A Systems Biology approach to understanding and monitoring chemical toxicity in the environment

Thesis submitted in accordance with the requirements of

the University of Liverpool for the degree of Doctor of Philosophy

By

Danilo Basili

May 2018

Abstract

Chemicals pose every day a continuous hazard to both human health and environment. Unfortunately, Information about chemicals Mode of Action (MoA) for most of these compounds is limited. Development of approaches able to elucidate chemicals mechanisms of action is needed in order to improve risk assessment. Environmental omics aims to provide tools and methodologies to address these goals. Omics technologies in combination with system biology approaches have the potential to provide a powerful toolbox for understanding chemicals mode of action and consequently the outcomes these compounds trigger.

The work presented in this thesis demonstrates the effectiveness of such approach in the context of environmentally relevant species. More specifically I focused on characterization of single chemical and chemical class toxicity mechanism in zebrafish embryos (*Danio rerio)* and in a fish gill cell line (*Rainbow trout*) and I demonstrated that the transcriptional state of an *in vitro* system exposed to a panel of environmentally relevant chemicals can be used as a biosensor to predict toxicity in an *in vivo* system. I also developed a computational model of ovary development in Largemouth bass (*Micropterus salmoides*) and used this to successfully identify chemical compounds with the ability to affect reproduction. Lastly, I developed a method to identify novel endocrine disrupting compounds in *Daphnia magna* supporting the use of this species for rapid screening in risk assessment.

My results demonstrated the potential of system biology and data-driven science in identifying novel mechanisms of environmental toxicity and to develop a set of biomarkers for monitoring purposes. Further development building on these findings could potentially lead to improvements in risk assessment.

Acknowledgements

I would like to dedicate this thesis to my parents, Daniela and Carlo, for always believing in me. Their love, care and sacrifice will guide me for the rest of my life. I would also like to thank my brother Simone for always caring about my studies and for being there when needed. To my grandparents who grew me up with love and taught me many important things during my life. To my uncles who have been giving me many special moments full of joy.

I would like to thank Michela for always pushing me beyond any difficulties by delivering the care, the smiles, the love and the understanding needed. A special thanks to Eloise for giving me love and support during the writing up stage which was crucial. Professionally I would like to thank my supervisor, Dr. Francesco Falciani for giving me the chance to undertake the PhD in a fascinating field. I would also thank Dr. Philipp Antczak for teaching me the skills needed for the achievement of this PhD. Thanks to all the other members of the team Erica, Peter, Kim, Xiaoliang, John Herbert and John Ankers for the help and the advices received. To Diego, Aitor, Silvia and all my dear friends in Liverpool for making these years enjoyable. Finally, to Martina, Giulio, Silvia, Gulia, Emilio and all the friends that for years have been supporting me making my achievements possible.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ANOVA, Analysis of Variance

AO, Adverse Outcome

AOP, Adverse Outcome Pathway

ARACNE, Algorithm for the Reconstruction of Accurate Cellular Networks

CTD, Comparative Toxicogenomic Database

DAG, Directed Acyclic Graph

DEGs, Differentially expressed genes

DAVID, Database for the Annotation, Visualization and Integrated Discovery

DGC, Daphnia Genomic Consortium

ED, Endocrine Disruptor

EPA, Environmental Protection Agency

ERA, Environmental Risk Assessment

ES, Enrichment Score

FET, Fish Embryo Toxicity Test

FDR, False Discovery Rate

GA, Genetic Algorithm

GO, Gene Ontology

GSEA, Gene Set Enrichment Analysis

HTS, High-throughput Screening

IVIVE, *In vitro* to *In vivo* Extrapolation

k-NN, K-Nearest Neighbour

KE, Key Event

KEGG, Kyoto Encyclopaedia of Genes and Genomes

LOESS, Local Regression

MAQC, MicroArray Quality Control Project

MI, Mutual Information

MoA, Mode of Action

MIE, Molecular Initiating Event

NC3Rs, National Centre for the Replacement, Refinement & Reduction of Animals in Research

NGS, Next Generation Sequencing

OECD, Organisation for Economic Co-operation and Development

PC, Principal Component

PCA, Principal Component Analysis

QSAR, Quantitative Structure-Activity Relationships

REACH, Registration, Evaluation, Authorisation and Restriction of Chemicals

REVIGO, Reduce and Visualize Gene Ontology

SAM, Significance Analysis of Microarray

SOM, Self-Organising Maps

SOTA, Self-Organisng Tree Algorithm

TDA, TimeDelay ARACNE

Chapter 1: Introduction and Background

1.1 A revolution in biology: towards a holistic understanding of complex biological systems

Traditional research in molecular biology has always been driven by a reductionist approach focusing on the mere characterization of the components of an organism. Reductionists analyse biological systems by breaking it down into its smaller components and determining the connection between the parts. The underlying key concept is that the individual components (i.e. the molecules and their structures) have enough explanatory power to provide an understanding of the whole system[1,2]. However, biological systems are extremely complex and their properties cannot be fully understood by simply looking at their individual components. Cellular and organismal constituents interact in many ways, including negative feed-back and feed-forward controls, hence their structures and dynamics might be investigated in intact cells and organisms. The last decade or so has witnessed the development of a more holistic approach embraced by the rise of System Biology[2,3]. As defined by Garcia-Reyero and Perkins[4], System Biology is "the study of an organism as interacting networks with the goal of understanding and predicting properties". More specifically, the goal of System Biology is to depict the interaction between genes, proteins and biochemical reactions in terms of interacting networks and to characterize the flow of information that links these elements to a biological process[5].

The shift from a reductionist approach to a holistic one along with the development of System Biology it has been made possible by the development of new technologies with the ability to enable a rapid and broad characterization of many different levels of biological organization. These functional genomics technologies, as transcriptomics, proteomics and metabolomics, allow measuring the expression of tens of thousands of genes, proteins and metabolites in single

experiments providing an overview of the molecular state of a given cell, tissue or even entire organs at a given time point[6,7]. Martyniuk et al.[8], applied a transcriptome-based approach to characterized molecular pathways and temporal gene expression patterns underlying oocyte maturation in wild largemouth bass (*Micropterus salmoides*). Using a proteomic approach, Chen et al., identified 361 proteins to be differentially expressed between mature and immature sperm in the catfish *Cranoglanis bouderius*[9]. They were involved in 235 pathways of which the tricarboxylic acid (TCA) and the EABB played an important role in the energy metabolism of sperm and the spermatogenesis, respectively. Metabolomics was used by Ali et al. to identify a panel of 19 metabolites able to discriminate between segment-elevation myocardial infarction (STEMI) and both unstable angina (UA) and healthy patients[10]. They identified hydrogen sulphide (H_2S), an endogenous gasotransmitter with profound effect on the heart, as promising predictive biomarker that will potentially allow for an earlier medical intervention.

This unprecedented technological development in Biology has been responsible for an explosion of multivariate molecular data, which in turn has stimulated the development of a data-driven approach to understanding biological systems. Data-driven Biology, differently from hypothesis-driven Biology, doesn't require the formulation of a hypothesis to be tested beforehand but it just uses omics technologies to generate thousands of data that will be mined to generate useful and specific leads for further study and validation. Given the large amount of data generated and the little knowledge about the way biomolecules interact with each other, the key question is whether we can improve our understanding of biological systems from high-throughput data. Clearly, there is a need to develop sophisticated computational tools able to integrate genome-wide measurements across different levels of biological organization, to identify key features of the biological system able to drive a specific biological function and to statistically infer the way these features works in concert to generate a given biological response[11].

1.2 The role of omics technologies in environmental biology

The application of high-throughput technologies to environmental biology and environmental ecotoxicology field stimulated the emergence of a new field of study called "Environmental OMICS". As defined by Ge et al. "Environmental omics is the application of OMICS technologies to better understand the environmental and genetic factors, toxicity mechanisms, and modes of action in response to both acute and chronic exposure to environmental chemicals and, in the long-term, development of diseases caused or influenced by these exposures"[12]. Environmental omics contributes to the understanding of chemical toxicity mechanisms, to develop biomarkers for environmental risk assessment and to predict effects on human health. In a longer term these tools are expected to result in a better understanding of ecological functions. Genomics and transcriptomics are probably the most widespread omics technologies. There are several examples of the application of these approaches in the field of environmental toxicology. Examples are, the identification of chemical Modes of Action (MoA), the definition of the specific molecular initiating events (MIE)[13] that trigger a physiological adverse outcome and for understanding the effects of chemical mixtures[14].

In a recent study, Brander et al. showed that bifentrhin exposure to an estuarine fish (*Menidia beryllina*) have the potential to alter metabolic processes and endocrine signalling and to decrease offspring[15]. Uren Webster and Santos investigated the potential toxicity effects of exposure in brown trout (*Salmo trutta*) to glyphosate and roundup, two herbicides used worldwide, and identify a similar mechanism of action through oxidative stress for environmentally relevant concentrations[16]. Kumar and Denslow provide a good overview of transcriptomic responses to chemical exposures in a variety of fish species and highlight future directions[17].

Galland et al. showed that PAHs/PCBs exposures in the European flounder (*Platichthys flesus*) resulted in the deregulation of proteins involved in oxidative stress and glutathione metabolism as well as betaine demethylation pathway and the methionine cycle[18]. Peng et al., investigated the effects of methamidophos, a worldwide used pesticide, on the brain tissue of the flounder (*Paralichthys olivaceus*) and find out that proteomic changes were associated with pathways involved in immunity and stress with HSP90 and GzmK protein to play a crucial role[19].

Yan et al., demonstrated that exposure of zebrafish embryos to the widely used pesticide fipronil resulted in the identification of 26 differential metabolites involved in 5 different biological pathways revealing new insight into fipronil toxicity[20].

However, transcriptomics, proteomics or metabolomics taken alone, do not always reflect relevant biological responses associated with toxicological effects. Important biological changes in proteins and metabolites are not detected by simply investigating the levels of mRNA. Hence, the need of integrating transcriptomics, proteomics and metabolomics towards multi-omics approaches able to give a better overview of the real changes undergoing in a cell or tissue as a result of chemical exposures. Qiao et al. investigated the effect of hepatotoxic exposures to depict the sexual dimorphism in adult medaka fish using a multi-omic approach[21]. Norris et al., successfully developed a high-throughput multi-omics platform to investigate cisplatin-induced molecular changes and cisplatin resistance[22].

Nowadays, many challenges in the field of environmental toxicology still need to be addressed. One of them is linked with the analysis of exposure to chemical mixtures. In the natural environment, we are exposed to multiple chemicals. Although approaches aiming to unravel mechanisms of toxicity of single chemicals have been successfully applied, they cannot be used to investigate chemical mixtures. The main reason is because synergistic effects are usually found

in chemical mixtures. Omics technologies offers the potential to overcome these limitations since they enable high-throughput analysis and are able to screen multiple molecular targets and environmental responses simultaneously which is critical to understand the components that cause toxicity in a mixture. Moreover, Omics technologies allow the identification and prioritization of chemical mixtures. Many different studies have already been performed aiming to characterize molecular effects of chemical mixtures. Wang et al. have recently demonstrated that coho salmon (*Oncorhynchus kisutch*) exposures to mixture of pesticides resulted in the disruption of cellular pathways underlying olfactory system[23]. Oliveira et al. showed that exposure to a mixture of benzo(a)pyrene, dichlorodiphenyltrichloroethane and tributyltin is more toxic than exposures to the single chemicals in silver fish (*Rhamdia quelen*)[24].

Another challenge lies in the combined analysis of multi-omics data (e.g. genomics, proteomics and metabolomics) and the integration with classical toxicological endpoints which would allow the detection of health adverse effects and the identification of biomarkers. This challenge is also linked with the lack of a well annotated genome sequence for many environmentally relevant species. However, with cost of sequencing steadily decreasing more genomes of non-model species are sequenced, providing new insights into toxicity mechanisms of action and risk assessment.

1.3 The Adverse Outcome Pathway (AOP) framework

In 2012, the Organisation for Economic Co-operation and Development (OECD)[25] launched a new program on the development of Adverse Outcome Pathway (AOP), first proposed by Ankley et al.[26]. An AOP is a conceptual framework that describes the sequential chain of causally connected Key Events (KEs) between two anchor points, a Molecular Initiating Event (MIE) and an Adverse Outcome (AO) that occur at a level of biological organization relevant to risk assessment (fig.

1.1). The goal of an AOP is to provide the framework to describe the KEs that lead from MIE to AO.

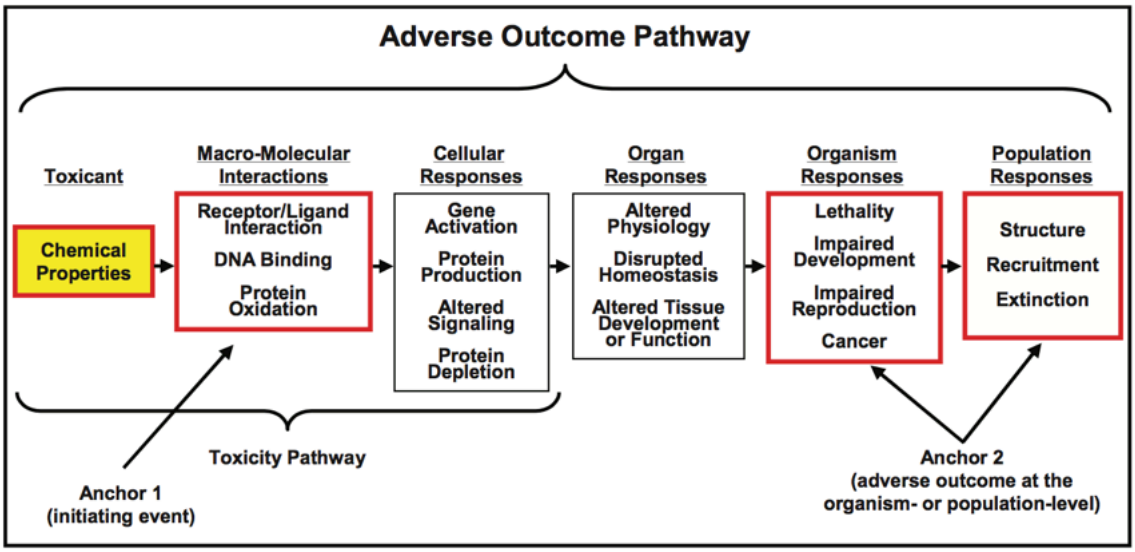


Figure 1.1: Conceptual framework of an AOP. A Molecular Initiating Event (MIE) is characterized by the interaction of the chemical with its target. A series of Key Events (KE) lead to an Adverse Outcome (AO) at organism or population level. Picture taken from Ankley et al.[26].

KEs have to be measurable and quantifiable and are used to test AOP hypothesis. Developing a robust, testable and functional AOP requires the generation of data within the laboratory. *In vivo* tests do not represent a good method for generating the data. *In vivo* endpoints are not amenable for direct prediction by in silico methods and *In vivo* effects are the result of the interaction between several different factors contrarily to *In vitro* tests which usually just measure one or very few of these factors[27]. Pathways information, crucial for the development of AOPs, will be inferred using the data generated. Depending on the final use of the AOP, the knowledge required will vary. For example, qualitative relationships are adequate for hypothesis generation but quantitative relationships will be needed for risk assessment. Studies aiming to develop AOPs are currently increasing. Recently, Song et al. proposed the first invertebrate ED-

AOP investigating the effects of endocrine disrupting chemicals (EDs) on arthropod molting[28]. Perkins et al., review four case studies to explore the degree of scientific confidence required for an AOP to be useful in regulatory applications[29]. Lee et al., demonstrated the usefulness of AOPs in biomarker-based Environmental Risk Assessment (ERA)[30].

1.4 Biological systems of relevance

Animal-based tests have always been the standard for assessing biological response either during diseases or chemical exposures. Especially in toxicology, testing is carried out by exposing the animal to a given chemical compound. Adverse effects triggered by these compounds can be displayed as result of acute (less than 1 month), sub-chronic (between 1 and 3 months) or chronic exposure (more than 3 months). Most of these tests examine specific types of adverse effect, known as endpoints, such as death, developmental abnormalities, behavioural changes, etc. Most of the toxicity assessments, especially in human, have been conducted using rats (*Rattus norvegicus*) or mice (*Mus musculus*). In the field of environmental toxicology *Danio rerio* (zebrafish), *Xenopus laevis* and *Daphnia magna* are of particular importance. In the next sections I will outline species that have been the bases of the work described in this thesis.

1.4.1 *Danio rerio*

The freshwater fish zebrafish (*Danio rerio)* is one of the most used model organisms in scientific research. As a model species presents advantages which makes it suitable in many different field of study. First of all, it has a rapid rate of reproduction: females usually lay hundreds of eggs each week which will reach adulthood in about three months. This is quite useful when introducing novel genetic modifications as several generations are required to achieve a stable modification. Moreover, zebrafish is a cost-effective species: a small amount of space for individual is needed, it is relatively inexpensive to feed and it is straightforward to ship samples across laboratories.

The zebrafish genome has been fully sequenced which is an important point for model organisms nowadays. This effort has showed that 70% of human genes have a zebrafish equivalent (84% when considering human disease-causing genes)[31]. These discoveries make the zebrafish a really suitable model for human disease investigation[32,33]. Zebrafish embryos are relatively large and transparent during early stages which makes their manipulation straightforward. This feature allows scientists to easily alter embryo properties, for example through the CRISPR-Cas9 technique which allow rapid phenotyping assessment of gene silencing in embryos[34,35]. Moreover, embryos are permeable to many environmentally relevant chemicals and drugs which makes them particularly suitable for screening a large number of toxicological samples or drugs candidates[36,37]. Particularly, the high sensitivity of zebrafish aids in monitoring environmental contaminants[38,39]. In May 2013, the use of zebrafish embryo for acute toxicity testing (FET assay) has been approved by the Working Group of the National Coordinators (WNT) of the OECD Test Guideline Program and published as OECD test guideline (TG) no. 236 on July 26, 2013[40]. Given its excellent correlation with conventional *in vivo* fish testing[41–43], the test is now accepted as a full replacement for the acute fish test[44].

