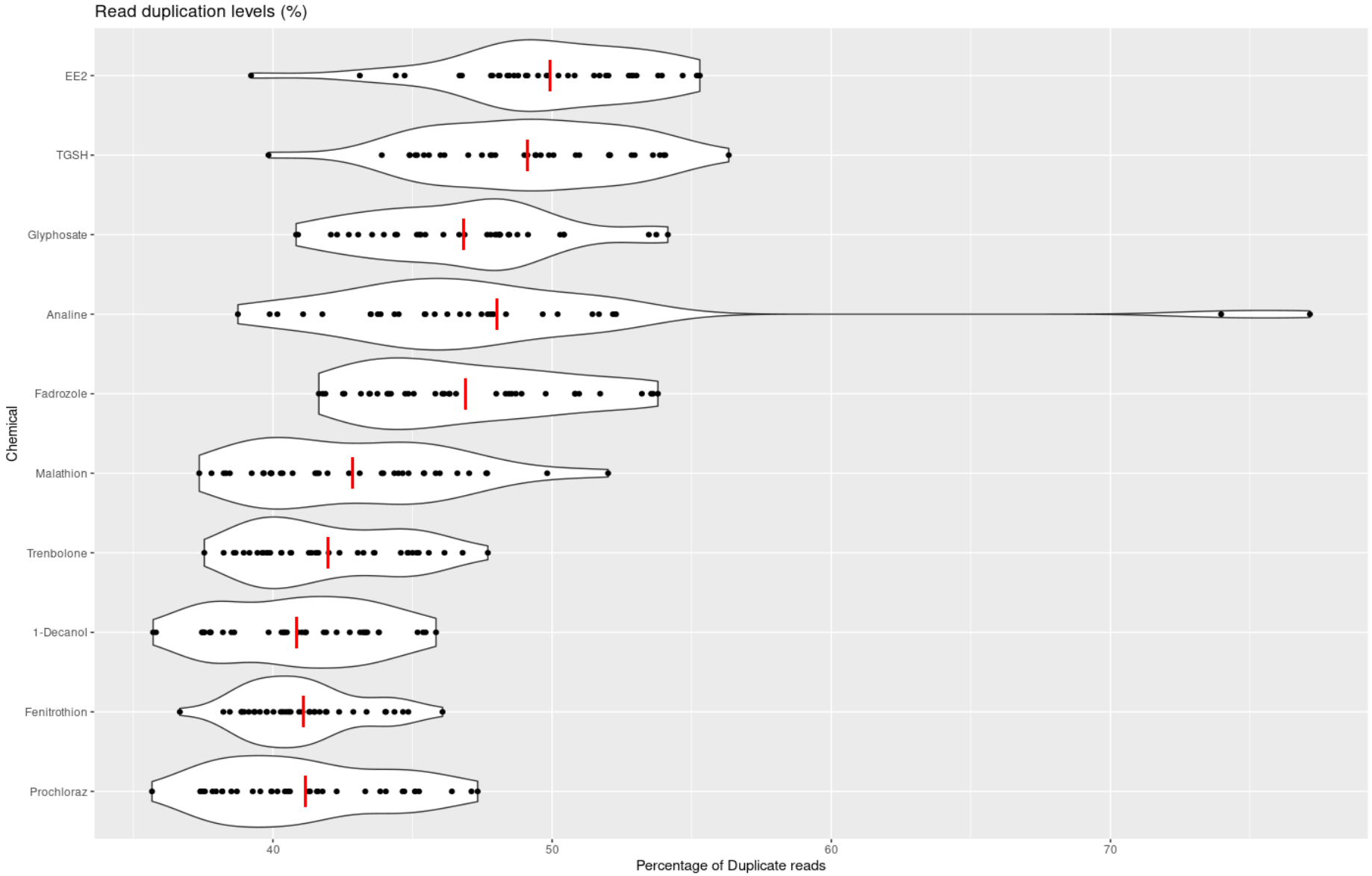
Supplementary Figure 27: The percentage of reads mapping to rRNA in the transcriptome per sample is summarized. The average percentage of reads mapping to rRNA per chemical group is shown with the vertical red line overlapping the violin plots. The violin plots show the distribution of reads mapping to rRNA. Outlier samples were omitted from the violin plots but are still included as points on the plot. The x-axis begins at the 0% contamination level and extends to 70%.



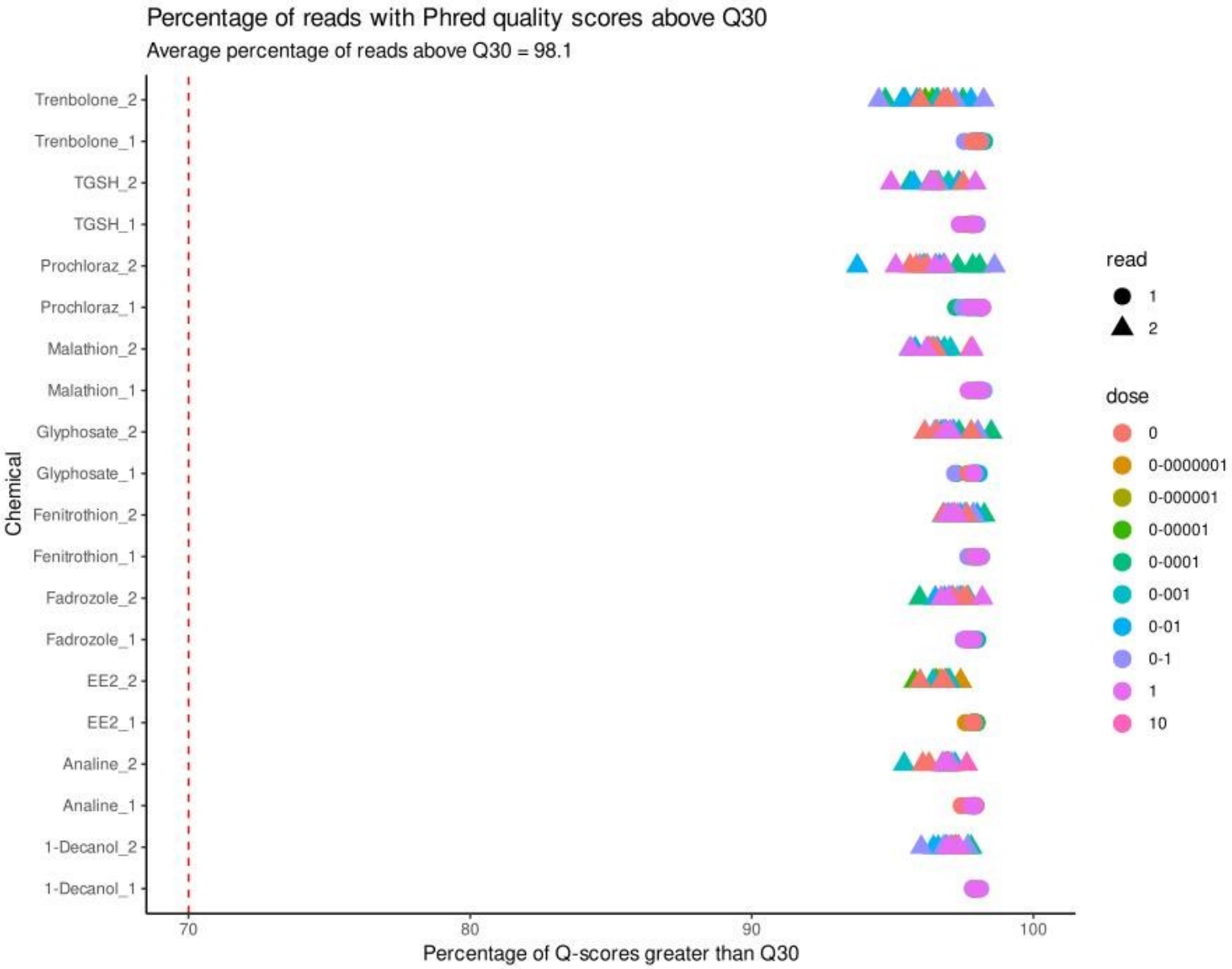
Supplementary Figure 28: The average quantified proportion of reads mapping rRNA in the transcriptome and broken down by biological domain are summarized to evaluate the overall contamination levels per chemical group. The majority of reads mapped to eukaryotic rRNA in the transcriptome.



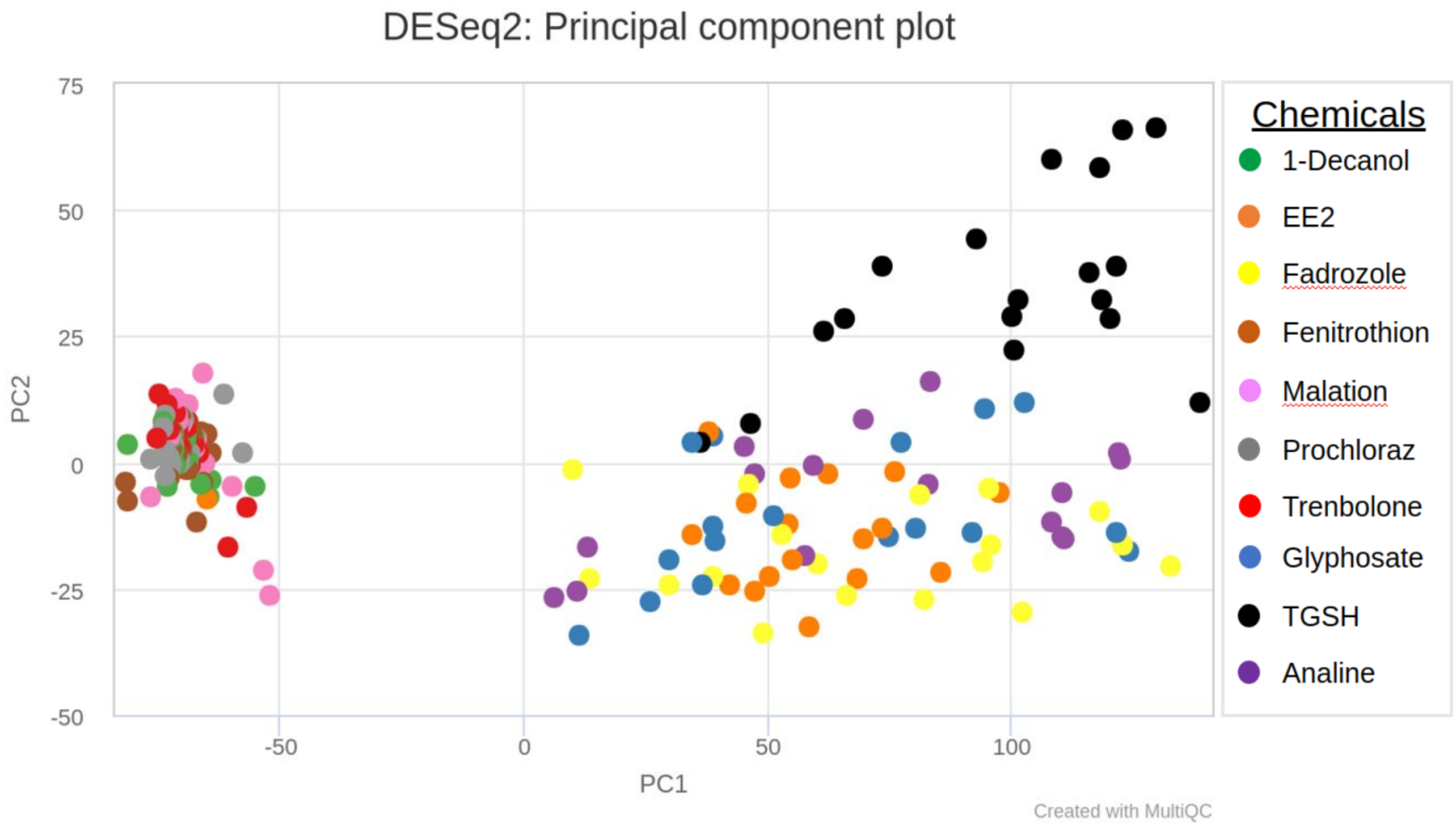
Supplementary Figure 29: The percentage of base pairs trimmed per read (R1 and R2 shown) by Cutadapt is summarized. The vertical red line overlapping the violin plot shows the average percentage of base pairs trimmed per chemical group. The violin plot shows the distribution of samples. The x-axis ranges from 0% to 10%. The average percentage of reads trimmed across all samples was 2.4%.



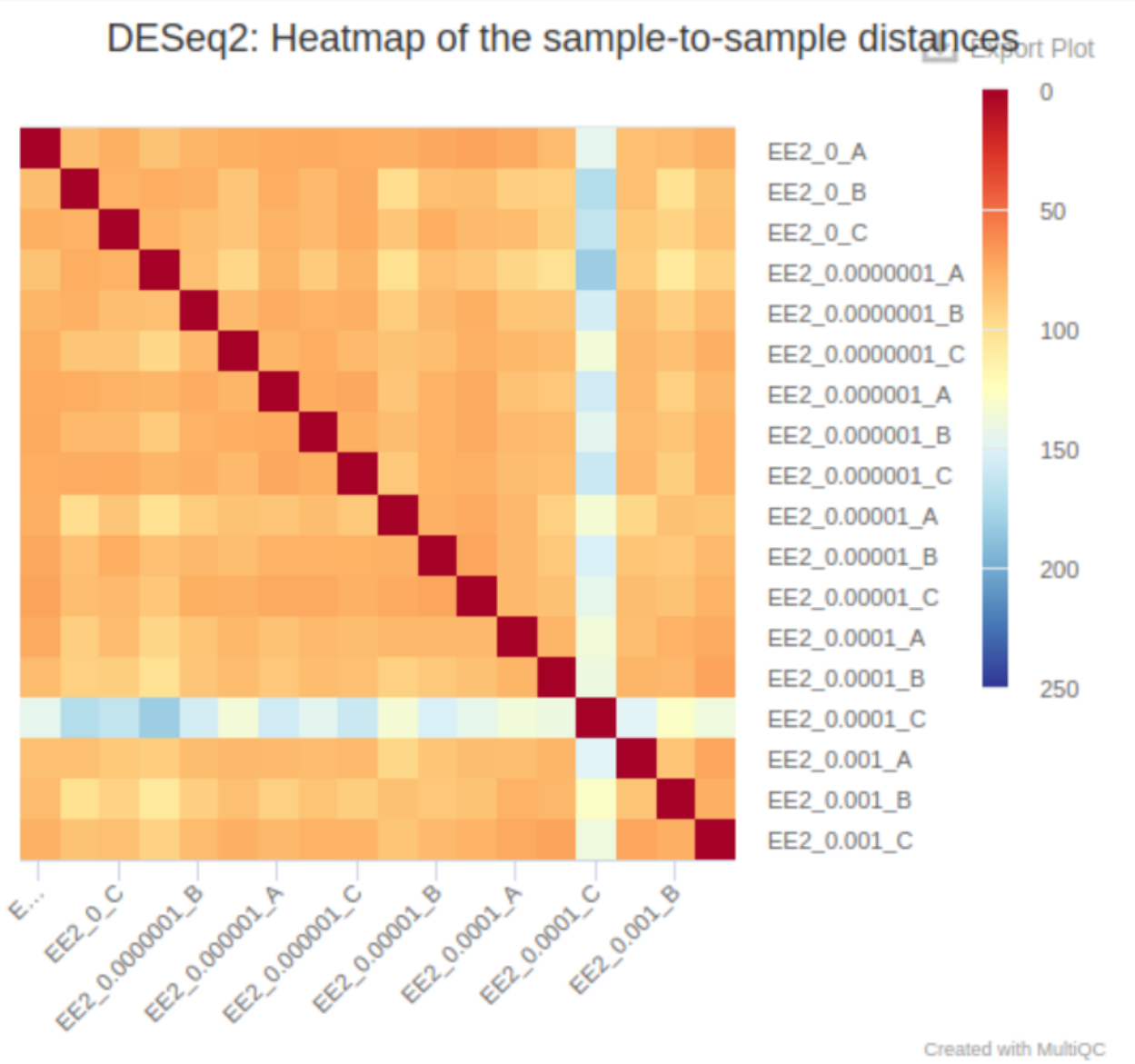
Supplementary Figure 30: The percentage of duplicate reads per read (R1 and R2 shown) is summarized per sample and by chemical group. The vertical red line overlapping the violin plot is the average percentage of duplicate reads per chemical group. The violin plot shows the distribution of samples. The x-axis ranges from 35% to 80%.