1.4.2 *Daphnia magna*

The freshwater water flea *Daphnia magna* is one of the oldest system used in biological research[45]. *D. magna* is a small planktonic crustacean that grows up to 5 mm and inhabits a wide range of freshwater environments. They occur in a highly diverse set of habitats ranging from freshwater lakes to saline ponds. As a result, they manifest extensive phenotypic diversity providing ample raw material to study gene function and genome by environmental interactions[46]. The phylogenetic position of *Daphnia* is ideal for comparative genomics. The number of model systems available for understanding genome function and evolution is rapidly expanding. This rapid growth is particularly noticeable in insect taxa. The close relationship of

the crustacean and insects is clearly supported by both molecular and morphological studies making *D. magna* a valuable outgroup for comparative genomic studies[47,48]. *D. magna* are an exceptional model for studying developmental and disease processes since they are transparent throughout life allowing for studies of tissue-specific gene expression at any life stage and direct observation of parasites and pathogens[49]. The reproductive cycle of *D. magna* is ideal for experimental genetics. Generation time in the lab rivals that of almost all other model eukaryotic systems. *D. magna* are easily cultured and maturation is reached within 5 - 10 days. Reproduction is normally clonal (allowing the maintenance of genetic lineages), but sexual one can be induced environmentally (allowing the production of inbred or outbred lineages). The clonal nature of the organism provides an exceptional opportunity to study genetic responses to environmental stimuli in a defined and constant genetic background with unlimited replication[50]. Moreover, Daphnia-based assays have been widely used for biomonitoring water quality. Le et al., provided a good overview of using *D. magna* in aquatic toxicological monitoring depending on omic approaches[51]. The Daphnia Genomic Consortium (DGC)[52] has already released the genome of *Daphnia pulex*[53], a close relative of *D. magna.* The study identified around 30,000 genes in the small *D. pulex* genome (only 200 megabases) as a result of an elevated rate of gene duplication and many of them are attributed to lineage-specific gene families. Moreover, more than a third of the genes have no detectable homologs in any other species. A genome draft of *D. magna* has been also released and it is publicly available in the interactive Daphnia genome database at wFleaBase.org[54]. Studies investigating transcriptomic response as a result of exposure to environmentally relevant compounds have already been performed[55].

1.4.3 *Micropterus salmoides*

The largemouth bass (LMB) (*Micropterus salmoides*) is widely distributed along the U.S.A. east coast and it has an important economic value due to its popularity as a sport fish and it is also ecologically relevant due to its trophic position in the freshwater environment as apex predator. For this reason, it has been often used as sentinel species for assessing chemical toxicity[56,57]. Moreover, it has been widely cultured in the world, especially in China where it has a rapid production growth[58]. Recently, the muscle transcriptome of LMB has been assembled to investigate genes and SNPs associated with growth traits[59]. Byadgi et al.[60], assembled a transcriptome starting from spleen tissue to investigate the molecular response of LMB when hosting a bacterial infection by *Nocardia seriolae*. Most of the studies however have been addressing questions linked to ovary development and the effects chemical compounds exert on it. Dominguez et al. demonstrated the expression of the VTG receptor to be stage-specific and to be controlled by insulin and sex steroids[61]. Martyniuk at al.[62], identified genes related to reproduction (granulosa function and oocyte development), endocrine function (steroid metabolism and hormone biosynthesis) and immune function (T cell suppression and leukocyte accumulation) to be differentially expressed in female LMB sampled from a highly polluted site. Collì-Dulà et al.[63], investigated the effects of perfluorinated chemicals (PFASs) in the liver and testis of male LMB and found out genes related to lipid metabolism, energy production, RNA processing, protein production/degradation and contaminant detoxification to be modulated by high concentrations of PFASs. The work presented in chapter 4 uses part of the data coming from Martyniuk et al, where they investigated the gene expression network underlying Largemouth bass ovarian development characterizing the molecular pathways involved in oocyte maturation[8].

1.4.4 Other commonly used species

In addition to the species outlined above, many other are frequently used for toxicity testing. In ecotoxicology different species of fish are used worldwide to investigate chemicals impact in their natural environment. The European flounder (*Platichthys flesus*) has been used by Williams et al. to successfully characterize chronic molecular responses to environmental mixtures[64].

The stickleback (*Gasterosteus aculeatus*) was successfully used by Petersen et al. to investigate the effect of the perchlorate on the embryonic androgen synthesis and more in general on reproductive development[65].

Another frequently used organism in the ecotoxicology field is the fathead minnow (*Pimephales promelas*). Cavallin et al, demonstrated endocrine disrupting chemicals coming from wastewater treatment plants (WWTP) to affect fathead minnow reproduction in a dose-dependent manner[66].

1.4.5 Alternatives to animal testing

One of the biggest challenges the toxicology community is trying to address is the reduction of toxicity testing using whole animals[67]. Each year, more than 100 million animals are killed in U.S. laboratories for biology lessons, medical training, curiosity-driven experimentation, and chemical, drug, food, and cosmetics testing. In the UK, Home Office statistics shows that 2.08 million experimental procedure were undertaken in 2015. In Canada, the Canadian Council of Animal Care (CCAC) reported a total of 3.57 million experimental procedures in 2015. The amount of animals used for toxicity testing is not compatible with the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH) legislation managed by the European Chemicals Agency (ECHA)[68] which is trying to address the issue linked to chemical impact on both human health and environment by reducing the use of whole animal experimental procedures.

The agreement signed in 2009 by United Nations, Canada, Japan and the European Union posed the basis for the development of alternative methods to assess the safety of industrial chemicals, consumer products and pharmaceuticals. The shared goal is summarized by the principles of the 3Rs framework proposed by the UK National Centre for the Replacement, Refinement and Reduction of Animal in Research (NC3Rs)[69] aiming to advocate scientific ingenuity to Replace, Reduce and Refine the use of animal (3Rs) without compromising scientific rigour[70]. Two main strategies were adopted for addressing this need: the development of *in vitro* system (i.e. cell culture) as a type of non-animal testing and the development of *in silico* methodologies able to generate toxicological prediction based on chemical properties (i. e. QSAR – Quantitative Structure-Activity Relationship). Sewell et al., provide a good outline of current and future opportunities to apply the 3Rs in safety assessment programs for pharmaceuticals and potential benefits to the industry[71]. The use of *in vitro* cell based system and *in silico* approaches is an active area of research in many different fields of study and their use as alternative to animal testing by pharmaceutical industries is fast developing[72]. In environmental toxicology, the use of fish cell lines (i.e. fish gill cell line) is increasing since they have many potential applications[73]. Tanneberger et al. assessed the toxicity of 35 chemical compounds and found out a good agreement between *in vivo* and *in vitro* effective concentrations[74]. Yazdani et al., showed that BPA exposures were able to induce oxidative stress and immune-related dysfunctions on the Atlantic salmon kidney cell-line[75]. A good overview of alternatives to animal testing is given by Scholz et al.[76]. In chapter 2 and 3 we present a work based on a trout fish gill cell line[77]. *In-silico* methods aims to complement existing toxicity tests to predict toxicity, prioritize chemicals and guide toxicity tests. Sangion and Gramatica recently developed QSAR models to predict acute toxicity of Active Pharmaceutical Ingredients (APIs) in key organisms belonging to different

aquatic trophic levels[78]. A comprehensive overview, explaining and comparing the strength and the weakness of the existing *in silico* methods for toxicity prediction is given by Raies and Bajic[79].

A good overview of the full scope of the 3Rs framework is given by Burden et al. which present examples of short, medium and long-term approaches to enable the uptake of the 3Rs across regulatory toxicity testing[80].

1.5 Omics technologies

The advent of omics technologies has revolutionized the way research is done. Differently from traditional studies which are purely hypothesis-driven, omics experiments are hypothesis-generating where all data are acquired and needed to define a hypothesis that can be further validated. Omics technologies span from genome sequence (Genomics) to expression (Transcriptomics, Proteomics and Metabolomics) and its regulation (Epigenomics, Cistromics and miRNAomics). These high-throughput technologies can simultaneously measure thousands of features in a single experiment and have many different applications (fig.1.2).

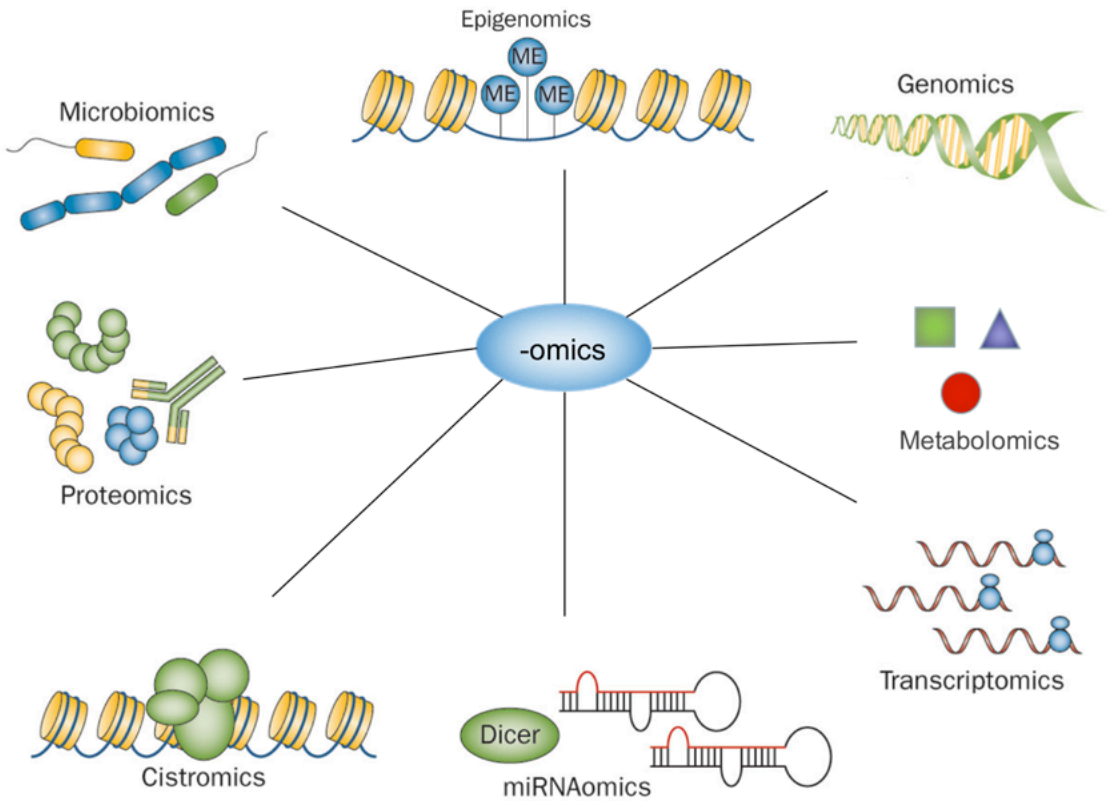


Figure 1.2: Omics technologies are rapidly evolving providing scientist with a wide range of technologies to carry out multi-level studies. Picture adapted from Weaver et al.[81].

As the amount of data generated is considerable, data analysis is really challenging and a lot of effort have been put into the development of suitable tools that help in the extrapolation of biological insight from quantitative measurements. Here we describe the main functional genomics technologies and review the most important approaches for data analysis, including data-driven pathway inference algorithms.

1.5.1 Transcriptomics

The transcriptome is defined as the collection of all mature RNAs in a given cell. The analysis of the transcriptome can help unravelling underlining processes within a cell. The early development of microarray technologies and then later of mRNA sequencing has made possible

to measure the whole transcriptome of a cell or a tissue in a single experiment. Most of the data available in the public domain has been developed using microarray technologies but mRNA sequencing is now the technology of choice for transcriptomics profiling.

Microarray Technologies

Most of the transcriptomics data available in the public domain has been generated with either Agilent or Affymetrix technology platforms. The general principle behind microarray technology is hybridization between two strands of DNA or RNA. Each array is a collection of microscopic spots, called features, containing thousands of copies of specific probes (gene sequences), usually representing the full or partial transcriptome of a given organism, attached to a solid surface. Fluorophore-labelled cDNA samples hybridize with the probes and their hybridization is then detected and quantified to determine relative abundance of nucleic acid sequences in the target.

Affymetrix arrays are built using a photolithographic synthesis process[82] that produces a high-density pattern of short oligonucleotide sequences (25 nucleotides in length) . Each of the gene transcripts is represented by oligonucleotides complementary to different exonic gene regions.

Agilent, a Hewlett-Packard (HP) offshoot, developed an innovative ink-jet technology to print sequences on a small glass surface. This method simply print nucleotide bases on a glass surface with micron precision to synthetize oligonucleotides. These can be up to 120 nucleotides long. This design gives the customer the possibility to print custom designed microarrays at no extra cost, a particularly useful solution for non-model species.

Next Generation Sequencing (NGS) technologies

Along with these omics technologies, next-generation sequencing technologies have quickly developed and cost have significantly decreased making them accessible by most of the laboratories worldwide. While the RNA molecules of interest need to be known previously in designing microarrays, NGS do not rely on previous knowledge of the RNA molecules. Here DNA or RNA molecules are sequenced by reading each base and reconstructing either the genome (DNA-seq) or the transcriptome (RNA-seq). Sequencing offers many advantages over microarray expression profiling technology[83]. Microarray gene expression measurements are limited by background at the low end and signal saturation at the high end. Sequencing technologies, on the other hand, offers a broader dynamic range by quantifying discrete read counts allowing the detection of very lowly expressed genes. Moreover, sequencing technology is particularly suitable for non-model species allowing the generation of the full transcriptome[84]. The challenge still lies in the analysis of the high amount of data these technologies are able to generate.

As for microarrays, molecules of mRNA are first reverse-transcribed into cDNA. Those molecules are then fragmented into smaller sequences and platform-specific adapters are attached to one or both ends. Those modified cDNA molecules are then sequenced and the resulting short sequences called “reads” are first assembled into longer fragments and finally aligned against a reference transcriptome. For non-model species it is possible to assemble the reads *de novo* to produce a new genome. The number of reads sequenced for each gene sequence reflects the abundance of the mRNA molecules in that particular tissue in the given experimental condition.

1.5.2 Proteomics

Proteomics is the study of the entire set of proteins produced by an organism. Proteomic studies aiming to map out the structure of protein complexes are known as structural proteomics and take advantage of X-ray crystallography and nuclear magnetic resonance spectroscopy (NMR). On the other hand, the quantitative study of protein expression between samples in different

experimental conditions is known as expression proteomics and use technologies as 2D-PAGE and mass spectrometry (MS). Proteomics offers the potential to be more informative of the molecular state of a particular cell or tissue than mRNA levels as proteins are translated products of gene expression and post-transcriptional modifications may reduce the amount of mRNA successfully translated into proteins. Proteomics studies have been successfully applied in many different fields and a significance contribution has been given to the ecotoxicology one[85,86].

1.5.3 Metabolomics

Metabolomics provide an overview of the full set of metabolites, intermediate or final product of metabolism, present in an organism under a given experimental condition. Differently from transcriptomics and proteomics, metabolomics provides information about whether a given molecular process has happened. Metabolomics studies can be categorized into untargeted and targeted analysis: Untargeted analysis aims to define the metabolic profile of the total complement of metabolites ("fingerprint") in a sample and take advantage of the NMR while targeted analysis focuses on the identification and quantification of selected metabolites and take advantage of MS.

NMR spectroscopy can be used with either liquid[87], gas phase[88] and tissue samples[89] and it can be used to investigate the chemical and physical properties of molecules such as molecular dynamics[90] and electron density[91]. One of the major disadvantages of NMR is its low sensitivity. On the other hand, MS is characterized by a high sensitivity and selectivity which make it the best analytical platform for profiling metabolites in mixed biological samples. However, one of the disadvantages of MS is that different methods have to be used to detect all classes of metabolites. A good overview of both the strength and weakness of NMR spectroscopy and MS applications in metabolomics research is given by Emwas[92]. Metabolomics is still in the early

stages of development however, it has already started to give relevant insight in different fields of research and recently interesting results in the field of ecotoxicology have been achieved[93,94].

1.6 Data acquisition

As the projects embedded in this dissertation all uses microarray technologies, I will briefly describe a typical sample-to-data workflow using Agilent technology (used throughout this thesis).

Figure 1.3 shows a schematic representation of a typical sample-to-data workflow using Agilent technology. Initially, the RNA or mRNA is extracted from the biological sample of interest using either the most recent spin column-based method or the more traditional phenol-chloroform based one[95,96]. Next, an oligo-dT primer is used to hybridize with the poly(A) tail of the desired RNA molecule. Then, a reverse transcriptase enzyme along with deoxynucleotides are added to the mixture allowing the creation of cDNA molecules based on the original sample. This initial linear mRNA amplification step reduces the amount of total RNA needed as input to as little as 25 ng. A T7 RNA polymerase in the presence of fluorescence dyes (Cyanine-5 or Cyanine-3) convert cDNA molecules into complimentary RNA (cRNA). The resulting cRNA is fragmented into shorter and less structured segments before the hybridization to the array which occurs overnight. The fragmentation step allow the reduction of structural effects as secondary and tertiary structures can significantly affect the hybridization efficiency[97]. Given the different wavelength emissions of the two dyes, double channel experiments are often used to represent the ratio of each gene given a treated and a reference sample. Once the unbound cRNA has been washed off, a laser-scanning machine capture the fluorescence signal from the labelled cRNA. The amount of fluorescence signal reflects the transcript abundance within the given cell or tissue. Finally, an image analysis procedure is used to extract a numeric representation of each

probe on the array based on pixel intensities. Once a numeric data table has been acquired, downstream analysis can start.

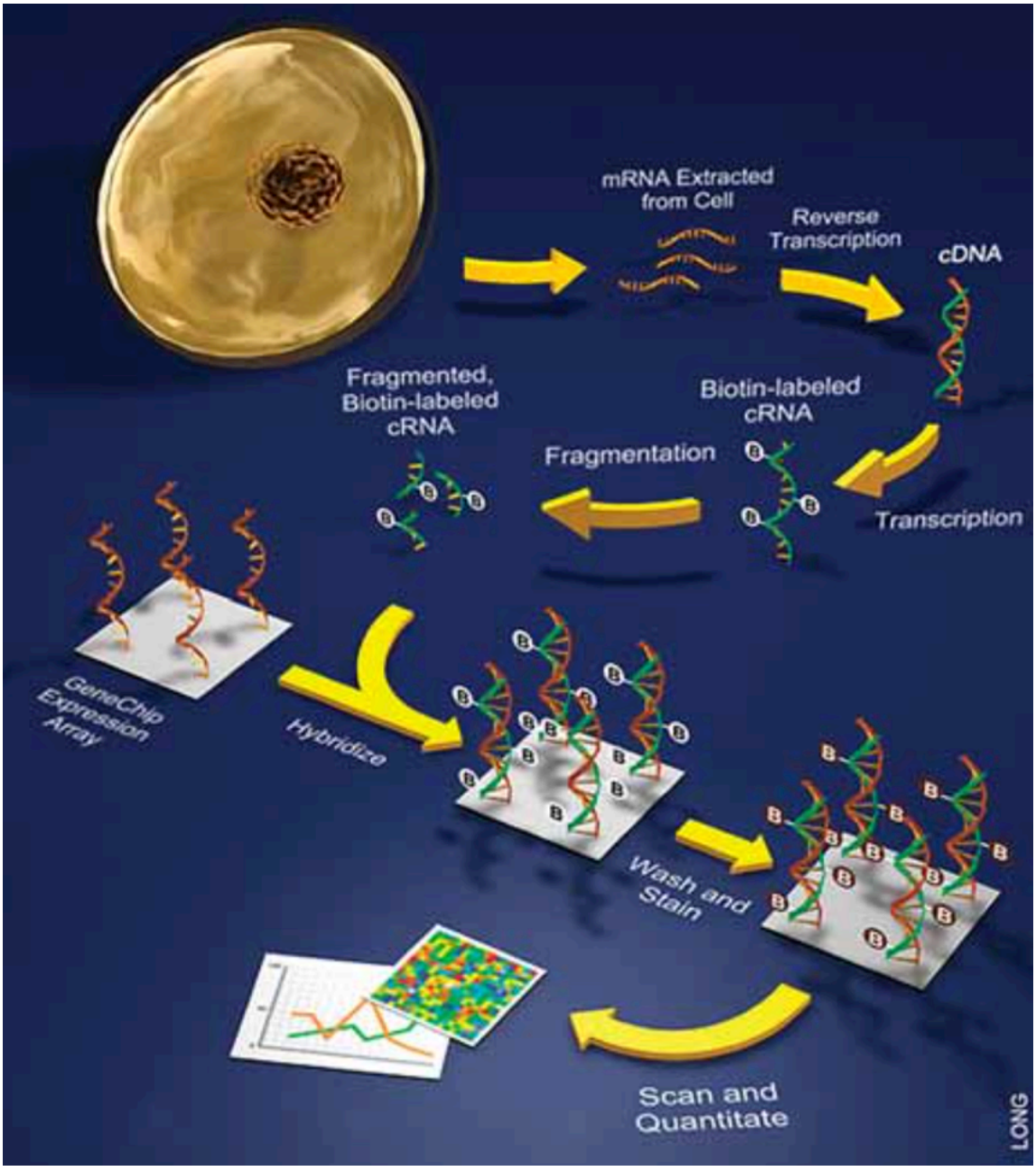


Figure 1.3: Schematic representation of the key steps of a typical sample-to-data workflow with microarray technologies. Picture taken by The SCQ[98].

1.7 Microarray data analysis

Microarray technology allows measurements of thousands of features that require advanced computational methodologies to extrapolate biological meaning out of them. In the last decades a wide range of computational tools for the analysis of omics data have become available offering statistical or visual results that help in the extrapolation of biological insights. Microarray data analysis always starts with raw data pre-processing and normalization to make the data ready for downstream analysis. Once the dataset has been normalized many different questions can be addressed using different approaches, of which differential gene expression analysis, exploratory data analysis, functional analysis, network inference and class prediction techniques are the most important.

1.7.1 Data pre-processing and normalization

The overall aim of microarray data processing is to remove noise and systematic technical variability to reveal the true biological differences between samples. RNA extraction, labelling, hybridization and scanning can introduce systematic errors. Normalization algorithms are used to minimize this variability. Usually, a $\log_2$ transformation is applied to the data to remove low signal intensity bias. This process is required because the distribution of raw signal intensity values is skewed towards the low signal intensity and the process forces values to be spread more evenly across the intensity range. Moreover, most of the statistical tests developed are parametric and requires the data to follow a normal distribution. Following the transformation, a normalization procedure is applied to the data. Many different normalization procedures have been developed but they all share the same goal to try to reduce the unwanted variability while preserving the biological heterogeneity[99]. The choice of method is dependent on the design of the array probes, the number of channels and the manufacturer. Moreover, normalization procedures are usually different for one-color arrays compared to two-color arrays. Two-color

arrays usually require a 'within-array' normalization to account for dye biases before the 'between-array' normalization. LOESS normalization[100] used in chapter 2, is a widely used method for two-color array normalization likely due to it being easy to implement. It does fit a smooth curve through a set of data points to build up a function that identify the variation in the data. Quantile normalization proposed by Bolstad et al.[101] used in chapter 2 is a popular method due to it being mathematically simple. It does force each channel to share the same empirical distribution.

Microarrays have several millions of features which allow for multiple probes per target gene. This is mainly due to the availability of multiple sequences for the same target gene. A summarization procedure is applied to produce a unique measure for every gene[102]. Two different methods to summarize multiple probes are commonly preferred: taking the average expression values of all the probes or select the one with the highest intensity value as features with low expression values are considered highly variable across biological replicates. An additional filtering step to remove lowly expressed genes may be also applied. One of the key assumption of microarray technology is that only a few genes (up to a few hundred) are expressed at different levels between samples. This means that most of the genes have identical expression across samples. For this reason, in order to increase sensitivity in finding differentially expressed genes, genes whose expression fall below a given detection limit can be removed. The resulting dataset is now ready for downstream analysis as outlined in the next sections.

1.7.2 Exploratory data analysis

Exploratory data analysis techniques encompass a wide range of tools which are important for the detection of outliers and for the identification of trends between genes or samples. Moreover, microarrays as well as the other omics technologies are characterized by a high dimensionality which cannot be visualized by the human eye. For this reason, different data

reduction techniques have been developed to provide a visual representation of the similarity or dissimilarity between genes or samples. Among the different techniques developed, principal component analysis (PCA) is probably the most widely used for the identification of dominating patterns in multidimensional data sets[103,104]. Moreover, it has also been successfully applied for the detection of outliers in multidimensional data sets[105,106]. PCA is a projection technique that aims at representing a highly multidimensional dataset into a small number (typically a maximum of 3) of principal components with minimal loss of information. This allows the visual inspection of the data and the identification of the variables mostly contributing to the variation in the data. The variance across the different samples is summarized into principal components where the first and the last component contain the most and the least of the variance of the original data set respectively. Importantly, the different principal components (PCs) are not correlated to each other ensuring they represent different features of the original dataset[107].

Other commonly used techniques for representing the relative similarity between genes and between samples are clustering algorithms, where some of them use a tree based representation[108,109]. Clustering methods use a dissimilarity matrix created by applying simple distance measures (Euclidean, Pearson or Spearman correlation, etc.) and then use it to identify groups of variables that have a similar expression profile. Traditional clustering methods can be classified into two main categories: hierarchical and non-hierarchical algorithms. Hierarchical algorithms groups similar objects together and provide a natural way for graphical representation which is called dendrogram where each branch forms groups of genes (or samples) sharing similar expression patterns. Classical hierarchical clustering methods lack robustness when dealing with noisy data. Self-Organising Maps (SOM), clustering approaches belonging to the family of the neural networks, have been frequently used as an alternative as they are more robust and accurate in clustering noisy data. However, SOM lack of the ability to

detect higher order relationships between clusters which is typical of hierarchical clustering methods. The Self-Organising Tree Algorithm (SOTA), which we employed in chapter 4, combine both the advantages of SOM and hierarchical clustering providing a hierarchical clustering achieved with the accuracy of a neural network[110]. In SOTA Kohonen's self-organising map (SOM)[111] is coupled with Fritzke's growing cell structures[112]. The method starts with a mother neuron which include all the data and after a training procedure as implemented in the SOM, the data are split into two "sons" neurons. Then, the most heterogeneous of these two neurons splits again and the training re-starts. This splitting scheme uses a binary tree topology and it is possible to stop the growth of the hierarchy at the desired level of variability which allows a better visualization of the different patterns. Non-Hierarchical algorithms (i.e. partitional), on the other hand, perform a partition of n genes (or samples) into K clusters where each cluster contains features with similar expression profiles. One of the most common non-hierarchical clustering methods is K-means[113]. In K-means, K random centroids are put into your data space. For each individual point (genes) the nearest centroid is found. The centroids are then re-computed by averaging all the points associated to that given centroid. The nearest centroid for each point is again found. The process is repeated as long as the centroid cannot be moved further. A good overview of clustering methods for microarray data is given by Belacel et al.[114].

1.7.3 Differential gene expression

Identifying genes whose expression is significantly different between experimental conditions is one of the most common question in the analysis of omics data. Among the first statistical approaches developed for the identification of differentially expressed genes there is the traditional t-test which compare the sample means of up to two classes specifying a p-value for each gene. However, there are issue when this is applied to omics datasets. One is associated with the data distribution as the t-test assumes that the data are normally distributed and omics

data do not always follow a gaussian distribution. Moreover, the t-test needs a procedure for correcting for multiple testing. With all statistical tests, is highly important to control the Type 1 error rate. This error controls the number of false positive within the statistical test. The problem arises when multiple comparisons are made. Several correction methods have been proposed in order to control Type 1 error rate[115]. The most commonly used was proposed by Benjamini and Hochberg[116] which they named the false discovery rate (FDR). It has been particularly designed to control the expected proportion of false discoveries. An FDR of 1%, for example, would be able to yield 99% of true and 1% of false positives. Significance analysis of Microarrays (SAM) address these limitations by using an improved statistics and by using a non-parametric method for computing the FDR[117]. SAM is a permutation-based method which use the d-statistic to identify genes differentially expressed between different experimental conditions giving estimates of FDR. Moreover, it has been optimized for many different experimental designs as paired, unpaired and time-course. Another widely used method is the analysis of variance (ANOVA), which is similar to the t-test but it allows the identification of features that are linked to specific factors affecting the experimental data. These factors can be either biological (i.e. different treatments) or technical (factors that may affect the results, i.e. different hybridization days in a typical microarray experiment).

1.7.4 Functional analysis

Differential gene expression analysis used in combination with cluster analysis and other data exploration techniques are able to identify genes whose expression is linked to specific experimental or environmental condition and to organise them in groups with similar expression profiles. While this is extremely useful, it is challenging to make biological sense of large gene lists. This challenge is addressed by functional annotation tools that test whether genes belonging to specific functional terms or biological pathways are over-represented in a given gee

list. The Database for Annotation, Visualization and Integrated Discovery (DAVID) is probably the most popular functional analysis tool for omics data [118]. One of the key functions of DAVID is the ability to integrate multiple functional annotation databases as Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) providing biological insight at multiple level of biological organization (either genes or pathways). DAVID implements an algorithm to identify functional terms that share a high percentage of genes and group them together to simplify the task of biological interpretation. The EASE score, a modified Fisher exact p-value, is used to measure gene-enrichment in annotation terms. Moreover, different methods have been implemented within the DAVID tools such as functional clustering, functional annotation chart and functional annotation table. Other commonly used functional annotation tools are FatiGO[119] and gProfiler[120]. Gene Set Enrichment Analysis (GSEA)[121] has been shown able to identify statistically significant functional enrichment even when not a single genes is found to be differentially expressed by conventional hypothesis testing statistical procedures. The method works by first ranking all the genes in the transcriptome depending on their differential expression between different phenotypes. The goal of GSEA is to determine whether members of a functionally defined given gene set (e.g. genes in a biological pathway) are enriched at the top or at the bottom of the ranked list, in which case the gene set is correlated with that phenotypic class. An enrichment score (ES), which represent the degree to which a given gene set is overrepresented at the extremes of the ranked list, is then calculated for each of the gene set. Finally, the probability values are adjusted for multiple testing using FDR. An overview of the GSEA methodology is reported in fig. 1.4.

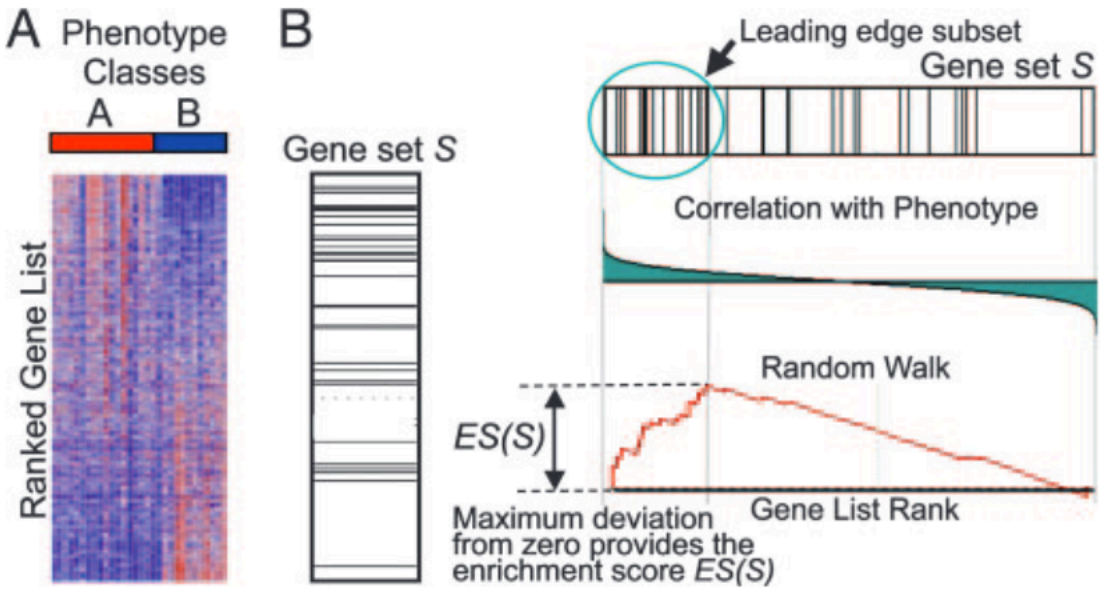


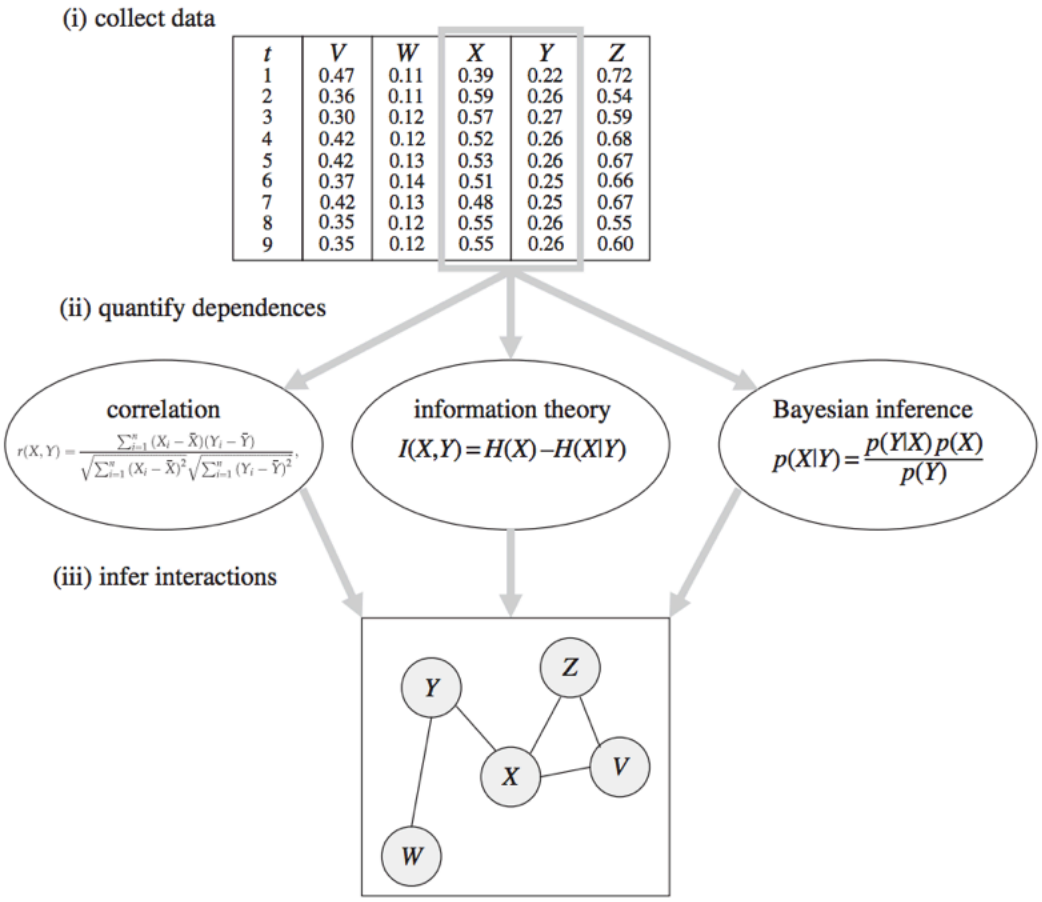
Figure 1.4: Schematic overview of the GSEA method. A) The expression dataset is sorted by correlation with the phenotype. B) Plot showing the calculation of the enrichment score. Picture taken from Subramanian et al.[121].

One of the weakness linked to these functional annotation tools is the inability to remove redundant terms. As previously reported, DAVID is able to deal with that however most of the classical annotation tools do not have an in-built function that address this issue. The Reduce and Visualize Gene Ontology (REVIGO) tool has been developed to specifically address this issue[122]. REVIGO identifies representative subset of terms using a clustering algorithm that relies on semantic similarity measures.

1.7.5 Network inference

Omics technologies are capable of measuring thousands of features representing the molecular state of a given biological system in a particular experimental condition. In transcriptomics, proteomics and metabolomics these features are represented by mRNA, proteins and metabolites respectively. Since transcription, translation and metabolism are integrated within the hierarchy of information leading from genotype to phenotype, it is reasonable to hypothesise

that the relationship between genes of similar functions may be learned from observational data[123,124]. In the last ten years, several research groups have worked at developing reverse-engineering approaches that can address precisely the above-mentioned challenge. Several methods have been developed to reverse engineer molecular pathways to represent biological pathways both as static as well as dynamical systems. The most common reverse-engineering approaches can be classified into three different classes: correlation-based, mutual information-based and Bayesian (fig. 1.5).



t	V	W	X	Y	Z
1	0.47	0.11	0.39	0.22	0.72
2	0.36	0.11	0.59	0.26	0.54
3	0.30	0.12	0.57	0.27	0.59
4	0.42	0.12	0.52	0.26	0.68
5	0.42	0.13	0.53	0.26	0.67
6	0.37	0.14	0.51	0.25	0.66
7	0.42	0.13	0.48	0.25	0.67
8	0.35	0.12	0.55	0.26	0.55
9	0.35	0.12	0.55	0.26	0.60

Figure 1.5: Schematic representation of the process to infer interaction networks applying three different reverse-engineering approaches: correlation, mutual information and Bayesian. Picture taken from[125].

Correlation-based methods for the reconstruction of genetic networks have been successfully applied[126]. These methods represent the right choice for measuring linear relationships however, their accuracy decrease when measuring non-linear interactions as spearman correlation coefficient is only able to spot monotone relationships. Mutual information, defined by Shannon[127], represent a more general measure with the ability to identify non-monotonic relationships. It measures the amount of information a variable contains about another applying the concept of the entropy which represent the uncertainty of a variable. The stronger the interaction between two variables the larger the value of mutual information; mutual information of unrelated variables will be theoretically zero. One of the most popular mutual information-based approaches for the inference of gene regulatory networks is the Algorithm for the Reconstruction of Accurate Cellular Networks (ARACNe)[128]. ARACNe predicts potential functional association between features (genes, proteins or metabolites) using MI. The power of the algorithm is coming by the exploitation of the Data Processing Inequality (DPI) to discard indirect interactions[129]. It reduces the number of false positive by examining gene triplets that have a significant value of MI and removing the edges with the smallest value. A time-delay version of ARACNe accounting for time-course data has also been developed[130]. One of the advantages of the MI-based algorithms is the low computational cost as the MI is calculated between pairs of features at a time hence making this tool suitable to deal with the huge amount of data generated by omics technologies. MI methods have been criticized for their inability to infer direction of regulations[131]. However, some improvements have been made and nowadays methods capable of recovering directions have been developed[132]. The third class of reverse-engineering approaches for network reconstruction use Bayesian inference. It uses probability as a measure of uncertainty allowing the incorporation of prior knowledge in order to update the probability as more information becomes available. Bayesian networks are probabilistic

graphical models that use Directed Acyclic Graph (DAG), a finite graph with no directed cycles, to fit the data. They present a few advantages as their ability to deal with small and incomplete datasets[133,134] and the possibility to incorporate prior knowledge[135]. An important limitation is that most of the biological networks contains loops and Bayesian networks, being acyclic, do not take them into account. Dynamic Bayesian networks have been developed to overcome this issue as they can include cycles and are particularly useful when time-course data are available[136,137]. The downside of these reverse engineering approaches is that they can be computationally expensive and they require a lot of measurements for the same node (>50 with ARACNe and >100 with Bayesian networks). Moreover, the lack of genome-wide measurements of factors we cannot collect measure for and the dynamics that elude cost-effective experimental designs still present some limitations[138]. Nevertheless, network inference approaches have been successfully applied in different fields of science providing new insights about the mechanism of action of both chemical exposures[8] and diseases[139].