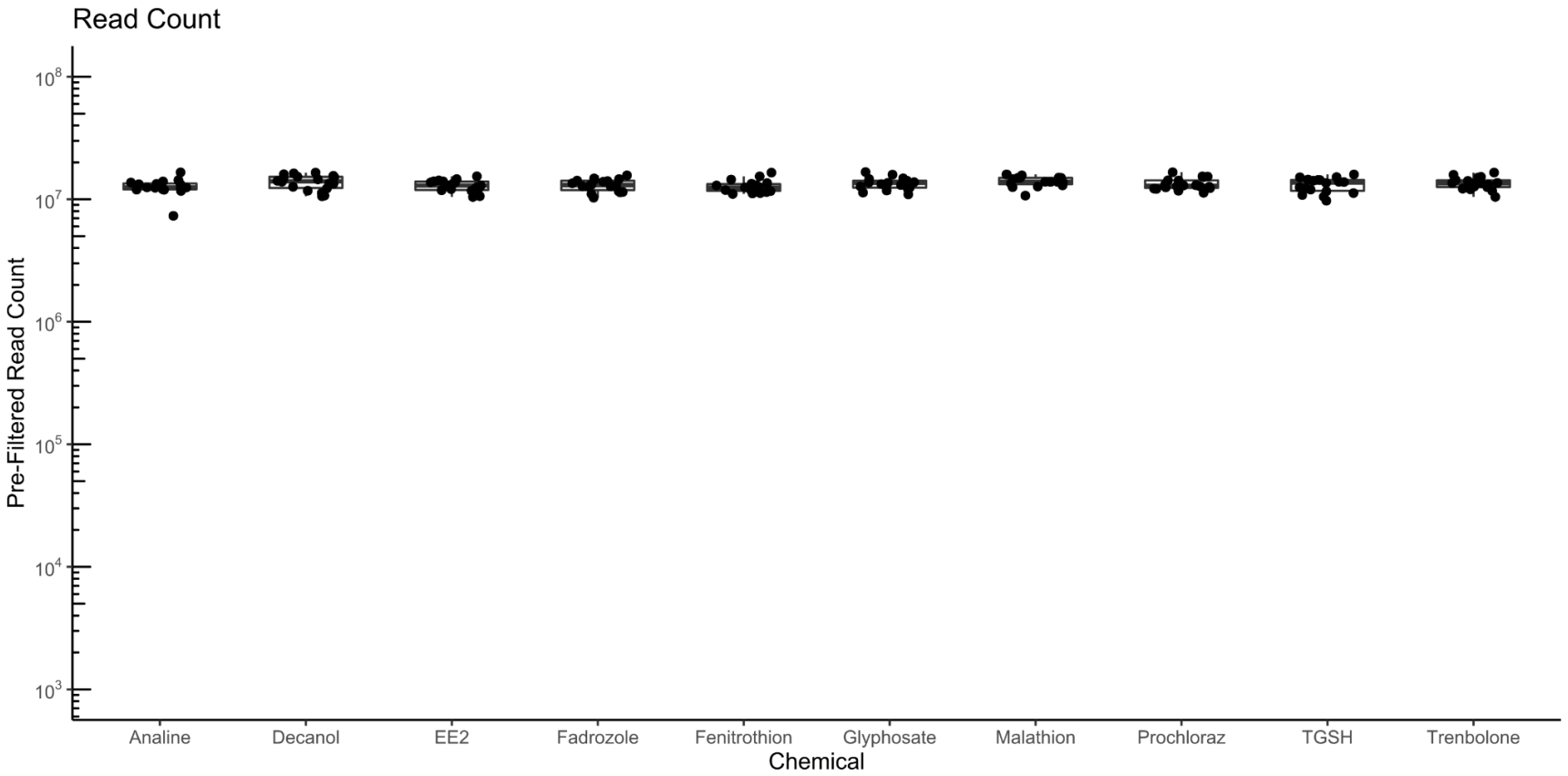
Supplementary Figure 31: The percentage of reads with Phred quality scores above Q30 is summarized. Read 1 is the reverse strand, and read 2 is the forward strand. The R-ODAF Q-score cutoff is shown with the vertical red dashed line at 70% on the x-axis.



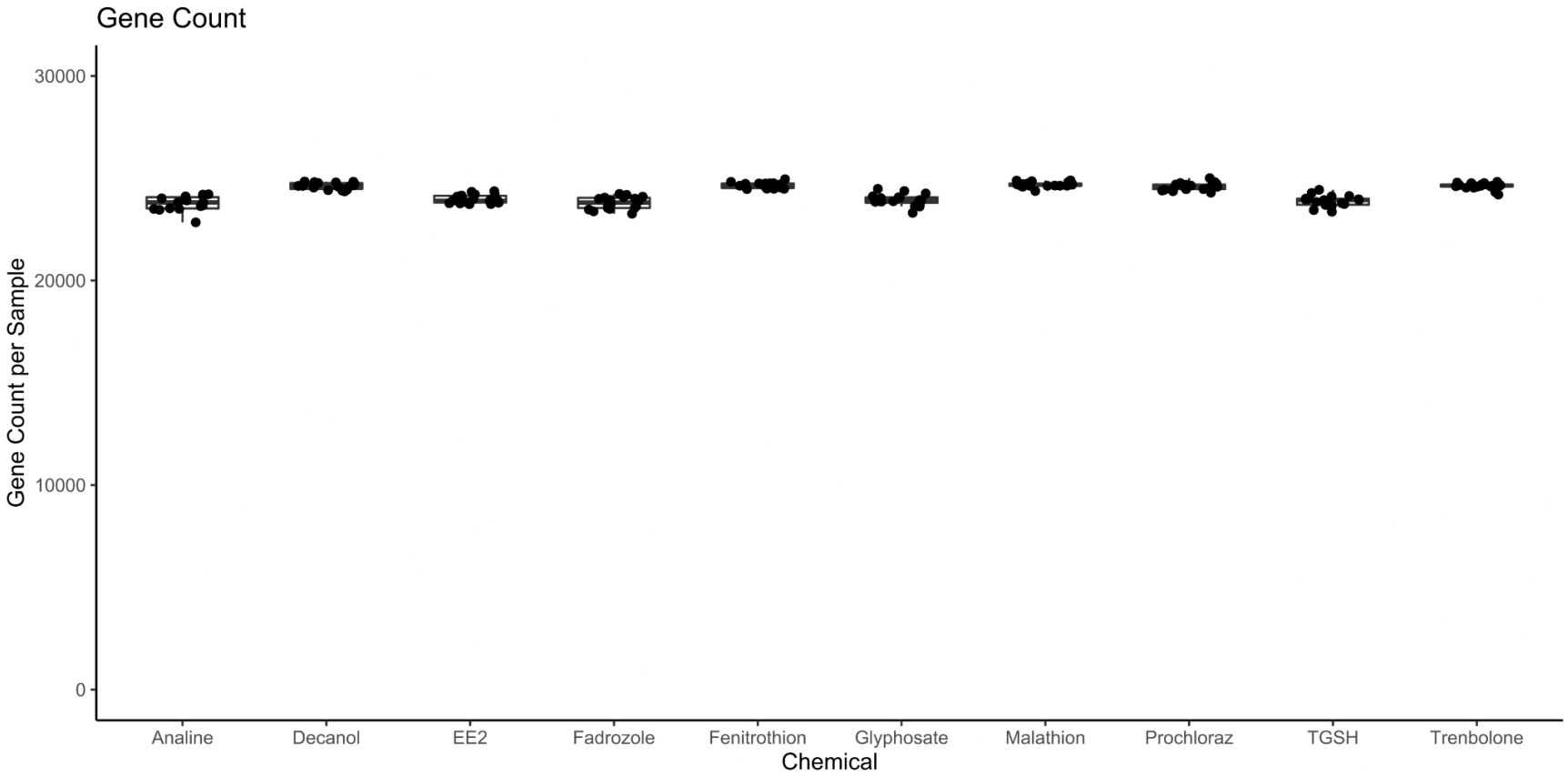
Supplementary Figure 32: After quantifying the RNAseq samples, a PCA was conducted to identify outlier samples. The PCA revealed two distinct clusters of samples driven by PC1.



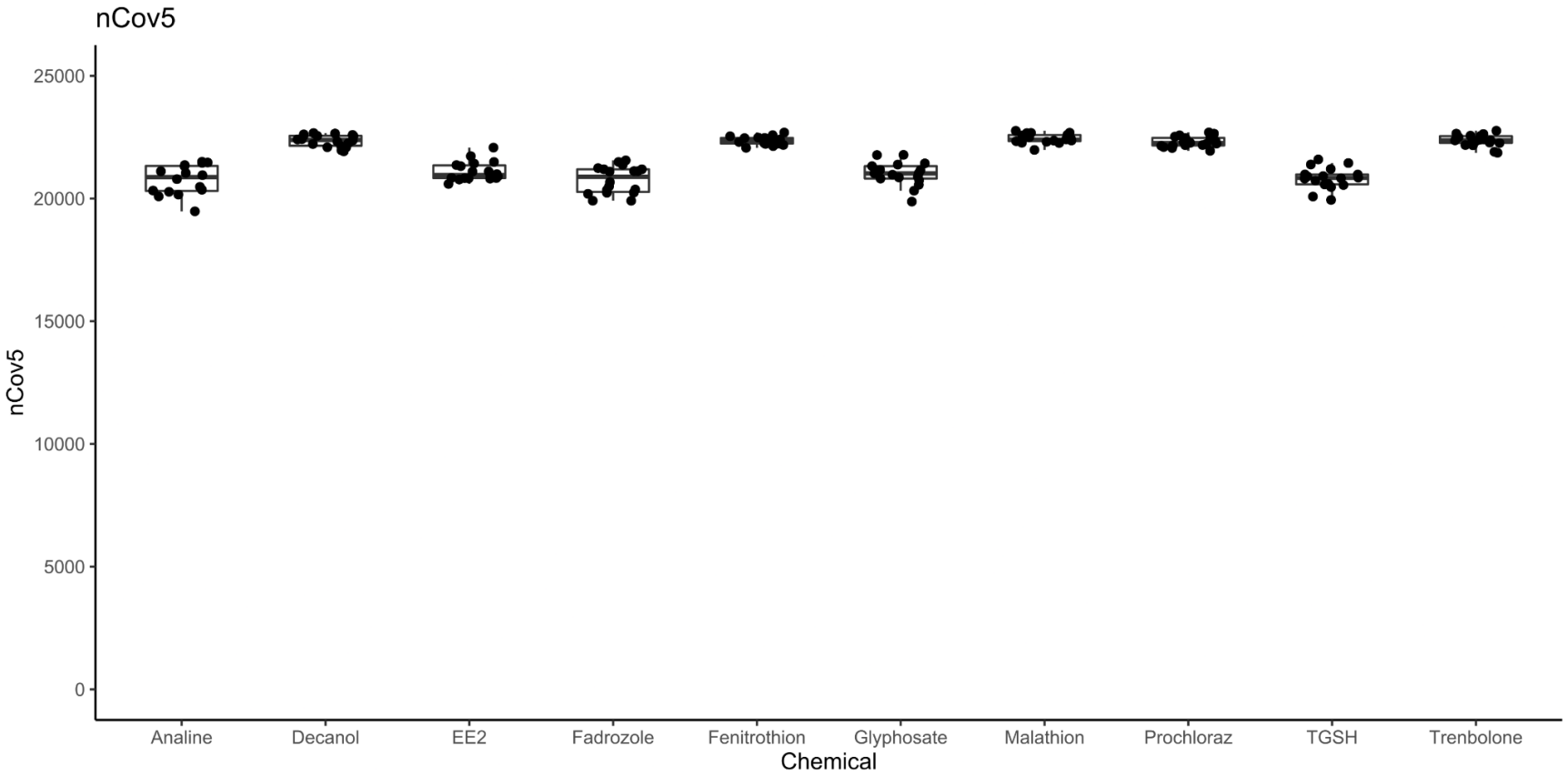
Supplementary Figure 33: A Euclidean distance clustering analysis was conducted after quantifying the RNAseq samples to identify outlier samples. This image shows a clear outlier in the EE2 experiment dose group 0.0001mg/L group. Outliers such as this were removed from the analysis.