1.7.6 Machine learning

Machine learning is a field of computer science that develops algorithms that can learn from data rules to predict unknown future events. A Machine learning workflow consist of three main phases including building the model from input data, evaluating and tuning the model and finally using the model for prediction-making[140]. These methods can be classified into three broad categories: supervised learning, unsupervised learning and reinforcement learning. In supervised learning, the system infers a function from labelled training data which is used to predict the value of a variable, called dependent variable, from a set of variables, called independent variables. In supervised learning, there are two different kind of tasks: regression and classification. Regression models predict numerical values while classification models try to predict distinct classes. Among the most common techniques, we find decision trees, rule

learning and instance-based learning such as *k*-Nearest Neighbors (*k*-NN)[141], Genetic Algorithms (GA), Artificial Neural Networks (ANN) and Support Vector Machines (SVM). *k*-NN, which we employed in chapter 2, is a simple and intuitive method widely used for classification problems. It offers many advantages over the other method as it being non-parametric and its low error rate. The classification algorithm finds the K nearest neighbours of a sample of unknown category by computing samples similarities (i.e. Euclidean distance). A given class is then assigned depending on the class majority of the K nearest neighbours. The choice of the K value is arbitrary its value should be odd for a two-class problem and must not be a multiple of the number of classes in order to avoid ties. Historically, the best K value is between 3 and 10.

In unsupervised learning, on the other hand, the system tries to discover hidden data structure or data association from unlabelled data. In unsupervised learning, there is no evaluation of the model accuracy as the data given to the learner are unlabelled. Some of the most common methods include clustering approaches, ANN and approaches for learning latent variable models as PCA. In reinforcement learning, the system tries to learn through direct interaction with the environment to maximize some notion of cumulative reward[142]. The most common approaches for reinforcement learning belong to the family of the Monte Carlo methods.

1.7.7 Variable selection

With the development of functional genomics technologies able to generate very large datasets, the selection of genes whose profile is associated with the known sample phenotype has become a challenging problem. Identifying subset of features with the ability to discriminate between samples type would allow generating new hypothesis, selecting biomarkers, acquiring more knowledge about MoA and choosing potential drug targets[143,144]. Methods able to "test" whether genes are related to sample phenotype can be subdivided in univariate and multivariate methods. Univariate methods test each variable at the time for its ability to distinguish between

group of samples. These methods first identify differentially expressed genes between group of samples and then use the ones most differentially expressed to build a statistical model. One of the most common univariate variable selection methods is PAMR[145]. As genes works together in the context of complex interconnected pathways it is their behaviour as a group that can be predictive of the phenotypic class and univariate methods do not take this into account. Multivariate selection methods have been developed to overcome this issue since the variables are tested in combination to identify interactions between genes. One limitation associated with multivariate selection methods is the lack of computational resources to evaluate the extremely large number of statistical models they produce since they test different combination of thousands of genes. A good solution is the use of searching algorithm that explore the data looking for better sets of variables. These methods, as Markov Chain Monte Carlo methods and GA, have been already successfully applied[146,147]. Multivariate methods present some pitfalls specially in the way they build the best predictive model. Sainani provide a good explanation of issues associated with variable selection and suggest alternatives to overcome these limitations[148]. GALGO, which we employed in chapter 5, is a method that uses a GA variable selection strategy to develop statistical models from large-scale datasets[149]. Briefly, given a dataset of two classes of samples, the GA is used to search and evolve combination of genes (called chromosome) that distinguish between classes using a classification method. A broad number of models are developed which may differ in gene content but they all have a similar high classification accuracy. Genes that appear multiple times in different models suggest they are important in defining that given phenotype. Then, a forward selection strategy is used to develop a representative model: genes are ranked based on their frequency in the population and the top 50 most present variables are incrementally tested, by adding each variable one by one. The final representative model is chosen as the one with the higher accuracy and the smaller

number of variables. GALGO has been already successfully applied in the field of environmental toxicology[150].

1.7.8 State-of-the-art environmental omics

Risk assessment of chemical exposures has traditionally relied on directly measured apical endpoints by applying *in vivo* toxicity test. These methods were good at investigating the adverse outcome but were not providing any information about molecular mechanisms driving the toxic response. To overcome this issue the AOP framework has been developed[26]. It aims at filling the gap between a MIE and the resulting AO by applying *in vitro* toxicity tests coupled with advanced *in silico* methodologies. In recent years, OECD in collaboration with Environmental Protection Agency (EPA), the European Commission's Joint Research Centre (JRC)[151] and the US Army Engineer Research & Development Center (ERDC)[152] developed the "Adverse Outcome Pathway Knowledge Base" (AOP-KB) to enable scientific community to share, develop and discuss AOP related knowledge. The AOP-KB project includes 4 different platforms: 1) the AOP-wiki[153], which is publicly available, it is an interactive encyclopedia for AOP development which allows users to develop new AOPs; 2) the AOP Xplorer (currently being developed), a computational tool that enables graphical representation of AOPs; 3) the Intermediate Effects DB (currently being developed), a database holding information on how chemical compounds trigger MIE and/or KE; 4) the Effectopedia[154], a modelling platform using modular structure to capture semantically annotated Knowledge. Currently, the AOP-wiki includes 203 AOPs of which 6 endorsed, 18 under review and 178 under development. It is a valuable source as incorporates assembly and evaluation of weight of evidence able to support the casual relationships between KE in the pathway and moreover, when available it can include quantitative understanding. For this reason, the AOP approach offers great potential to aid in the development of computational prediction models in regulatory toxicology. This is possible because AOP focuses on a limited

number of molecular processes hence reducing an initially complex biology which in turn will avoid model overload when building predictive models[155]. Furthermore, the AOP concept may help for assessing the risk of chemical mixtures by integrating the knowledge coming from different KE and AOPs[156]. The application of the AOP framework for risk assessment is however at present challenging. The information on the chain of events linking the MIE to the outcome is in fact missing for many of the biological processes of interest in ecotoxicology and gaining that knowledge can be very time consuming. However, the application of omics technologies and computational learning of biological pathways has provided us with the necessary tools for the rapid discovery of toxicity mechanisms.

1.8 Aim of the thesis

The work presented in this thesis aims at achieving the overarching goal of demonstrating the potential of systems biology and data driven science in identifying novel mechanisms of environmental toxicity and to develop a set of biomarkers for monitoring proposes. This overarching aim has been achieved by the following specific objectives:

1. Identifying specific biomarkers correlated to chemical MoA
2. Address the question whether a cell culture system can be informative of whole organism toxicity
3. Development of a system that allow the identification of novel chemicals with endocrine disruption activity in fish and in an invertebrate species

These objectives were demonstrated using three different biological systems, either model and non-model species. Principle 1 and 2 were demonstrated using both *D. rerio* embryos and a Trout gill cell line (*R. trout*). Results show I can identify biomarkers correlated to a specific chemical

MoA. Furthermore, I demonstrated that *in vitro* molecular signatures linked to compound class are informative of toxicity in a more complex organism hence providing evidence that a trout gill cell line has the potential to replace the zebrafish embryos in toxicity testing. Application of principle 3 was demonstrated using a vertebrate fish species and an invertebrate species. Application of the principle to Largemouth bass (*M. salmoides*) led to the reconstruction of a dynamic model of ovary development. Moreover, using information coming from the CTD database I successfully identified novel compound with the ability to interfere with biological pathways underlying ovary development. The invertebrate species used to demonstrate principle 3 was *D. magna* and results led to the identification of novel compounds with the ability to affect moulting and juvenile hormone pathways.

The work presented in this dissertation offers essential knowledge that can be used to refine further study and to improve risk assessment and environmental monitoring.

Chapter 2: Molecular signatures for the classification of environmentally relevant chemicals: *in vitro* and *in vivo* studies

The whole experimental design was performed by our collaborating labs while I have developed data analysis strategies and performed all the in silico analysis with the exception of the microarray annotation which was achieved by John Herbert.

2.1 Abstract

Alternatives to *in vivo* animal testing in ecotoxicology aim to increase the throughput of chemical safety assessment whilst reducing the number of animals used. The use of *in vitro* systems is more cost-effective, practical and expedient. However, it is still unclear whether such alternative methods provide the level of information gained from the use of a whole-living system. In this chapter, I set to characterize the molecular response following exposure to environmentally relevant chemicals in two alternatives for toxicity testing as the zebrafish embryo and a Rainbow trout gill cell line. By applying high-level computational techniques, I identified and compared *in vitro* and *in vivo* molecular signatures triggered by chemical exposure and demonstrated that these can be indicative of the chemical MoA. I identified biomarkers of compound MoA that can be potentially used to improve ERA. More specifically, I identified glycolysis/gluconeogenesis and steroid biosynthesis to be diagnostic of uncoupler chemicals in zebrafish while ribosome biogenesis was found to be diagnostic of reactive chemicals in trout.

2.2 Introduction

Chemical safety assessment is extremely important in modern industrialised countries. Until recently, this heavily relied on the use of animal testing. In the UK, Home Office statistics for 2015 shows that 2.08 millions of animal procedures were undertaken for toxicological testing, including the testing of chemicals, pharmaceuticals and personal care products, or for monitoring the aquatic environment. Of these, 14% (~294K procedures) involved fish, specifically in conjunction with environmental monitoring within current legislation. Moreover, the use of fish in experimental procedures had increased by 14% (~35K procedures) compared to 2013. With an increasingly stringent legislative context and because of a constant increase in the number of novel chemicals released in the environment, there is a clear need to identify alternatives to animal testing that make the process of safety assessment faster, cheaper, and more ethically acceptable. This has triggered several initiatives to promote the development of alternative methods. In UK, the Research Councils have developed the 3Rs as a framework for animal research[157,158]. The principle of the 3Rs is based on developing methods that avoid or replace the use of animals (Replacement), methods that minimise the number of animals used per experiment (Reduction) and methods that minimise suffering and improve animal welfare (Refinement). This has triggered the development of toxicity testing procedures that rely on cell lines or early-stage fish embryos.

The first decade of the 21st century has seen a major transition from use of whole animal experiments to a mechanism-based assessment based on *in vitro* experimentation. This shift has been driven by some key considerations. First, the concordance between human and animal toxicity has been demonstrated to be poor[159,160]. Second, *in vitro* methods allow tests at more well-defined chemical concentrations. Third, *in vivo* methods cannot account for human variability in response and susceptibility[67]. Lastly, there is the need to reduce the use of animals

in toxicity testing as required by the 3Rs framework. Cell cultures represent a promising tool able to embrace these requirements. The application of vertebrate cells for predicting the toxicity in whole animals, is based on the belief that the interaction of a substance with an organism take place first at the cellular level. Cell culture methods present many technical advantages over the use of *in vivo* methods in reducing the cost for animal care and maintenance, in reducing the amount of chemical needed for a given toxic effect and also the amount of chemical waste produced, in reducing the time needed for testing and thereby increasing the throughput for evaluating multiple chemicals. Moreover, the easier access to dose-response relationships, enzyme activities, and global methods for assessing protein and transcript abundance, all aids the identification of the chemical MoA. However, concerns have been expressed that cell cultures will never be able to fully replace the breadth of information generated from whole animal experiments. This is due to the fact that functional organs *in vivo* exhibit complex relationships and crosstalk from multiple and different cell types and physiological processes are modulated by complex dynamics that are not evident in cell cultures. Also organs are able to compensate for stressful situations that *in vitro* may not be evident[161–163]. Because of these issues, extrapolation of *in vitro* data to the whole integrated animal is very challenging. Moreover, cell lines are also less sensitive to perturbation than whole, integrated organisms[164]. The bioavailability of compounds in a test system has been demonstrated to be reduced when considering lipophilic or volatile compounds, as compounds tend to bind to the serum protein contained in the culture media, and to evaporate, respectively, leading to the underestimation of toxicity[165,166]. Potential solutions to improve cell line sensitivity have been proposed by Schirmer[166]. Fish cell lines, as the RTgill-W1[77], have been widely used for assessing toxicity after exposures to environmentally relevant chemicals[73,167]. Fish cell line are expected to better represent fish phenomena than the more commonly used mammalian cell lines. Moreover, they

do not require CO_2 atmosphere as for mammalian cell lines. Also, they tend to immortalise spontaneously thanks to a relatively high telomerase expression[168]. The reasoning behind the choice of a gill cell type is due to it being the site of oxygen, ion and metabolic waste product exchange with the environment[169]. It is also one of the first organs to be affected by many chemical compounds given that it is the site of toxicant uptake. The potential of the RTgill-W1 cell line has been demonstrated by Tanneberger et al. who showed a good correlation between fish gill cell line EC_{50} values and the LC_{50} values from the fish acute toxicity test[74]. Moreover, Knobel et al., suggested that the RTgill-W1 cell line assay has the potential to serve as an adequate alternative to whole organism toxicity testing[42]. A good overview of the gill cell culture system as a model for aquatic environmental monitoring is given by Bury et al.[167].

The fish embryo acute toxicity test (FET) with zebrafish was originally designed as an alternative to the adult fish acute test[170]. Initially designed as a tool for defining the best range of chemical doses for the definitive *in vivo* test, its robust correlation with *in vivo* fish testing[42,43] led to an OECD recommendation to be an adequate full replacement for the fish acute test[171]. Zebrafish presents many advantages over other animal models. First, they have a short generation time, they are small, fast to develop and mature females typically lay up to 10,000 eggs yearly[172]. Another advantage is represented by the transparency of their early-stages, from fertilization to when the tissue starts to become dense (approximately at 30-72 hpf), which allows the collection of numerous data points. The fact that development occurs entirely outside the body of the mother makes it the most convenient vertebrate system to study toxicant effects on developmental processes. Given these experimental advantages the zebrafish embryo system offers the potential to investigate mechanisms of toxicity[173,174]. However, despite its advantages, there are few points that need to be considered in using the zebrafish FET for screening toxicity. First, chemical compounds with a high molecular weight are not able to pass the chorion prior

to hatch and Braunbeck et al.[175] showed that the barrier function of the chorion increases with the lipophilicity of the chemical compound. To overcome this issue the FET was extended in duration from 48 h to 96 h to encompass the hatching process and to capture effects of those compounds less able to cross the chorion. Another solution has been proposed by Henn and Braunbeck who proposed the removal of the chorion (dechorionation), as a means of improving the sensitivity of fish embryo to toxicant effectss[176]. A second limitation is presented by the presence of solvent vehicles that are able to modify the ease of uptake of the chemicals into the embryo. DMSO, a commonly used solvent, has been shown not to affect chemical uptake into the embryo at a maximum concentration of 0.01% (0.1 mL/L)[177]. A third concern is that neurotoxic compounds have been shown to be less toxic in the embryo than in adult fish[178]. Finally, there is evidence of a limited capacity for biotransformation of compounds by early embryonic stages of zebrafish development, which results in the detoxification to less toxic metabolites and, in some other cases, it can also result in more toxic molecules[179]. With respect to the FET, so far, allyl alcohol is the only compound known to be less toxic in zebrafish embryos due to a lack of bioactivation[42,180]. A good overview of the FET, its origin, application and future consideration is given by Braunbeck et al.[44]

Identifying the MoA for a compound is a crucial step in determining whether the effects we observed in lower vertebrates can be of relevance to humans. Moreover, it would enable the design of *in vitro* assays, read-across and development of suitable biomarkers. New post-genomic technologies, particularly including gene expression profiling, have been identified as important tools to deliver mechanistic information at a global level which undoubtedly increases the impact of environmental risk assessment[181,182]. In this chapter, I compared two alternative approaches to toxicity testing, namely the zebrafish embryo and a fish gill cell line, based on biomarker gene signatures generated by toxic exposure, and which can be predictive of a

compound's MOA. Applying advanced statistical techniques, I have been able to characterize the functional profile of single chemical exposures based on transcriptomic signatures diagnostic of chemical class. These findings provide support for the use of both zebrafish embryo and gill cell line methods for the identification of potential toxicity biomarkers that can be used to improve ERA.

2.3 Materials and methods

2.3.1 Experimental design

Experimental design along with data acquisition from microarray experiments were carried out by our collaborating labs (Melanie Knobel and Kristin Schirmer from EAWAG, and Ashley Sawle and Andrew Cossins from the University of Liverpool). Briefly, 12 environmentally relevant chemicals[183] were selected to represent 4 mode of action categories (narcotic, neurotoxic, reactives and uncouplers) and to cover a wide range of lipophilicity (expressed as a function of the octanol-water partition coefficient, K_{ow}) (Table 2.1). Both zebrafish embryos and rainbow trout gill cell line (RTgill-w1) were exposed at 2 different concentrations of each toxicant. Concentrations used were defined as EC_{10} and 1/10 of the EC_{10} where EC_{10} is the effective concentration at which 10% of the population exhibit a response. This response was intended as any developmental abnormalities in the zebrafish embryo, and cell membrane or lysosomal integrity in the cell line. Moreover, zebrafish embryos were exposed over a 96h period as indicated by the OECD guidelines, while rainbow trout gill cell line were exposed for 24h.

				Trout		Zebrafish	
Chemical	ID	MoA	KoW	Solvent	EC10 (mM)	Solvent	EC10 (mM)
Diethylphthalate	DEP	Narcotic	2.47	DMSO	0.13	Water	0.041
Dimethylphthalate	DMP	Narcotic	1.6	DMSO	0.76	Water	0.099
Sodium dodecyl sulfate	SDS	Narcotic	1.6	DMSO	0.003	Water	0.01
2,4-dichlorophenol	DCP	Uncoupler	3.06	DMSO	0.006	Water	0.018
Penta-chlorophenol	PCP	Uncoupler	5.12	DMSO	0.00003	DMSO	0.0002
Dinoseb	Dino	Uncoupler	3.56	DMSO	0.00003	DMSO	0.0001
Permethrin	Perm	Neurotoxic	6.5	DMSO	0.001	DMSO	0.001
Esfenvalerate	Esfen	Neurotoxic	6.22	DMSO	0.003	DMSO	0.00005
Lindane	Lind	Neurotoxic	4.14	DMSO	0.005	DMSO	0.0003
Acrolein	Acro	Reactive	-0.01	DMSO	0.003	Water	0.023
Allyl alcohol	AA	Reactive	0.17	DMSO	9.1	Water	0.724
Flucythrinat	Fluc	Reactive	6.2	DMSO	0.003	DMSO	0.00004

Table 2.1: Panel of chemicals used in the study. For each of the compound the ID, the MoA and the relative logP value (K_{ow}) are reported. Solvent and dose used for each compound within each of the systems are showed.

Samples were processed for genome-wide expression profiling using 60K Agilent microarray for both organisms. A two-colour hybridisation strategy was used based on an interwoven ANOVA-based statistical design[184] (Fig. 2.1). This was a relatively popular experimental design in the early days of microarray technology since it allowed a better control of experimental variability. A total of 108 arrays for the zebrafish embryos and 96 for the trout gill cell line were used.

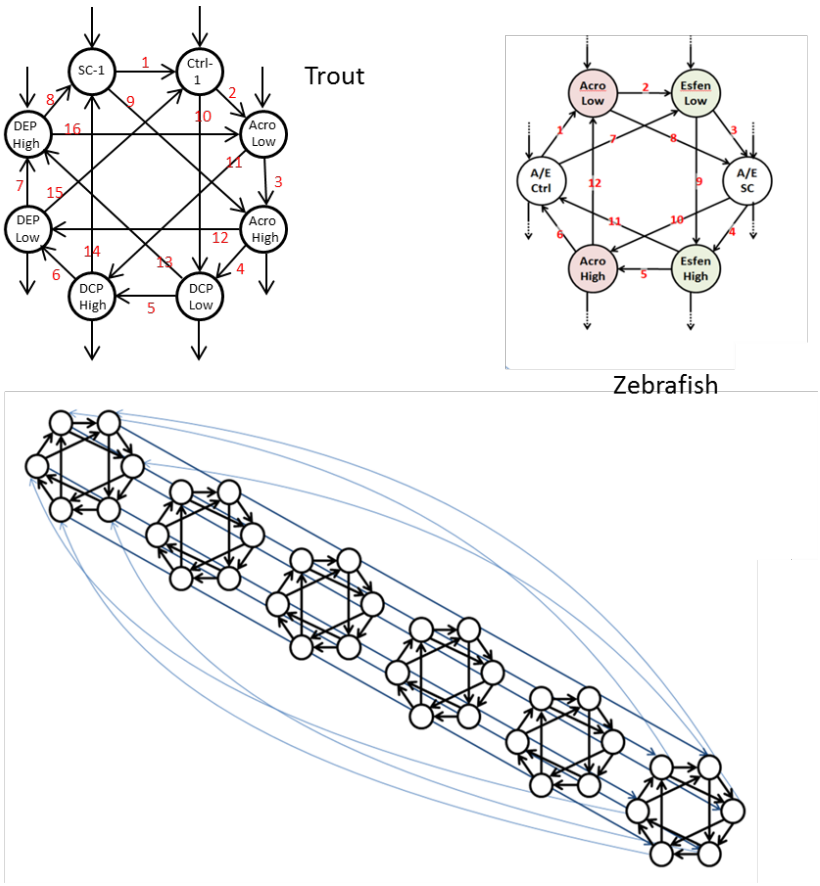


Figure 2.1: Interwoven loop design employed for microarray analysis. Samples are represented by circles and labelled either with Cy3 or Cy5. Hybridisation groups were built differently for trout and zebrafish but all the comparison needed where achieved for both the species.

2.3.2 Annotation of the zebrafish and trout arrays

Zebrafish arrays were purchased from Agilent in the format of 8x44K (in 2007 it was the largest zebrafish array). To reach the 60K probes, the remaining 16K probes were designed by Cossins and collaborators against Ensemble transcripts. To improve our ability to interpret biologically the results of the analysis, we set to re-annotate the array. The array-design sequences were blasted to the Refseq RNA transcriptome (downloaded the 6th February 2015) setting the minimum alignment length to 58bp (all the probes were 60bp) and the identity of match >= 98%.

44,057 probes were successfully assigned to a single gene from the collection of 22,114 non-redundant zebrafish genes.

For the trout an in-house oligoarray designed by Cossins and collaborators was used. It was printed by Agilent in the format of 8x60K. We set to re-annotate the trout array especially because now a full genome is available for trout[185]. Zebrafish orthologs genes were assigned to zebrafish transcripts by BLAST. Using the RBH (reciprocal best hit) technique[186], a successful orthologue assignment was attained. For those that failed, probe mapping to gene models was used where a single gene model was overlapped by a probe, and for that model a successful RBH result was attained. 45,124 probes were assigned to a gene (12,228 non-redundant trout genes). A total of 7698 genes were found to be in common between the zebrafish and the trout dataset. Array annotation was successfully achieved by John Herbert from Francesco Falciani's lab.

2.3.3 Data pre-processing

I adopted the same data pre-processing and downstream analysis workflows for both zebrafish and trout datasets, all within the statistical environment R. The key steps were as following:

Step 1: Positive and negative control spots (1319) were removed from the dataset.

Step 2: Genes whose average expression across all samples was in the lower 20th percentile (17870 for the trout and 18138 for the zebrafish) were removed.

Step 3: Data were normalized using a LOESS regression method followed by a quantile method to correct for dye bias and for between-samples variation, respectively.

Step 4: Duplicated probes were summarised by average intensity.

Step 5: Probes having missing values in at least one sample were removed (5,147 in trout and 2,491 in zebrafish).

Step 6: Probes without biological annotation were removed.

The final dataset for zebrafish comprised 32,625 probes representing 17,387 non-redundant genes, while the trout dataset comprised 27,308 probes representing 9,262 non-redundant genes. Duplicated genes were summarised keeping those with the highest averaged intensity across all the samples. A channel decomposition routine using the "limma" package[187] was applied on the dataset to obtain individual sample measurements. I then assessed whether the decomposition procedure introduced artefacts by first identifying differentially expressed genes from the full loop design using "limma" and on the separated channel dataset using SAM (Significance Analysis of Microarray)[117] and second, I investigated the degree of overlap between the two different methods.

A Principal Component Analysis (PCA) was applied to the individual samples to identify presence of artefacts, such as outliers and batch effects. Using the "prcomp" function within the R statistical environment I successfully identified and removed outliers in both the systems. Moreover, the PCA allowed me to identify the presence of a batch effect in both datasets due to the different cell line passage in trout and to the different hybridisation groups in zebrafish. The batch effect was successfully taken into account by computing ratios of each chemical with its specific solvent control sample.

2.3.4 Identifying potential effects of DMSO

Ability of the DMSO solvent to elicit transcriptional response was investigated first by using a 2-factor ANOVA to assess the relative importance of the treatment (DMSO or water) as well as the potential confounding factors as the different cell line passages and the different hybridisation group as previously defined (10% FDR). Then, I investigated whether this difference between DMSO and water was detectable within each of the hybridisation groups or cell line passages

using a "2-class SAM" method[117] and keeping a high FDR threshold (10% FDR). Finally, I ran a "1-Class SAM" on the DMSO/water ratios and we recorded the d-statistic for each gene. I then ranked the genes and used the ranked list as input into Gene Set Enrichment Analysis (GSEA)[121] to identify potential KEGG pathways affected by DMSO. GSEA allow to identify hidden biological features as doesn't require a threshold to be set beforehand as for the more conventional statistical methods (i.e. SAM).

2.3.5 Single chemical analysis

Differentially expressed genes were identified using the separated channel dataset. Transcriptional response of each of the single chemicals was achieved by running a 2-class SAM analysis between the chemical and its related control, followed by applying a cut-off threshold of 5% FDR.

2.3.6 Statistical power

To address the important question of whether the limited statistical power significantly affected our ability to detect the molecular effects of chemical exposures, the GSEA[121] was employed. I first recorded the d-statistic of the differentially expressed genes associated with each chemical compound and then used the resulting ranked list as input into GSEA[121] to determine enrichment within KEGG pathways. A 1% FDR threshold was applied. Results have been reported as heat-maps.

2.3.7 Functional analysis

Gene lists derived from differential expression analysis of single chemicals were tested for enrichment in genes belonging to Gene Ontology and KEGG pathway functional terms using the DAVID web-service[118]. Biological GO terms with an FDR <= 5% and functional pathways with an FDR<= 1% were considered significant.

2.3.8 Analysis of chemical classes

Chemical classes and logP gene signatures were identified by applying a 2-factor ANOVA coupled with a TUKEY post-hoc test and using a 5% FDR threshold. Functional analysis of gene signatures associated to the lipophilicity was achieved retrieving GO terms using the "Gprofiler" package within the statistical environment R and applying an FDR threshold of 5%. Redundancy of GO terms was removed using REVIGO[122]. Functional analysis of chemical classes was achieved at a pathway-level retrieving KEGG pathways using DAVID web-service and applying a 1% FDR threshold.

Key pathways were further investigated for their ability to be diagnostic of chemical MoA and were selected based on the differential expression between at least two chemical classes. Once identified, all the genes contained within the given pathway were used regardless of the fact that they were differentially expressed. A heat-map was used to visualize the gene expression profile in each of the chemical classes using Spearman correlation as distance method and average as the clustering method for the genes. Bar-charts were then built to show the overall behaviour of that given pathway in each chemical class based on the number of genes showing up- or down-regulation. Heat-maps and bar-charts of key pathways were obtained within the statistical environment R and using the "heatmap.2" and "barplot" functions.

2.4 Results

2.4.1 Channel decomposition

The two-channel microarray dataset was constituted by ratio data. However, since most of the objectives of my analysis strategy required a dataset with measurements on individual samples I applied a channel decomposition routine. Since the channel decomposition procedure may introduce some artefacts I decided to test whether a methodology that analyses the full loop

design may give the same results of a methodology applied on the separated channel data. I therefore identified differentially expressed genes from the full loop design and from the separated channel dataset and we looked at the degree of overlap (Table 2.2).

A

	Trout 5% FDR					
	2-Class SAM		Limma		Overlap	
	up	lo	Up	Lo	Up	Lo
AAHigh	814	1723	815	1477	535	1225
AALow	0	0	0	0	0	0
AcroHigh	472	828	130	137	107	99
Acrolow	0	0	0	0	0	0
DCPHigh	767	387	641	549	450	263
DCPLow	0	0	0	0	0	0
DEPHigh	1152	849	1235	1114	863	649
DEPLow	2	0	0	0	0	0
DinoHigh	1	0	0	0	0	0
DinoLow	0	0	0	0	0	0
DMPHigh	1960	1431	725	533	690	477
DMPLow	0	0	0	0	0	0
EsfenHigh	262	111	112	123	69	53
Esfenlow	15	0	1	0	1	0
FlucHigh	260	346	7	3	6	3
FlucLow	9	0	0	0	0	0
LindHigh	59	104	0	0	0	0
LindLow	0	0	0	0	0	0
PCPHigh	26	3	0	0	0	0
PCPLow	0	0	0	0	0	0
PermHigh	12	0	0	0	0	0
PermLow	0	0	0	0	0	0
SDSHigh	551	180	698	638	425	135
SDSLow	4	0	0	0	0	0

B

	Zebrafish 5% FDR					
	2-Class SAM		Limma		Overlap	
	Up	Lo	Up	Lo	Up	Lo
AAHigh	1861	2111	1067	1012	925	876
AALow	2	1	21	7	1	0
AcroHigh	0	0	18	110	0	0
Acrolow	2	0	0	0	0	0
DCPHigh	1356	1537	226	682	197	671
DCPLow	6	113	8	171	0	61
DEPHigh	114	7	85	22	58	5
DEPLow	25	6	26	3	16	2
DinoHigh	9	1	9	0	5	0
DinoLow	22	13	17	2	10	0
DMPHigh	997	176	585	246	493	66
DMPLow	184	0	793	362	140	0
EsfenHigh	0	0	0	0	0	0
Esfenlow	0	0	0	0	0	0
FlucHigh	0	4	0	0	0	0
FlucLow	0	3	0	0	0	0
LindHigh	36	5	3	1	3	0
LindLow	1	0	0	0	0	0
PCPHigh	632	782	213	540	171	420
PCPLow	718	26	41	7	39	6
PermHigh	2648	3134	1636	2318	1491	2113
PermLow	54	1	64	3	37	0
SDSHigh	78	2	29	4	20	0
SDSLow	206	1	17	2	15	1

Table 2.2: Comparison between full loop design analysis (limma) and independent channel analysis (SAM). Differentially expressed genes identified by the two methods and the relative overlap are shown for the trout (A) and the zebrafish (B).

Results showed that the two analyses were in close agreement for most of the compounds. The High dose of Fluc, Acro and DMP (FlucHigh, AcroHigh and DMPHigh) in the trout system showed differences that may have affected the downstream analysis. However, since only three out of the twelve compounds in the trout system (only considering the High dose, which is the one eliciting a transcriptional response) and none of the compounds in the zebrafish system present some differences between the two analysis, I reasoned that channel decomposition could be achieved without the risk of introducing artefacts.

2.4.2 Identification of outliers and batch effects

As a first step, I set to explore the quality of our dataset looking for potential artefacts that may affect our downstream analysis. A PCA analysis was undertaken on the individual samples to identify the presence of outliers. I successfully identified and removed six and two outliers in the trout and in the zebrafish system, respectively (Fig. 2.2). While in trout the six outliers corresponded to exactly three arrays, in zebrafish the two outliers belonged to two different arrays hence we lost two additional samples when removing them.

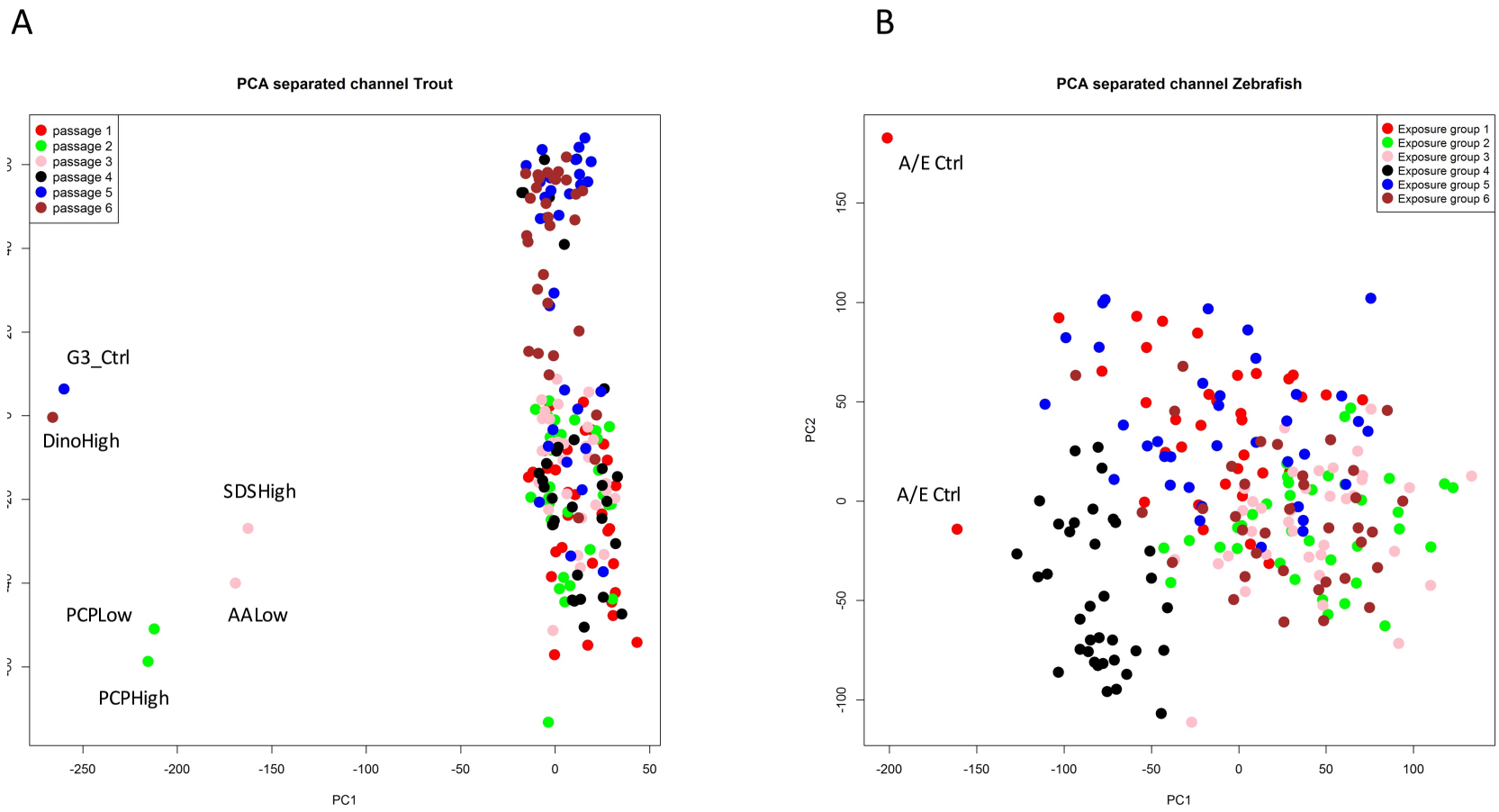


Figure 2.2: PCA plots reveal presence of outliers as well as the presence of a batch effect due to the different cell line passage in trout (A) and the different exposure groups in zebrafish (B).

Moreover, the PCA analysis also allowed me to identify the presence of a batch effect in both the systems due to the different cell line passage in trout and the different hybridisation groups used in the loop design in zebrafish. To account for this batch effect, chemical ratios were reconstituted simply by using the appropriate solvent control sample taking into account the cell line passage and the hybridisation group in trout and zebrafish, respectively.

2.4.3 The effect of DMSO as a solvent was negligible in the trout cell line and in the zebrafish embryos

DMSO is a commonly used solvent for toxicity testing, especially in the FET, due to its low toxicity and its ability to cross biological membranes without affecting their structural integrity. However, little is known about its ability to modulate the uptake of chemicals into the organism. Kais et al., demonstrated that DMSO can be safely used at a maximum concentration of 0.01% (0.1 mL/L)[177]. I tested whether the low final concentration of DMSO used to solubilise the chemical may elicit a transcriptional response. This is because while the DMSO concentration used in zebrafish embryos was 0.01% for all the compounds solubilized, but in the trout gill cell line was 0.05%. I first assessed the relative importance of the different treatments (DMSO/water) as well as the cell line passage or hybridisation groups. The results indicated there was no effect associated with the presence of DMSO but there was a large effect due to the cell line passages and/or hybridisation group as previously identified by the PCA analysis (Table 2.3).