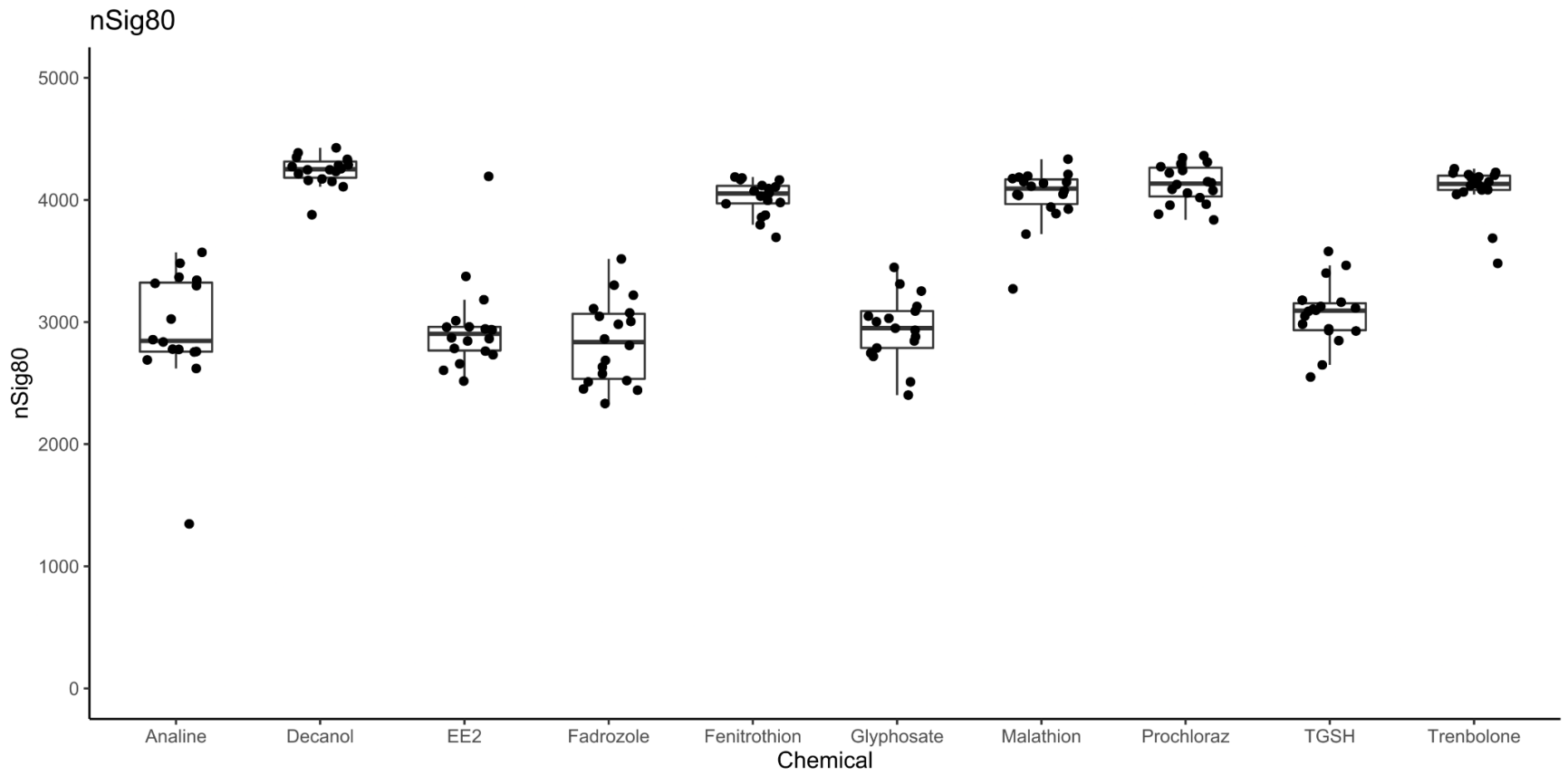
Supplementary Figure 34: The read count of each sample is summarized. None of the samples were below the 5 million reads R-ODAF cutoff threshold.



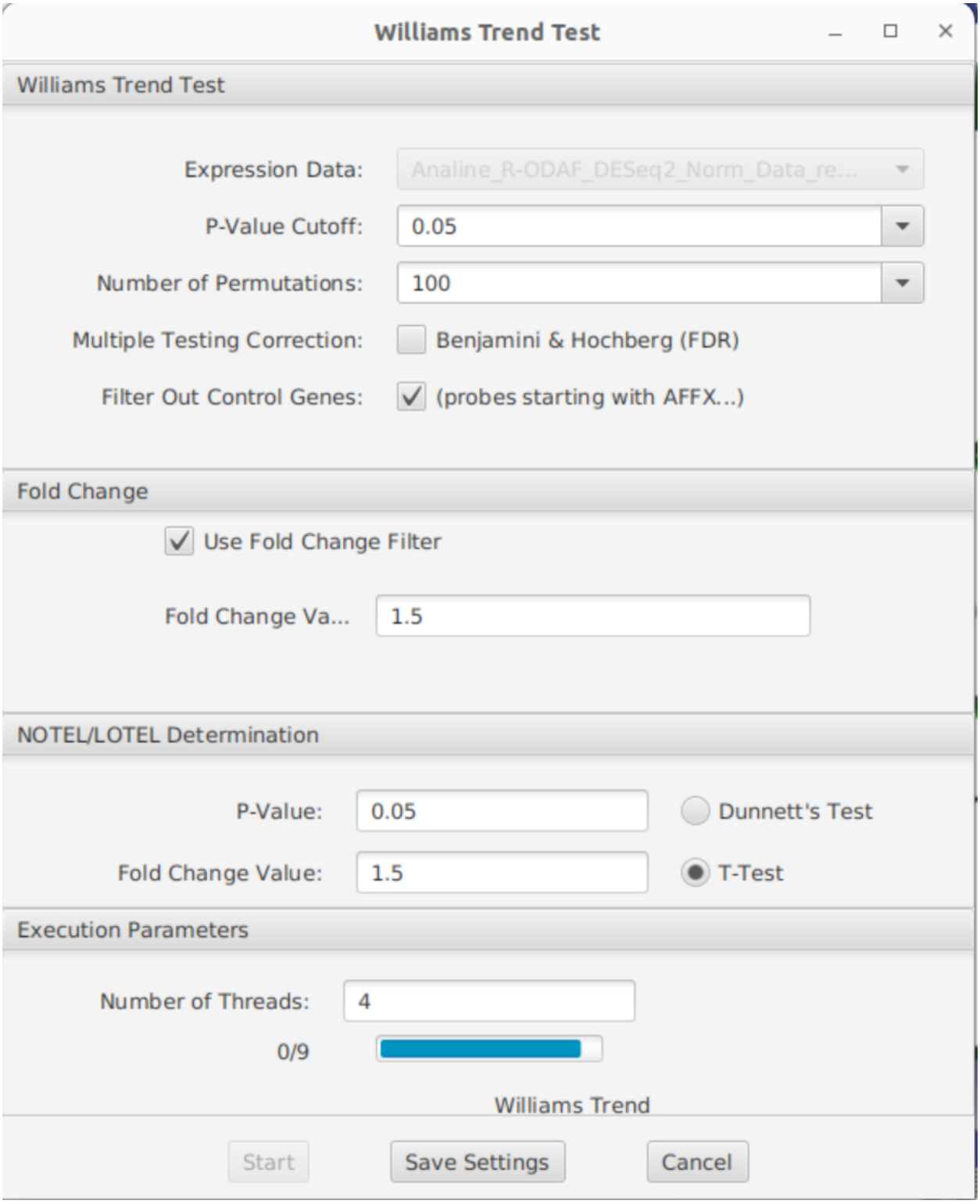
Supplementary Figure 35: The quantified gene count per sample is summarized. The average number of quantified genes per sample is ~22,754 genes.