	2-way ANOVA 10% FDR	
	Trout	Zebrafish
Treatment (DMSO/water)	60	19
Hybridisation group/Cell line passage	3861	14451
Interaction	0	2

Table 2.3: Results of the 2-way ANOVA used to evaluate the ability of DMSO to elicit transcriptional responses in both the systems. Results reveal that most of the genes were associated with the confounding factors.

Considering the strong influence on the transcriptional response of these confounding factors, I set to investigate whether this difference between DMSO and water was detectable within each of the hybridisation groups or cell line passages. The comparison between DMSO and water only gave a significant amount of differentially expressed genes in 1 out of 6 hybridisation group (277 genes in the group 5) (Table 2.4).

	2-Class SAM (10% FDR)			
	Trout		Zebrafish	
	Up-regulated	Down-regulated	Up-regulated	Down-regulated
Hybridisation group/Cell line passage 1	0	0	2	0
Hybridisation group/Cell line passage 2	4	0	0	0
Hybridisation group/Cell line passage 3	11	3	2	2
Hybridisation group/Cell line passage 4	0	1	46	0
Hybridisation group/Cell line passage 5	0	0	277	0
Hybridisation group/Cell line passage 6	13	0	1	0

Table 2.4: Results of the comparative analysis between DMSO and water ran within each hybridisation group or among each cell line passage. Only within the zebrafish system and for a hybridisation group alone we identified a significant number of DEGs.

As a final analysis, I have employed the GSEA to identify potential pathways affected by the DMSO. It is important to note that GSEA identified a number of significant pathways suggesting that although undetected by conventional hypothesis testing statistical techniques, DMSO may affect the transcriptional state of embryos and cells (Fig. 2.3). Also, important to note is the fact that although present, these effects may be very small.

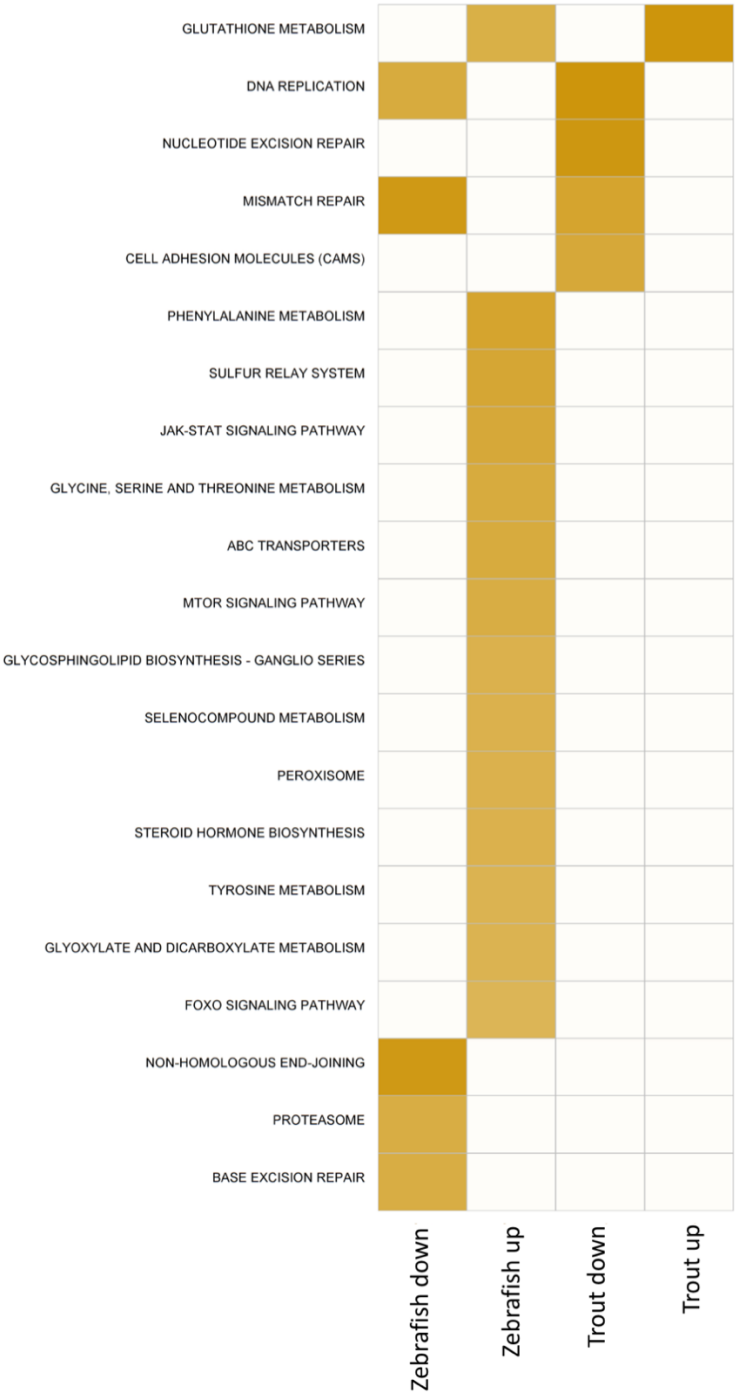


Figure 2.3: GSEA analysis reveals biological pathways potentially affected by DMSO. The darker the orange, the higher is the normalised enrichment score (NES).

2.4.4 Single chemical response is heterogeneous

I first asked whether chemical exposure induces a detectable transcriptional response in either the *in vitro* and *in vivo* systems. I discovered that this is indeed the case. However, the extent of

response (defined by the number of genes differentially regulated) was dose-dependent and it varied considerably across the chemicals (Fig 2.4). Low dose chemicals were generally unable to elicit a measurable transcriptional response. Moreover, not all the compounds in the study elicited a significant transcriptional response, even at high dose exposure. For example, PCP, dinoseb and permethrin elicited a poor transcriptional response in the cell line experiment (29, 9 and 12 genes respectively) while acrolein, flucythrinat, dinoseb, esfenvalerate and lindane were unable to induce a significant transcriptional response in the zebrafish embryos (0, 4, 10, 0 and 41 genes respectively). This observation was consistent across a spectrum of FDR thresholds (Table 2.5).

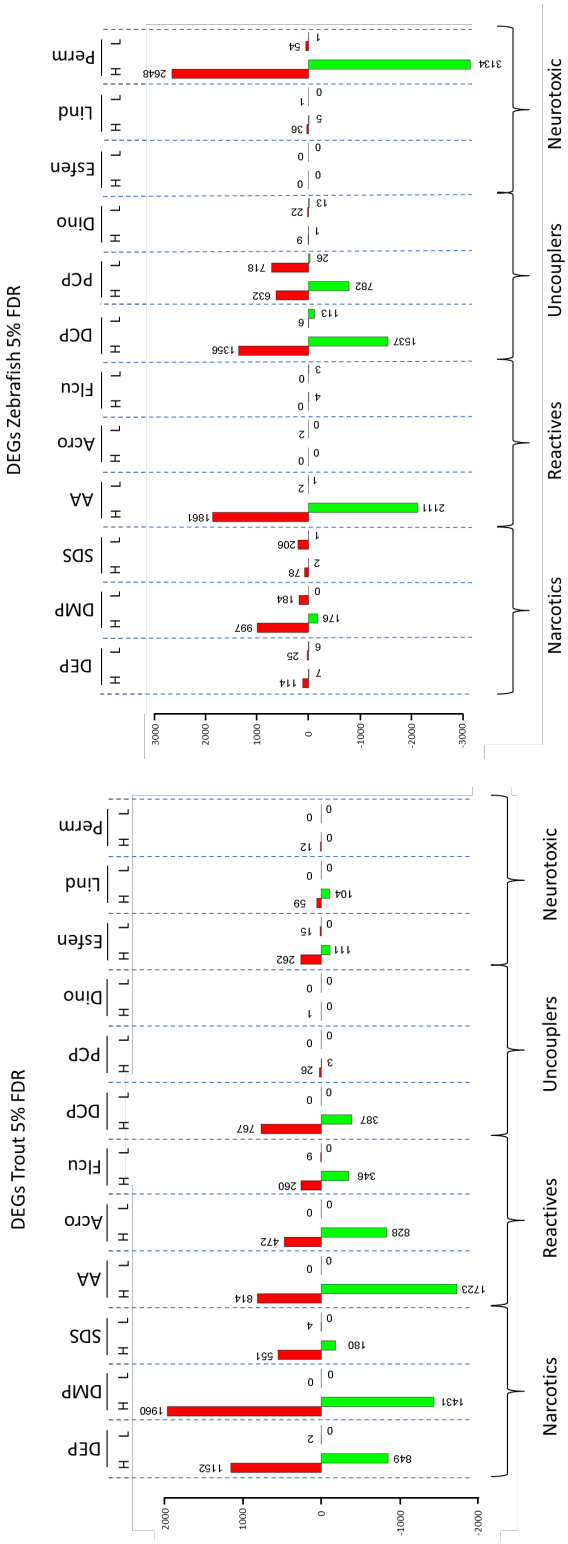


Figure 2.4: Number of differentially expressed genes identified for each of the compounds in the study. Red and green bars refer to up and down regulated genes respectively. Chemicals ID and mode of action are reported.

	Trout							
	1% FDR		5% FDR		10% FDR		20% FDR	
	up	lo	up	lo	up	lo	up	lo
AAHigh	335	878	814	1723	1327	2283	2012	2898
AALow	0	0	0	0	0	0	0	0
AcroHigh	148	304	472	828	922	1525	1640	2368
Acrolow	0	0	0	0	0	0	0	0
DCPHigh	350	178	767	387	1318	700	2003	1067
DCPLow	0	0	0	0	0	0	5	0
DEPHigh	420	274	1152	849	1734	1330	2569	1891
DEPLow	2	0	2	0	2	0	8	7
DinoHigh	1	0	1	0	14	0	57	0
DinoLow	0	0	0	0	0	0	0	0
DMPHigh	981	555	1960	1431	2654	2122	3405	2986
DMPLow	0	0	0	0	14	0	77	0
EsfenHigh	77	43	262	111	537	204	1274	417
Esfenlow	15	0	15	0	19	0	19	0
FlucHigh	92	142	260	346	502	587	966	1011
FlucLow	9	0	9	0	11	0	15	0
LindHigh	27	54	59	104	102	163	289	348
LindLow	0	0	0	0	0	0	0	0
PCPHigh	7	3	26	3	115	6	352	66
PCPLow	0	0	0	0	0	0	0	0
PermHigh	12	0	12	0	51	0	257	0
PermLow	0	0	0	0	0	0	0	0
SDSHigh	182	28	551	180	878	350	1350	622
SDSLow	4	0	4	0	4	0	8	0

	Zebrafish							
	1% FDR		5% FDR		10% FDR		20% FDR	
	up	lo	up	lo	up	lo	up	lo
AAHigh	976	922	1861	2111	2091	2624	2856	4012
AALow	2	1	2	1	11	2	22	2
AcroHigh	0	0	0	0	0	0	16	1
Acrolow	2	0	2	0	2	0	21	0
DCPHigh	357	1081	1356	1537	3438	1976	4928	2249
DCPLow	1	17	6	113	11	388	20	673
DEPHigh	37	4	114	7	178	12	305	36
DEPLow	15	5	25	6	34	7	90	25
DinoHigh	9	1	9	1	11	5	24	6
DinoLow	20	10	22	13	44	63	65	133
DMPHigh	339	24	997	176	1376	856	1926	2971
DMPLow	34	0	184	0	375	0	791	12
EsfenHigh	0	0	0	0	0	0	0	0
Esfenlow	0	0	0	0	0	0	0	0
FlucHigh	0	4	0	4	0	4	0	35
FlucLow	0	3	0	3	0	3	0	6
LindHigh	19	1	36	5	60	5	89	9
LindLow	1	0	1	0	1	0	1	0
PCPHigh	118	167	632	782	885	1025	1875	1691
PCPLow	353	6	718	26	920	42	1506	3016
PermHigh	1509	1654	2648	3134	3209	3914	4575	5551
PermLow	29	1	54	1	75	1	118	1
SDSHigh	29	2	78	2	186	2	370	17
SDSLow	50	1	206	1	366	4	506	7

Table 2.5: Number of differentially expressed genes as a result of chemical exposure across different FDR thresholds. Some of the compounds are not able to induce a significant transcriptional response event when considering high FDR cut-off.

I discovered that the transcriptional response elicited by neurotoxic chemicals was 10-fold higher in zebrafish than in the cell line (5823 genes in zebrafish while just 548 in the cell line). Conversely, narcotics chemicals are more effective in eliciting a transcriptional response in the cell line than the embryo. Interestingly, all the reactive compounds in the panel were able to induce a transcriptional response in the cell line whilst just one out of the three compounds, allyl alcohol (AA), was effective in the embryo system. Given the poor transcriptional response elicited by low dose chemicals, I focused only on the high dose and all the subsequent downstream analyses were carried out using just the high dose.

2.4.5 Alternative methods to identify hidden transcriptional signatures

Here I addressed the important question whether the limited statistical power significantly affected our ability to detect the transcriptomic effects of chemical exposures. The use of GSEA allows the identification of hidden biological effects even in the low dose chemicals for which SAM was not successful (Fig. 2.5-2.6).

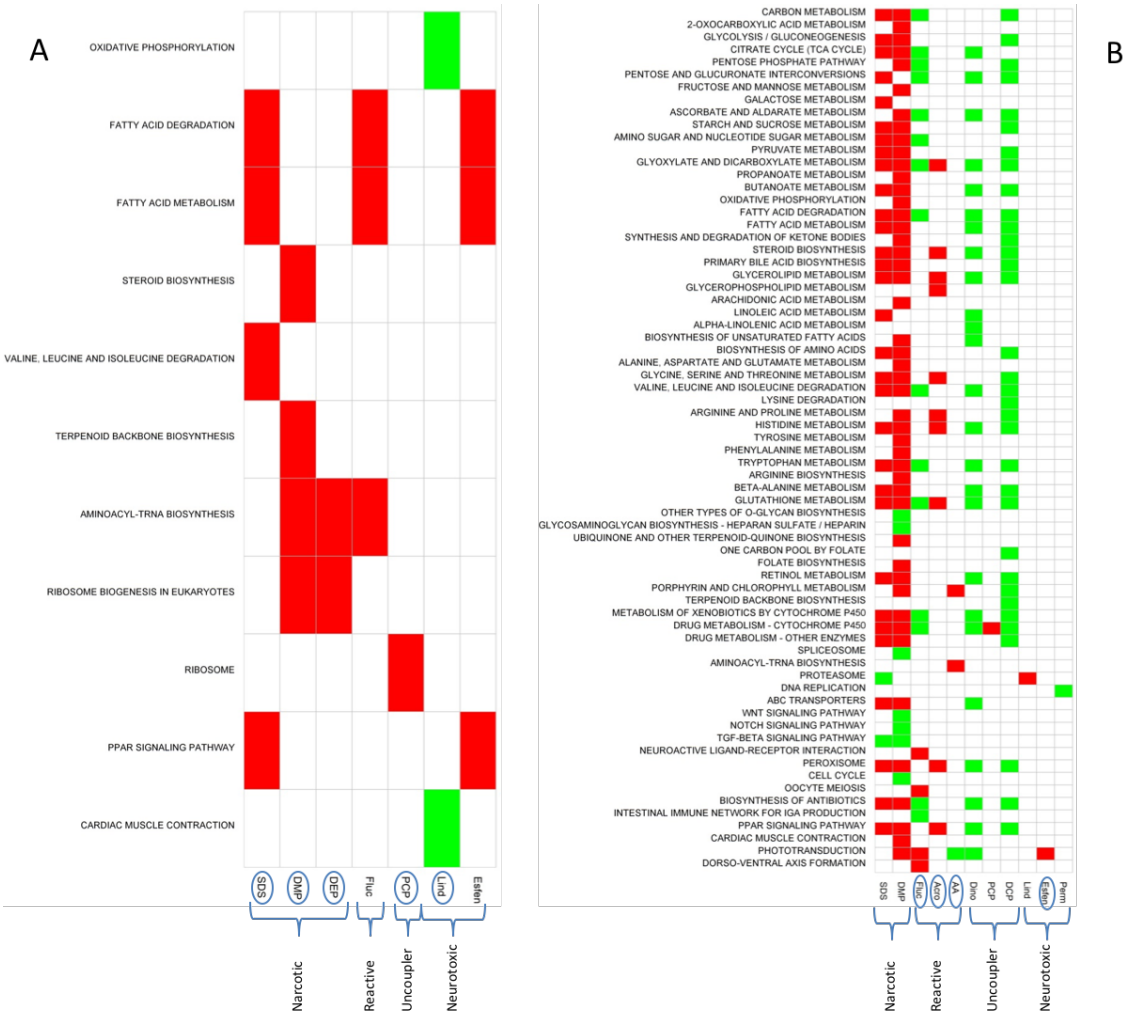


Figure 2.5: GSEA results (1% FDR) for the low dose chemicals (A is trout and B is zebrafish). Chemicals circled are those for which SAM spotted less than 5 genes to be differentially expressed. Red and green are up-regulated and down-regulated genes, respectively.

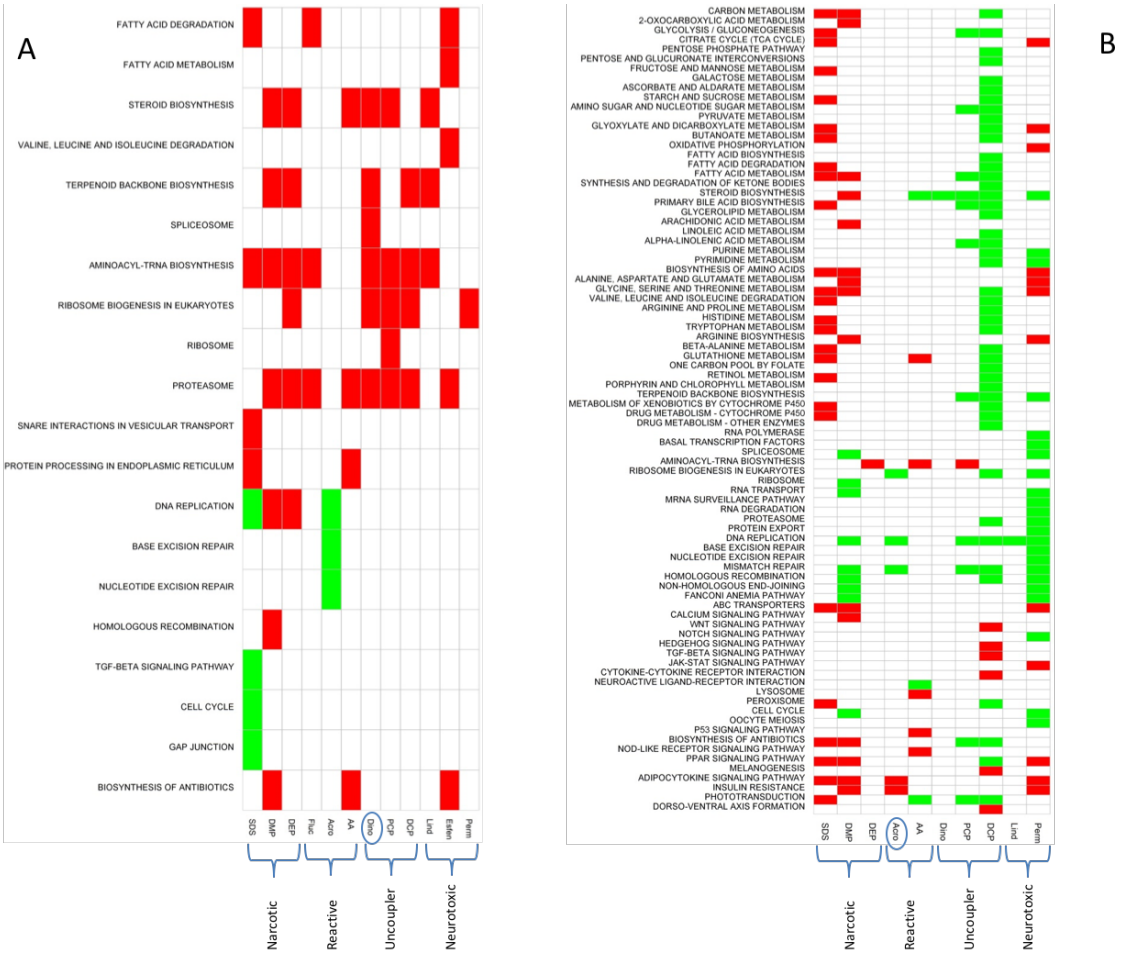


Figure 2.6: GSEA results (1% FDR) for the high dose chemicals (A is trout and B is zebrafish). Chemicals circled are those for which SAM spotted less than 5 genes to be differentially expressed. Red and green are up-regulated and down-regulated genes, respectively.