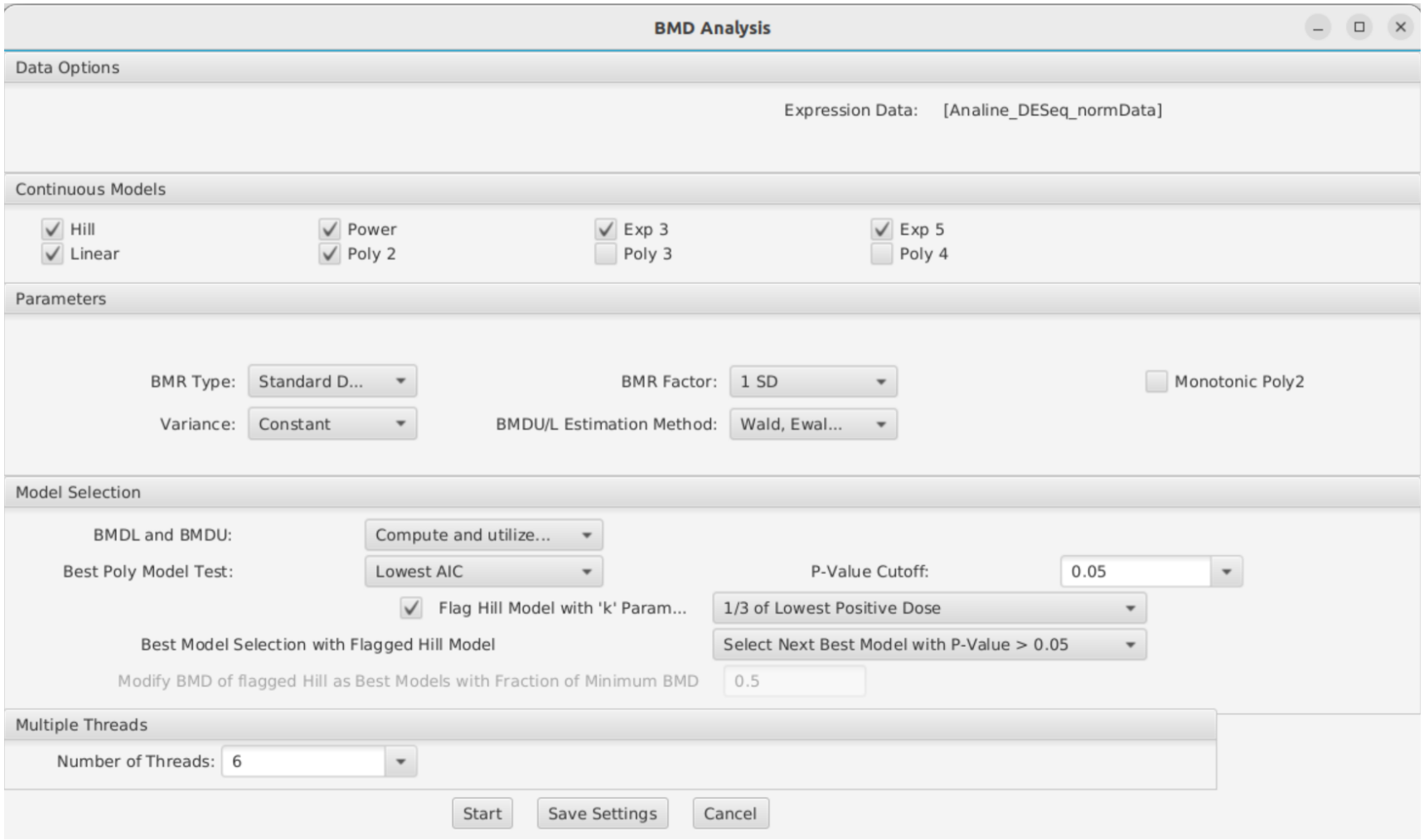
Supplementary Figure 36: The number of quantified genes with at least 5 mapped transcripts per sample is summarized.



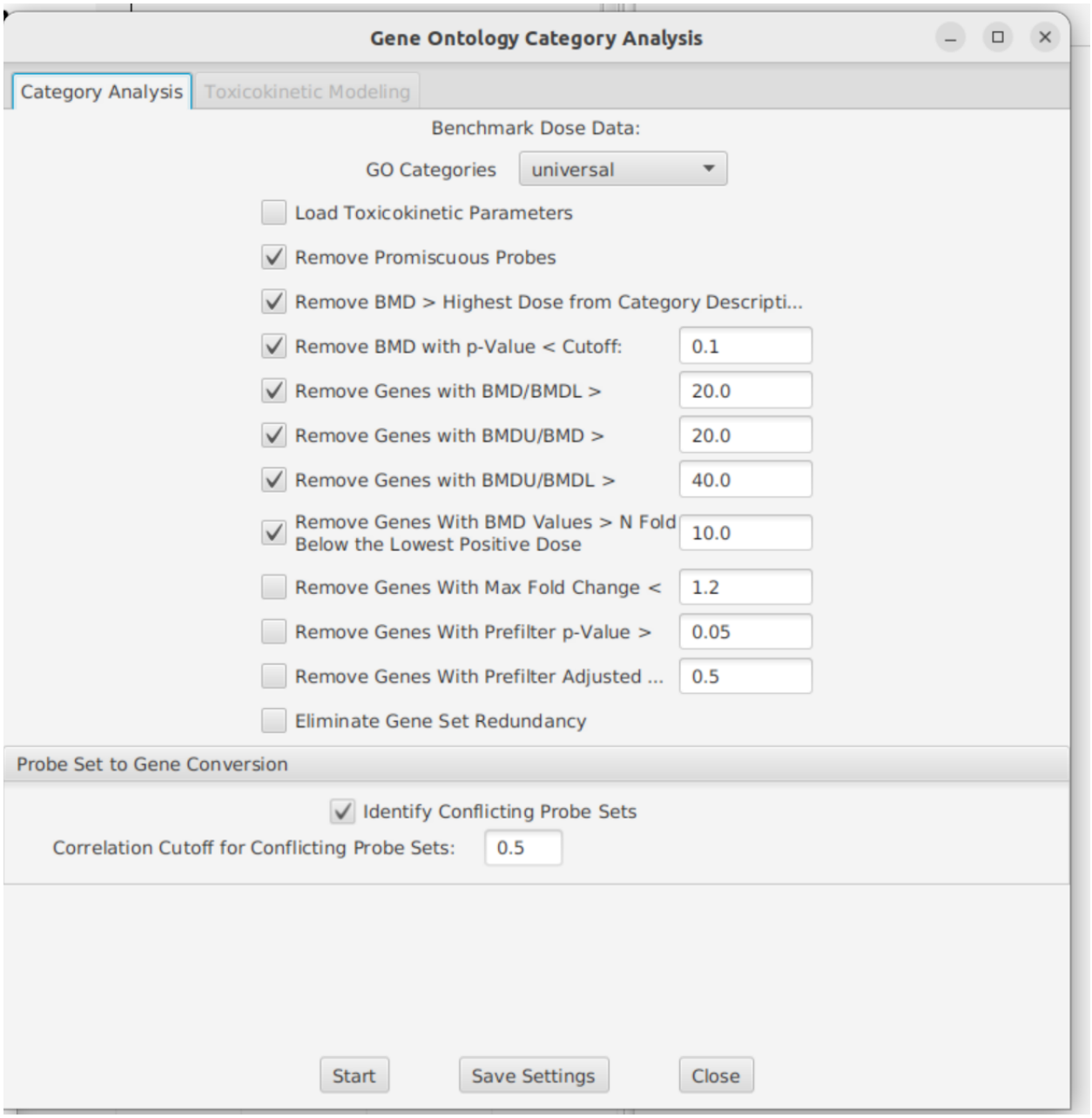
Supplementary Figure 37: The number of highly expressed genes comprising 80% of all quantified transcripts per sample is summarized. The nSig80 revealed a distinct sample clustering effect