2.4.6 Functional analysis of single chemical signatures

In order to biologically interpret these results, I performed a functional analysis to identify biological functions (GO terms) enriched in the list of differentially regulated genes (5% FDR) (Fig. 2.7).

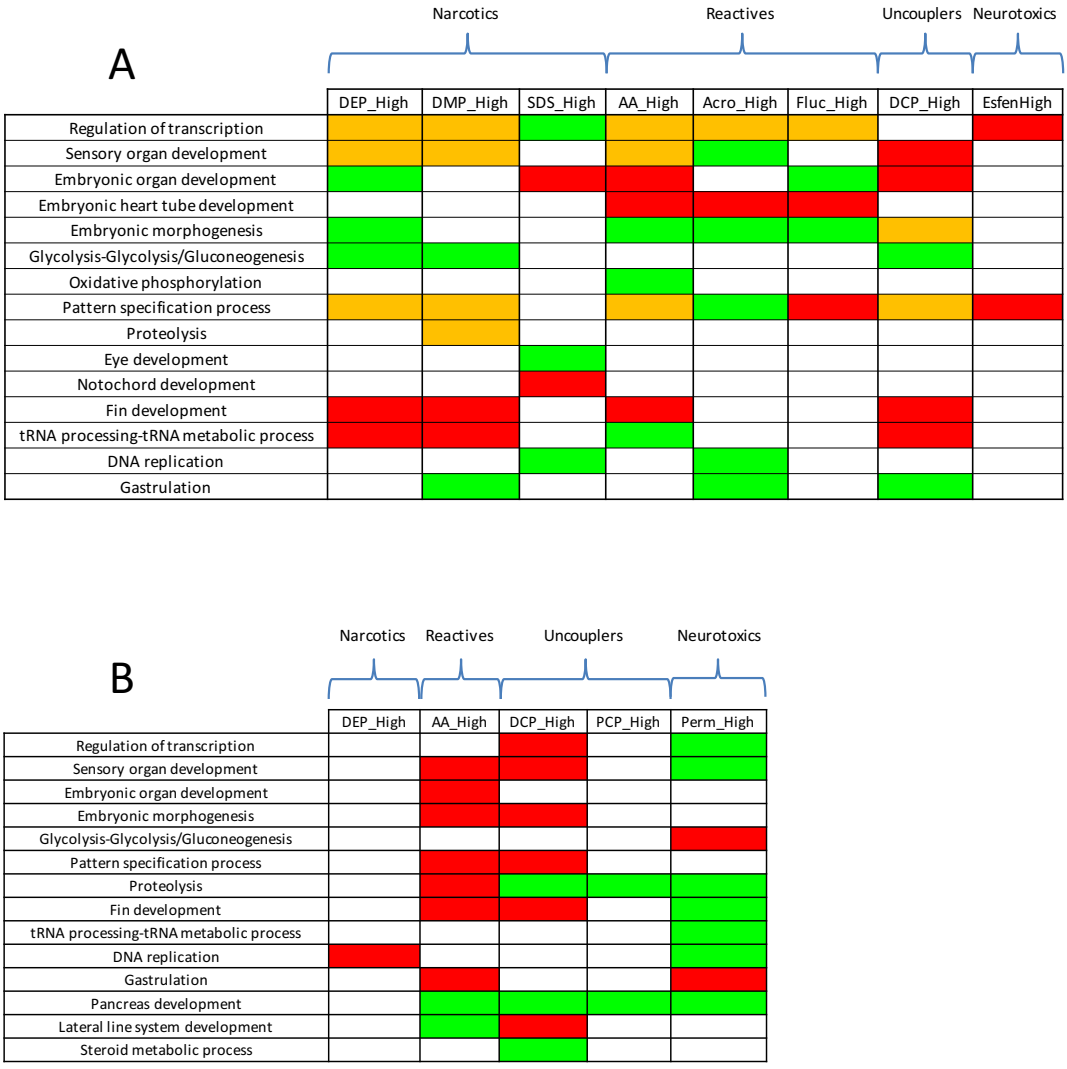


Figure 2.7: Biological GO terms significantly enriched in each of the chemical compounds (5% FDR). Gene level analysis in the trout cell line (A) and in the zebrafish embryos (B) are displayed. Red and green show up- and down-regulated functions respectively while orange shows functions which includes genes either up- or down-regulated. FDR values and genes included in each of the GO terms are provided in the supplementary material.

This functional analysis interestingly reveals that most of the chemicals affect biological functions associated with embryo developmental processes in both the cell line and the zebrafish embryo. In the cell line, DEP, DMP, SDS, AA, Acro, Fluc and DCP hit functions such as sensory organ development, embryonic organ development, embryonic heart tube development, embryonic morphogenesis, pattern specification process, eye development, fin development and

notochord development. Some of these functions, such as sensory organ development, embryonic organ development, embryonic morphogenesis, pattern specification process and fin development, were also enriched in the zebrafish embryo by genes associated with AA, DCP and Perm. Interestingly, AA and DCP are shown to affect biological functions linked to embryonic development regardless the species.

To have a better overview of the biological pathways affected by each of the single chemical in the two biological systems, I tested whether KEGG pathways were enriched of differentially expressed genes associated with each chemical compound (1% FDR) (Fig. 2.8).

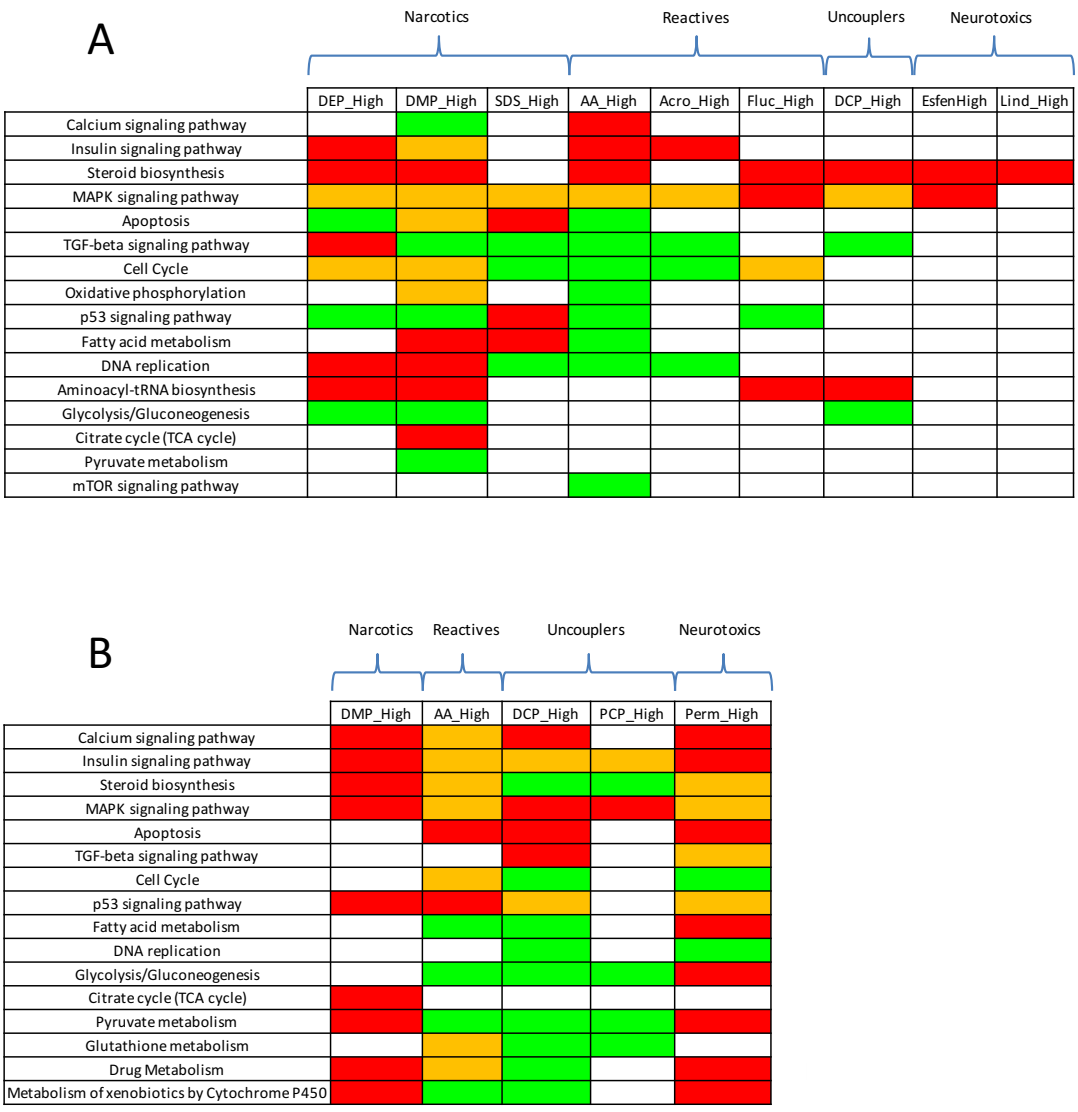


Figure 2.8: KEGG pathways significantly enriched in each of the chemical compounds (1% FDR). Pathway level analysis in the trout cell line (A) and in the zebrafish (B) embryos are displayed. Red and green show up- and down-regulated functions, respectively, while orange shows functions including genes both up- or down-regulated. FDR values and genes included in each of the KEGG pathways are provided in the supplementary material.

Pathway level analysis reveals a degree of functional similarity between chemicals belonging to the same chemical class suggesting they may share, as expected, mechanisms of action. Moreover, a partial overlap was observed also among chemicals belonging to different chemical classes. In trout, I observed a poor response by uncouplers and by neurotoxic compounds, as

previously stated. On the other hand, narcotics and reactives displayed massive functional response. Pathways enriched were mainly linked to signal transduction (MAPK and TGF-beta signalling pathways), cell growth and death (apoptosis, cell cycle and p53 signaling pathway), replication and transduction (DNA replication and aminoacyl-tRNA biosynthesis) and energy metabolism (calcium and insulin signalling pathways, oxidative phosphorylation and glycolysis/gluconeogenesis).

In zebrafish, less compounds per class were functionally enriched, but the functional overlap was more conserved across the 4 chemical classes. Functions represented were mainly linked to signal transduction and cell growth and death as in trout. Energy metabolism was again represented with the same functions plus the pyruvate metabolism. Replication and transduction pathways were poorly enriched compared to the trout. Interestingly, in zebrafish I observed a significant enrichment of functions linked to xenobiotics metabolism (e.g., metabolism of xenobiotics by cytochrome P450, drug metabolism and glutathione metabolism).

2.4.7 Identification of signatures linked to specific Mode of Action (MoA)

The functional analysis of the response of both cell line and embryos to chemical exposure suggests that similar pathways were affected by chemicals with similar MOA. I therefore decided to test whether an organism's transcriptional responses can separate chemicals on the basis of their MoA (Fig. 2.9).

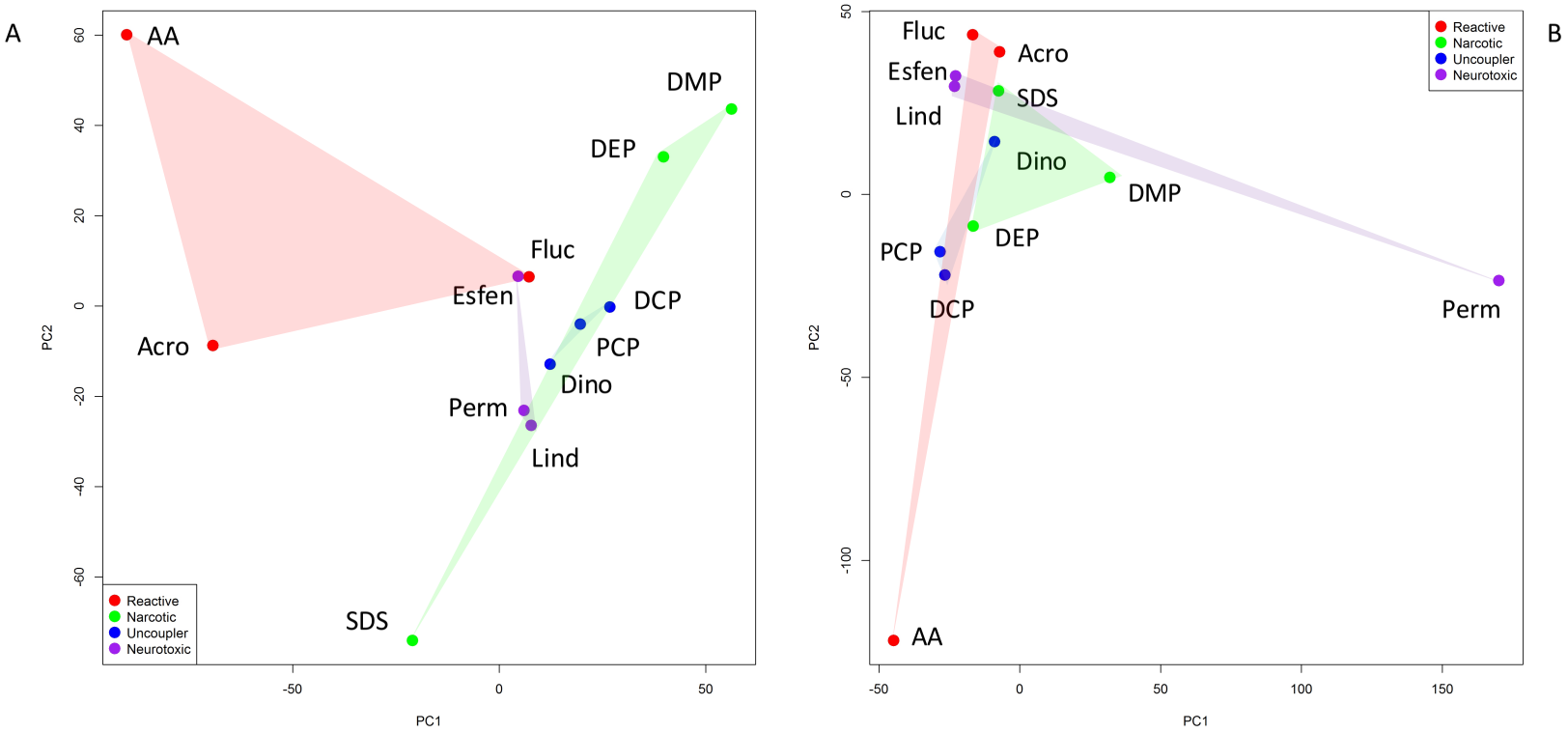


Figure 2.9: PCA plots of the trout gill cell line (A) and the zebrafish embryos (B) transcriptional response based on the differentially expressed genes between each of the single chemicals and its relative control. The plots suggest we were able to discriminate between the different MoAs except for the uncouplers. Moreover, few compounds such as AA, Acro and SDS in trout, and AA and Perm in the zebrafish, elicit specific responses. Chemicals are: Allyl Alcohol (AA), Acrolein (Acro), Flucytrhinat (Fluc), Esfenvalerate (Esfen), Permethrin (Perm), Lindane (Lind), Dinoseb (Dino),penta-chlorophenol (PCP), 2,4-dichlorophenol (DCP), Dimethylphthalate (DMP), Diethylphthalate (DEP), Sodium dodecyl sulfate (SDS).

PCA plots reveal that this was indeed the case. Moreover, both the zebrafish and trout PCA plots showed few chemicals were able to trigger a very specific response (AA and Perm in zebrafish and AA, Acro and SDS in trout).

Since the exploratory analysis suggested that transcriptional response was indicative of a chemical MoA, I set out to identify specific gene signatures that were diagnostic for each of the four chemical classes. I identified genes associated with each of the 4 chemical classes (Table 2.6).

	Trout		Zebrafish	
	Up	Down	Up	Down
Neurotoxic-Narcotic	547	684	919	778
Reactive-Narcotic	726	1642	1005	858
Uncoupler-Narcotic	206	282	996	1328
Reactive-Neurotoxic	451	1234	1145	1179
Uncoupler-Neurotoxic	175	210	1345	1765
Uncoupler-Reactive	1423	666	750	1164
LogP	1068		133	

Table 2.6: Number of differentially expressed genes for either each chemical class comparison and lipophilicity identified applying a 2-way ANOVA coupled with a Tukey post-hoc test and using an FDR threshold of 5%.

Lipophilicity highly affect chemical uptake by cells/tissue and always needs to be considered. Interestingly, I identified a higher amount of genes associated with lipophilicity in the cell line than the embryo. To characterise the biological response linked to lipophilicity a functional analysis on the genes identified as being associated with LogP was performed. This lipophilicity functional profile revealed a massive regulation of transcriptional and translational processes in the Trout gill cell line (Table 2.7A). Functions affected in zebrafish were mainly associated with immune system processes and development, and with homeostasis (Table 2.7B) (5% FDR).