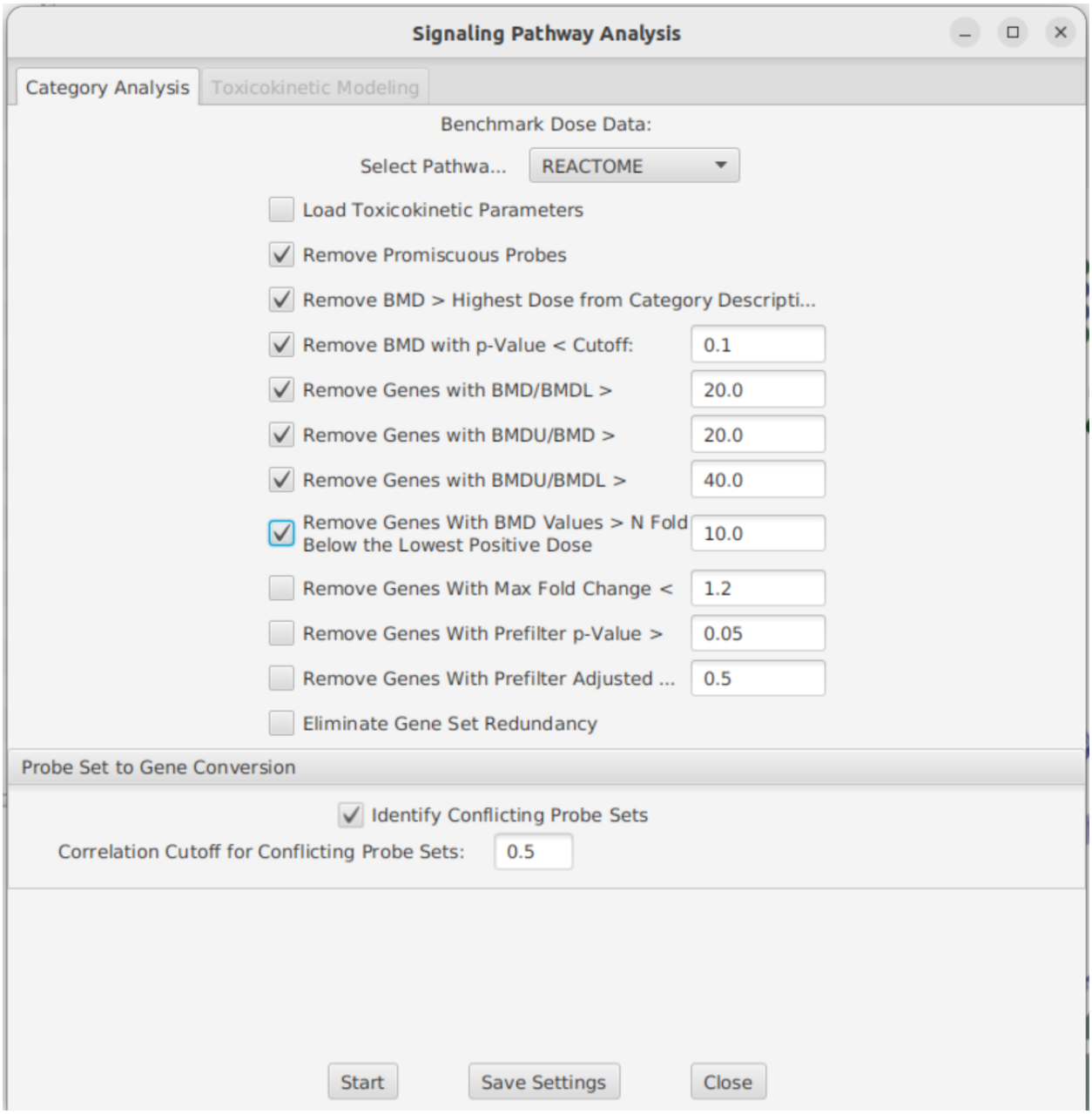
Supplementary Figure 38: A screenshot of the parameters used in BMDExpress 3.0. William's trend test and 1.5 times fold-change filter were used in BMDExpress 3.0 to identify DEGs.



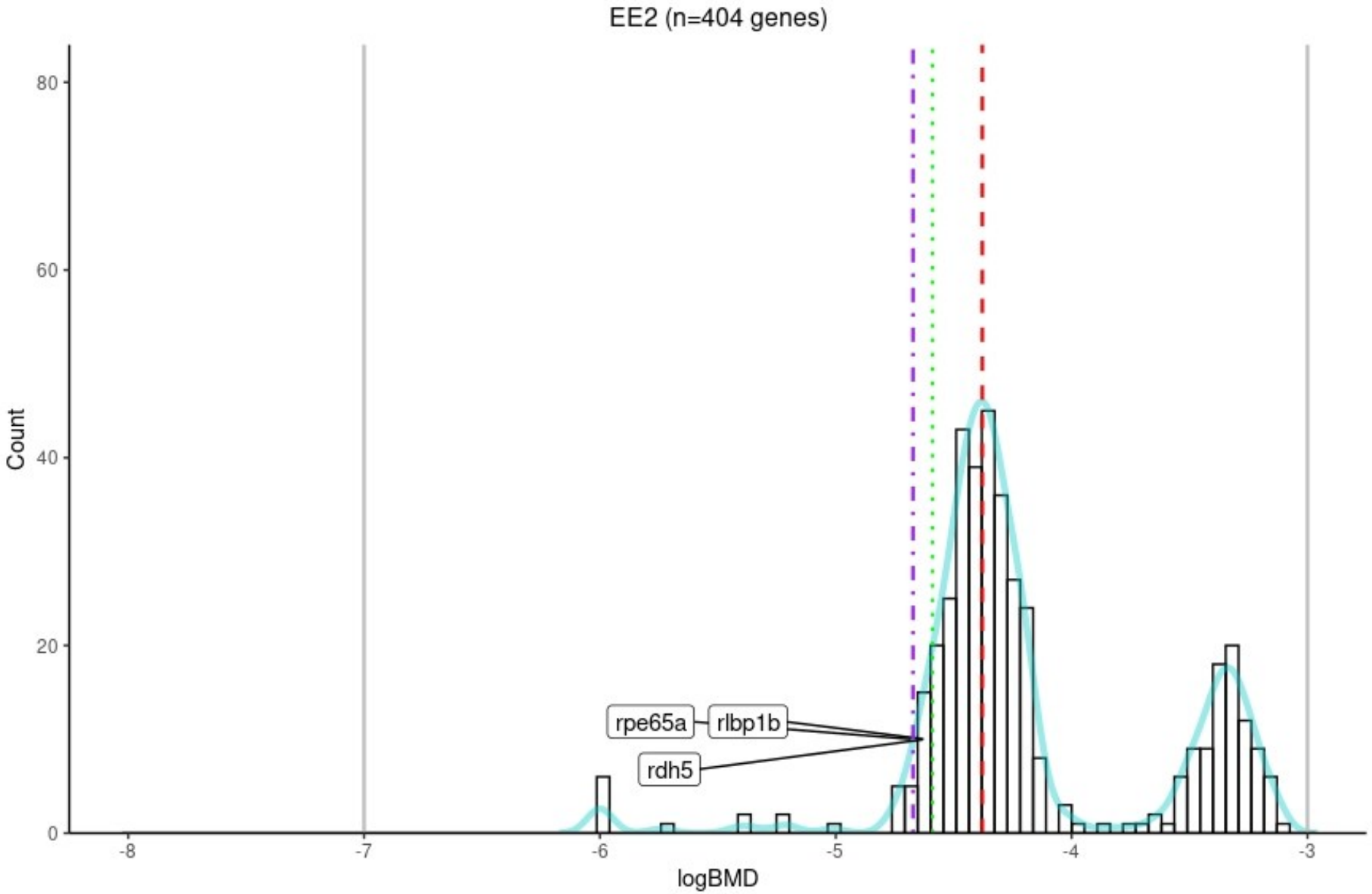
Supplementary Figure 39: A screenshot of the parameters used in BMDExpress 3.0 for the BMD Analysis. These parameters were used to calculate BMDs.



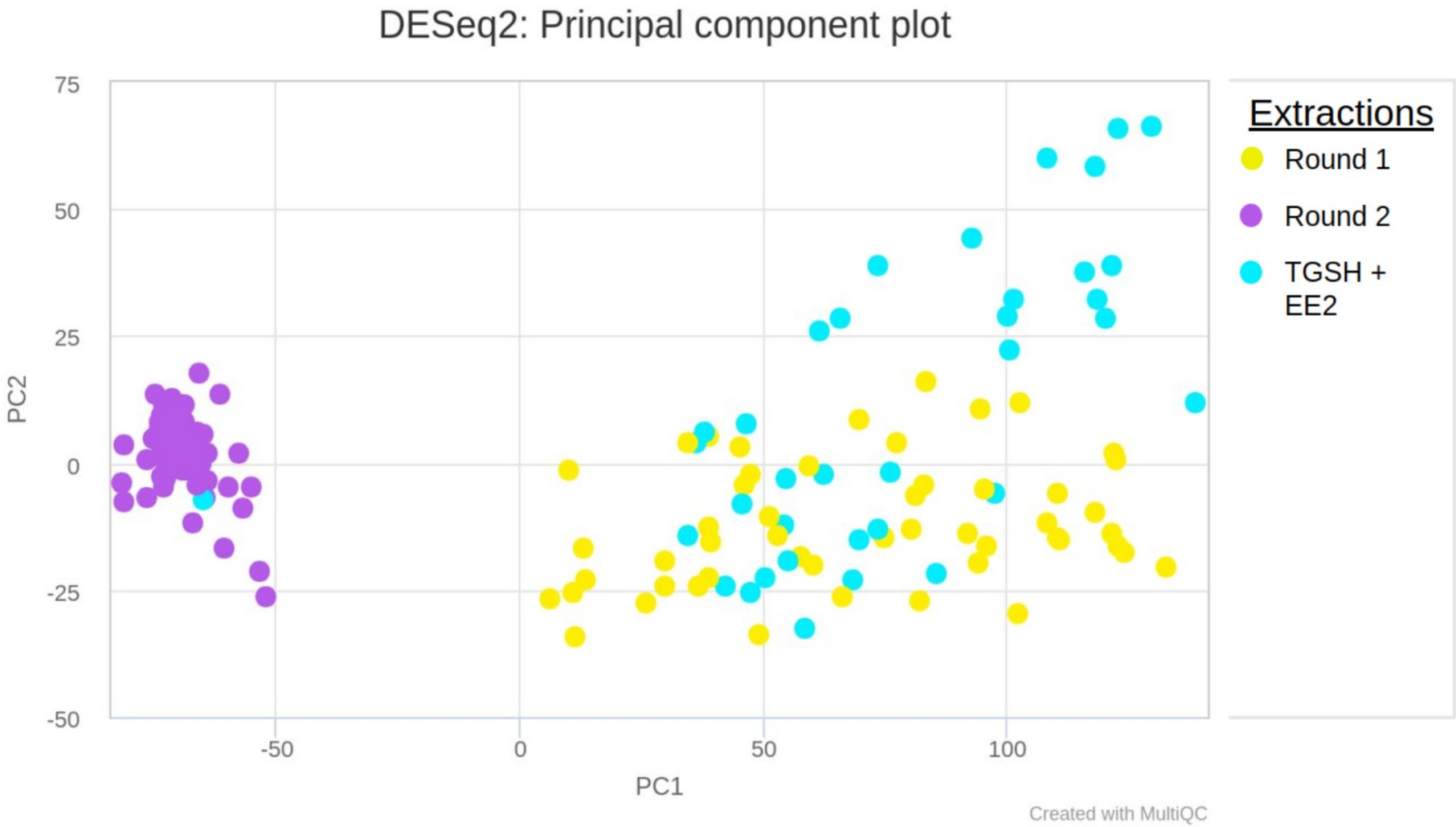
Supplementary Figure 40: A screenshot of the parameters used in BMDExpress 3.0 for the GO term enrichment analysis.



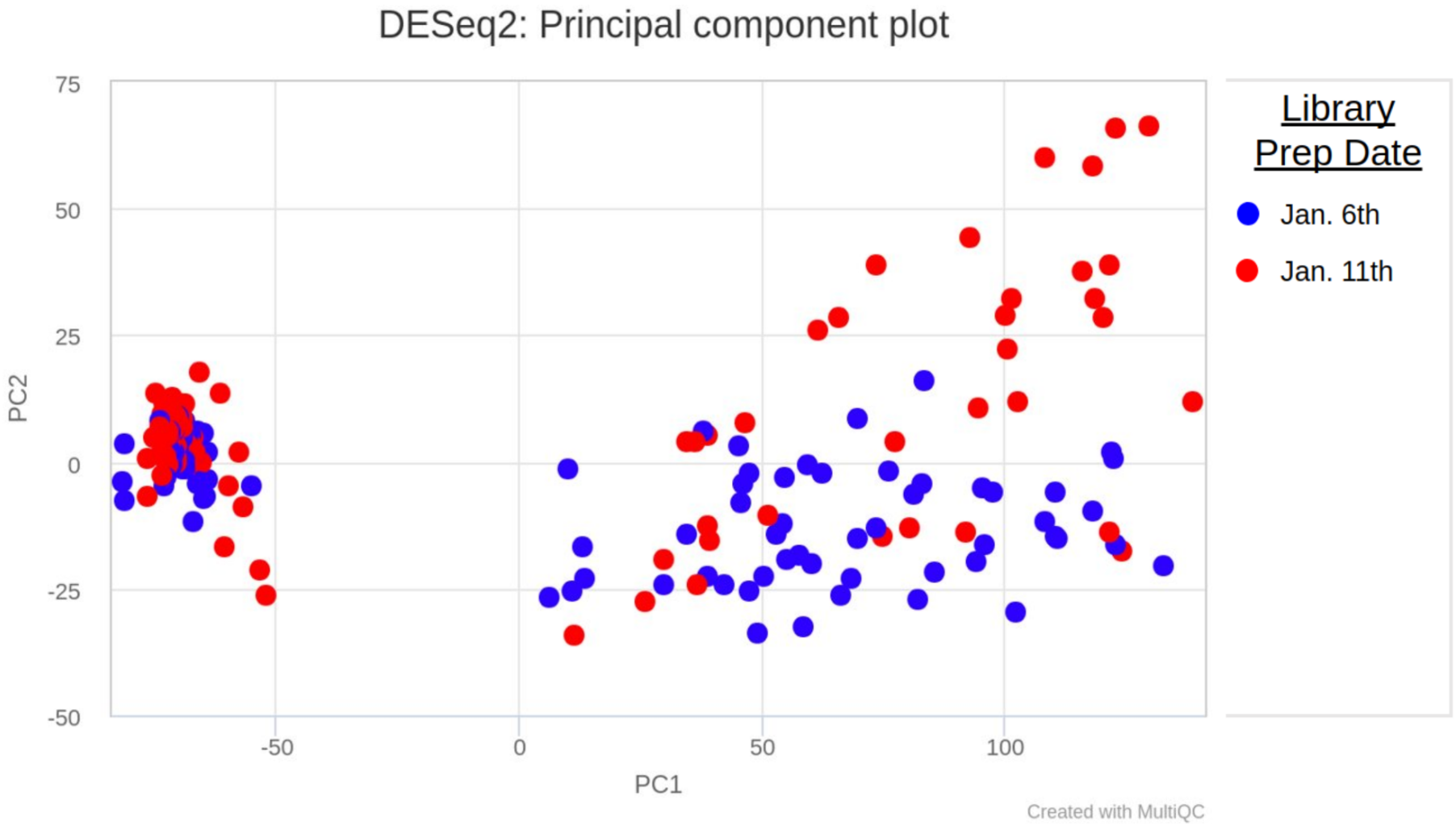
Supplementary Figure 41: A screenshot of the parameters used in BMDExpress 3.0 for the REACTOME pathway enrichment analysis.



Supplementary Figure 42: The BMDs of the genes enriching the sensory perception/visual phototransduction/twilight vision REACTOME pathway and where they fall on the distribution of all BMDs.



Supplementary Figure 43: A PCA was conducted to identify outlier samples after quantifying the RNAseq samples. Samples were highlighted by extraction 'rounds,' TGSH and EE2 were highlighted separately because they were initially extracted by hand in the uOttawa Mennigen lab for a different Master's project.



Supplementary Figure 44: A PCA was conducted to identify outlier samples after quantifying the RNAseq samples. Samples were highlighted by library preparation date.