A

term_ID	description
GO:0034660	ncRNA metabolic process
GO:0042254	ribosome biogenesis
GO:0071697	ectodermal placode morphogenesis
GO:0006807	nitrogen compound metabolic process
GO:0071363	cellular response to growth factor stimulus
GO:1901360	organic cyclic compound metabolic process
GO:0006725	cellular aromatic compound metabolic process
GO:0046483	heterocycle metabolic process
GO:0034641	cellular nitrogen compound metabolic process
GO:0010467	gene expression
GO:0071696	ectodermal placode development
GO:0009451	RNA modification
GO:0016070	RNA metabolic process
GO:0090304	nucleic acid metabolic process
GO:0034470	ncRNA processing
GO:0006396	RNA processing
GO:0006139	nucleobase-containing compound metabolic process
GO:0051252	regulation of RNA metabolic process
GO:0001510	RNA methylation

B

term_ID	description
GO:0002376	immune system process
GO:0002520	immune system development
GO:0042752	regulation of circadian rhythm
GO:0048872	homeostasis of number of cells
GO:0002262	myeloid cell homeostasis

Table 2.7: Biological functions (GO terms) found to be affected by lipophilicity in the Trout gill cell line (A) and in the zebrafish embryo (B) (5% FDR). Most of the functions are associated with immune system processes in zebrafish and with the regulation of transcription and translation in trout. FDR values and genes included in each of the GO terms are provided in the supplementary material.

Having defined the influence of lipophilicity on the transcriptional response of the two biological systems, I sought to identify biological functions associated with each of the 4 different MoA's. Differentially expressed genes identified as a comparison between pairs of chemical classes, were assessed for enrichment of KEGG biological pathways (1% FDR). The results implicated a wide range of biological functions. To facilitate representation, results have been arranged into tables, each reporting pathways belonging to a specific KEGG level of organisation (table 2.8).

Energy metabolism	Neuro-Narc		Rea-Narc		Rea-Neuro		Unc-Narc		Unc-Neuro		Unc-Rea	
	T	Z	T	Z	T	Z	T	Z	T	Z	T	Z
Glycolysis/Gluconeogenesis		↑				↓		↓		↓		↓
Citrate cycle (TCA cycle)						↓				↓		
Penthose phosphate						↓				↓		
Fructose and mannose						↓				↓		
Ascorbate and aldarate								↓		↓		
Starch and sucrose						↓		↓		↓		
Galactose								↓		↓		
Amino sugar and nucleotide sugar								↓		↓		↓
Pyruvate						↓				↓		
Glyoxylate and dicarboxylate								↓		↓		↓
Propanoate						↓		↓		↓		
Butanoate						↓		↓		↓		
Oxidative phosphorylation				↓		↓				↓		
Nitrogen										↓		
Pentose and glucuronate interconversion								↓				

Lipid metabolism	Neuro-Narc		Rea-Narc		Rea-Neuro		Unc-Narc		Unc-Neuro		Unc-Rea	
	T	Z	T	Z	T	Z	T	Z	T	Z	T	Z
Fatty acid metabolism				↓		↓		↓		↓		↓
Fatty acid degradation				↓		↓		↓		↓		
Fatty acid elongation								↓				
Synthesis and degradation of ketone bodies								↓		↓		
Steroid biosynthesis								↓		↓		↓
Steroid hormone biosynthesis								↓				
Primary bile acid biosynthesis								↓		↓		↓
Glycerolipid							↑	↓		↓		
Glycerophospholipid												
Arachidonic acid								↓		↓		
Sphingolipid												
Biosynthesis of unsaturated fatty acids								↓		↓		
Alpha-linoleic acid								↓				
Linoleic acid								↓		↓		
Purine		↓										
Pyrimidine		↓										

(Continue next page)

(Continued)

Amino acid metabolism	Neuro-Narc		Rea-Narc		Rea-Neuro		Unc-Narc		Unc-Neuro		Unc-Rea	
	T	Z	T	Z	T	Z	T	Z	T	Z	T	Z
Alanine, aspartate and glutamate					↓		↑			↓		
Glycine, serine and threonine								↓		↓		
Valine, leucine and isoleucine degradation				↓		↓		↓		↓		
Arginine and proline				↓				↓		↓		
Histidine								↓		↓		
Tryptophan				↓		↓		↓		↓		↓
Beta-alanine						↓		↓		↓		
Glutathione						↑		↓				↓
Biosynthesis of amino acids				↓		↓				↓		

Other metabolism	Neuro-Narc		Rea-Narc		Rea-Neuro		Unc-Narc		Unc-Neuro		Unc-Rea	
	T	Z	T	Z	T	Z	T	Z	T	Z	T	Z
Riboflavin metabolism							↑					
Nicotinate and nicotinamide												
Retinol				↓			↑	↓		↓		
Porphyrin and chlorophyll												
Caffeine								↓				
Glycosphingolipid biosynthesis – lacto and neolacto series							↑					
Carbon				↓		↓		↓		↓		↓
2-oxocarboxylic acid						↓				↓		

Xenobiotic metabolism	Neuro-Narc		Rea-Narc		Rea-Neuro		Unc-Narc		Unc-Neuro		Unc-Rea	
	T	Z	T	Z	T	Z	T	Z	T	Z	T	Z
Metabolism of xenobiotics by cytochrome P450								↓		↓		↓
Drug metabolism – cytochrome P450								↓		↓		
Drug metabolism – other enzymes								↓		↓		

(Continue next page)

(Continued)

Replication and repair	Neuro-Narc		Rea-Narc		Rea-Neuro		Unc-Narc		Unc-Neuro		Unc-Rea	
	T	Z	T	Z	T	Z	T	Z	T	Z	T	Z
DNA replication		↓				↑						
Base excision repair				↑								
Nucleotide excision repair		↓										
Mismatch repair		↓										

Transcription	Neuro-Narc		Rea-Narc		Rea-Neuro		Unc-Narc		Unc-Neuro		Unc-Rea	
	T	Z	T	Z	T	Z	T	Z	T	Z	T	Z
RNA polymerase		↓	↓									
Spliceosome						↑						

Translation	Neuro-Narc		Rea-Narc		Rea-Neuro		Unc-Narc		Unc-Neuro		Unc-Rea	
	T	Z	T	Z	T	Z	T	Z	T	Z	T	Z
Aminoacyl-tRNA biosynthesis	↓	↓	↓			↑			↑	↑	↑	
Ribosome biogenesis in eukaryotes		↓	↓			↑				↑	↑	
RNA transport		↓	↓			↑				↑		

Folding, sorting and degradation	Neuro-Narc		Rea-Narc		Rea-Neuro		Unc-Narc		Unc-Neuro		Unc-Rea	
	T	Z	T	Z	T	Z	T	Z	T	Z	T	Z
Ubiquitin mediated proteolysis				↑								

Signaling	Neuro-Narc		Rea-Narc		Rea-Neuro		Unc-Narc		Unc-Neuro		Unc-Rea	
	T	Z	T	Z	T	Z	T	Z	T	Z	T	Z
ABC transporters								↓				
MAPK signaling							↑					
Calcium signaling						↓						
Wnt signaling							↑	↑				
Notch signaling								↑				
Hedgehog signaling								↑				
TGF-beta signaling				↑				↑				
Jak-STAT signaling							↑					
Cytokine-Cytokine receptor interaction							↑					

(Continue next page)

(Continued)

Transport and catabolism	Neuro-Narc		Rea-Narc		Rea-Neuro		Unc-Narc		Unc-Neuro		Unc-Rea	
	T	Z	T	Z	T	Z	T	Z	T	Z	T	Z
Lysosome												↓
Peroxisome								↓		↓		↓
Cell community	**Neuro-Narc**		**Rea-Narc**		**Rea-Neuro**		**Unc-Narc**		**Unc-Neuro**		**Unc-Rea**	
	T	Z	T	Z	T	Z	T	Z	T	Z	T	Z
Cell adhesion molecules (CAMs)							↑					
Tight junction							↑					
GAP junction							↑					
Cell growth and death	**Neuro-Narc**		**Rea-Narc**		**Rea-Neuro**		**Unc-Narc**		**Unc-Neuro**		**Unc-Rea**	
	T	Z	T	Z	T	Z	T	Z	T	Z	T	Z
Cell cycle		↓										
Oocyte meiosis							↑					
Immune system	**Neuro-Narc**		**Rea-Narc**		**Rea-Neuro**		**Unc-Narc**		**Unc-Neuro**		**Unc-Rea**	
	T	Z	T	Z	T	Z	T	Z	T	Z	T	Z
RIG-I-like receptor signaling				↑								
Endocrine system	**Neuro-Narc**		**Rea-Narc**		**Rea-Neuro**		**Unc-Narc**		**Unc-Neuro**		**Unc-Rea**	
	T	Z	T	Z	T	Z	T	Z	T	Z	T	Z
PPAR signaling								↓		↓		↓
Melanogenesis								↑				
Circulatory and sensory system	**Neuro-Narc**		**Rea-Narc**		**Rea-Neuro**		**Unc-Narc**		**Unc-Neuro**		**Unc-Rea**	
	T	Z	T	Z	T	Z	T	Z	T	Z	T	Z
Cardiac muscle contraction				↓		↓						
Phototransduction				↓				↓				

Table 2.8: Pathways differentially regulated between pairs of chemical classes in both specie. Columns report pairs of chemical classes compared in trout (T) and zebrafish (Z) while red and green arrows represent up-regulation and down-regulation respectively and always refer to the first of the chemical class stated. FDR values and genes included in each of the Kegg pathways are provided in the supplementary material.

Interestingly, most of the functional enrichment has been achieved for the zebrafish embryo while only a few pathways in the trout gill cell line were found to be differentially expressed between at least a pair of chemicals. In zebrafish, the greatest functional difference was observed between uncouplers and narcotics/neurotoxics while the difference with reactives chemicals was just limited to a few functional pathways. Energy metabolism (Glycolysis/gluconeogenesis, amino sugar and nucleotide sugar and glyoxylate and dicarboxilate pathways), lipid metabolism (steroid biosynthesis and fatty acid metabolism), transport and catabolism (peroxisome), endocrine system (PPAR signaling) xenobiotics metabolism (metabolism of xenobiotics by cytochrome P450) displayed a significant down-regulation compared to the other chemical classes following exposure to uncoupling chemicals. Differences between reactives and neurotoxics/narcotics were mainly associated with energy and amino acid metabolism. Pathways involved with translation (Aminoacyl-tRNA biosynthesis, ribosome biogenesis in eukaryotes and RNA transport) were down-regulated following exposure to neurotoxics compounds when compared with the other chemical classes. In trout, as previously mentioned, functional enrichment was extremely poor. The biological pathways showing the greatest differences between pairs of chemical classes were translation and signaling pathways.

To further investigate the differences between the molecular mechanisms underlying the MoAs for the different chemical classes I focused on some key pathways (defined as those whose regulation was statistically different in more than 2 chemical classes) and the degree of differential expression in each of the different chemical classes was assessed.

Oxidative phosphorylation, within the dataset of the zebrafish embryo system, was down-regulated in uncouplers compared to neurotoxics, and in the reactives compared to both narcotics and neurotoxics. The heat-map confirms these findings (Fig. 2.10). Moreover, I found

that 65% and 82% of the genes involved in this pathway to be down-regulated in the uncoupler and reactives, respectively, while narcotics and neurotoxics had 82% and 96% of up-regulated genes, respectively (Fig. 2.11).

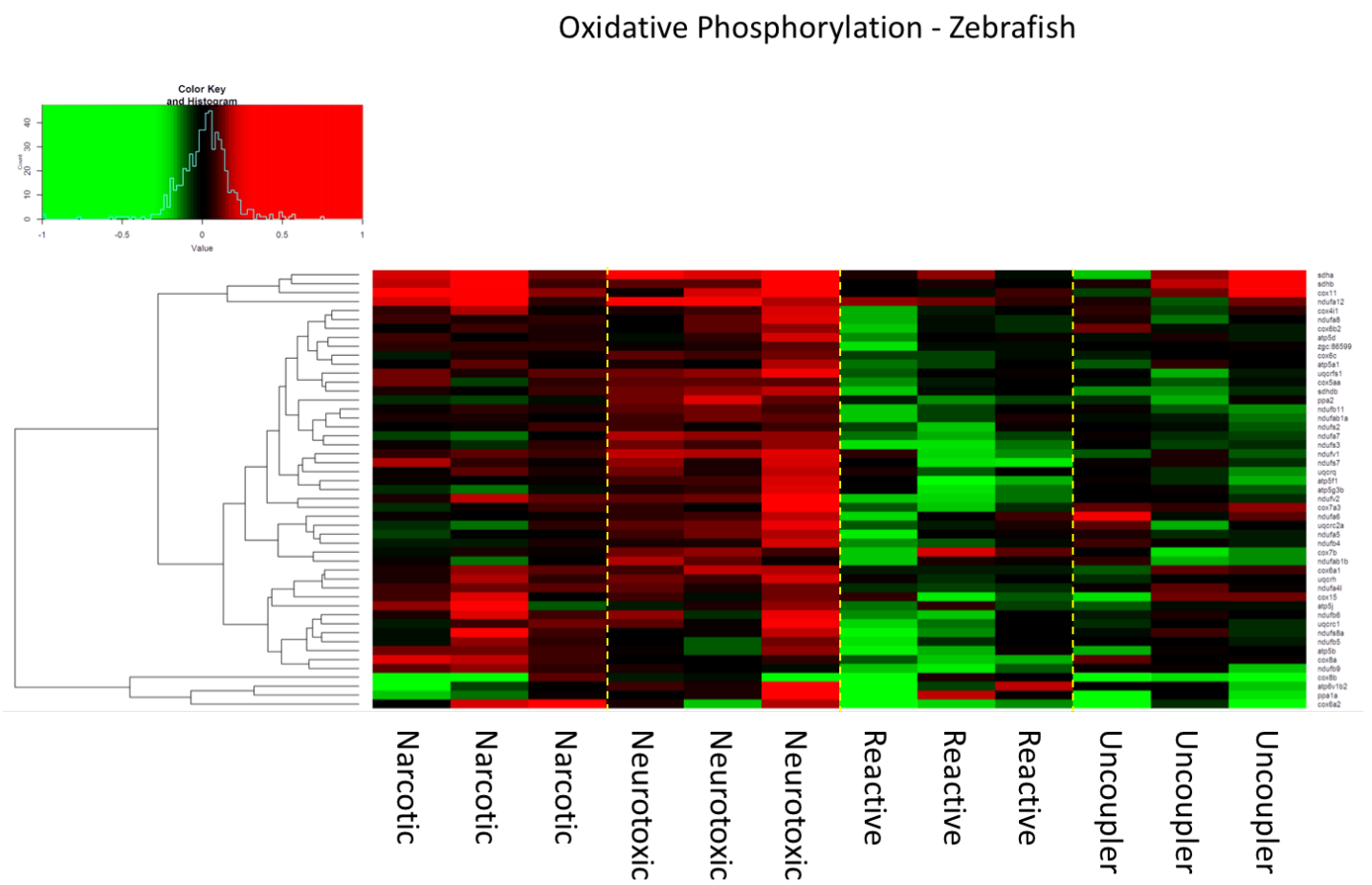


Figure 2.10: Gene expression levels of the genes involved in oxidative phosphorylation in response to each of the listed chemical classes. Red and green refers to up- and down-regulation, respectively. The heatmap reflects the results obtained by our gene expression and functional analysis with uncouplers and reactives showing a down-regulation compared to narcotics and neurotoxics. Heat-map has been performed using Spearman correlation as distance method and average as the clustering method for the genes.

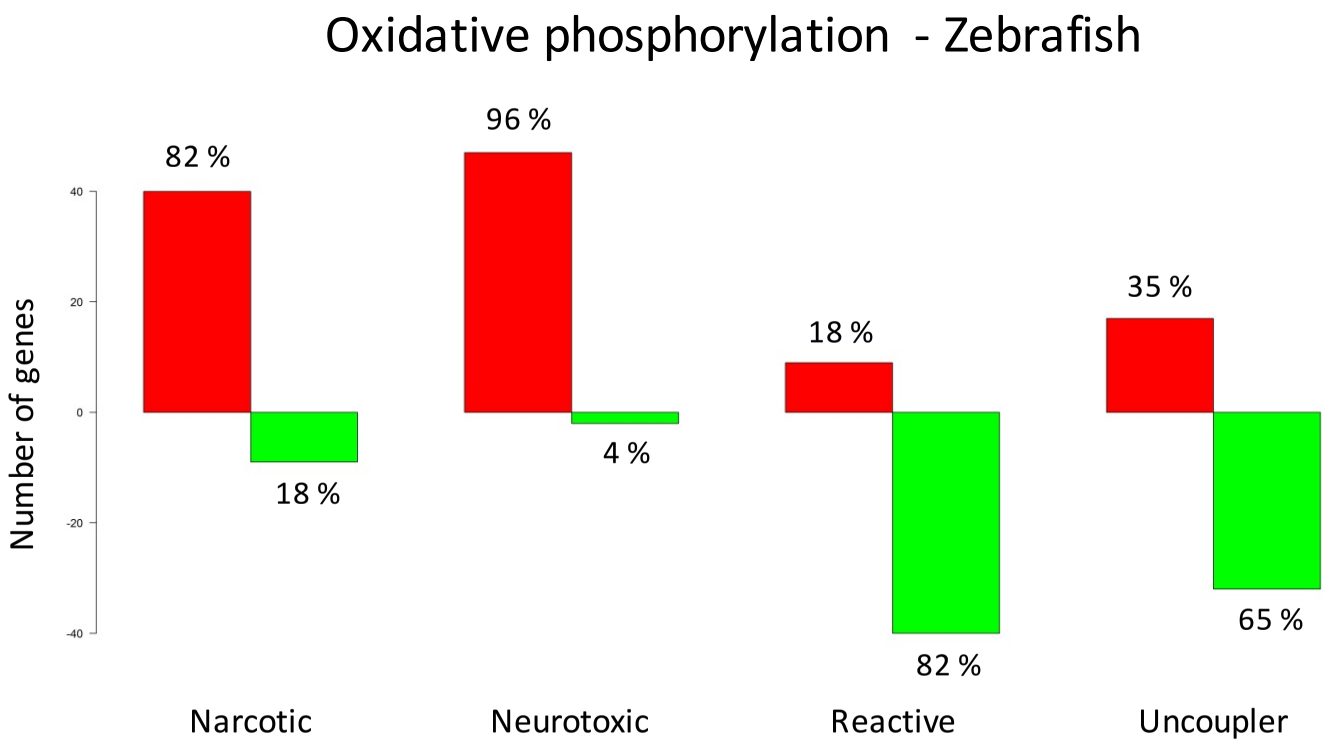


Figure 2.11: Number of differentially expressed genes involved in oxidative phosphorylation responding to each of the different chemical classes. Red and green indicate up- and down-regulated genes, respectively. Most of the genes have a positive expression in narcotic and neurotoxic and a negative expression in reactives and uncouplers. Y-axis shows the number of genes differentially expressed (negative numbers to be considered as absolute).

The glycolysis/gluconeogenesis pathway within the dataset of the zebrafish embryo system was found to be down-regulated in uncouplers compared to all the other classes, while in reactives and narcotics it was down-regulated compared to neurotoxics as confirmed by the heatmap (fig. 2.12). I identified 78% and 72% of the genes involved in this pathway to to be down-regulated in uncouplers and reactives, respectively, while 64% and 83% of the genes were up-regulated in narcotics and neurotoxics, respectively (Fig. 2.13).

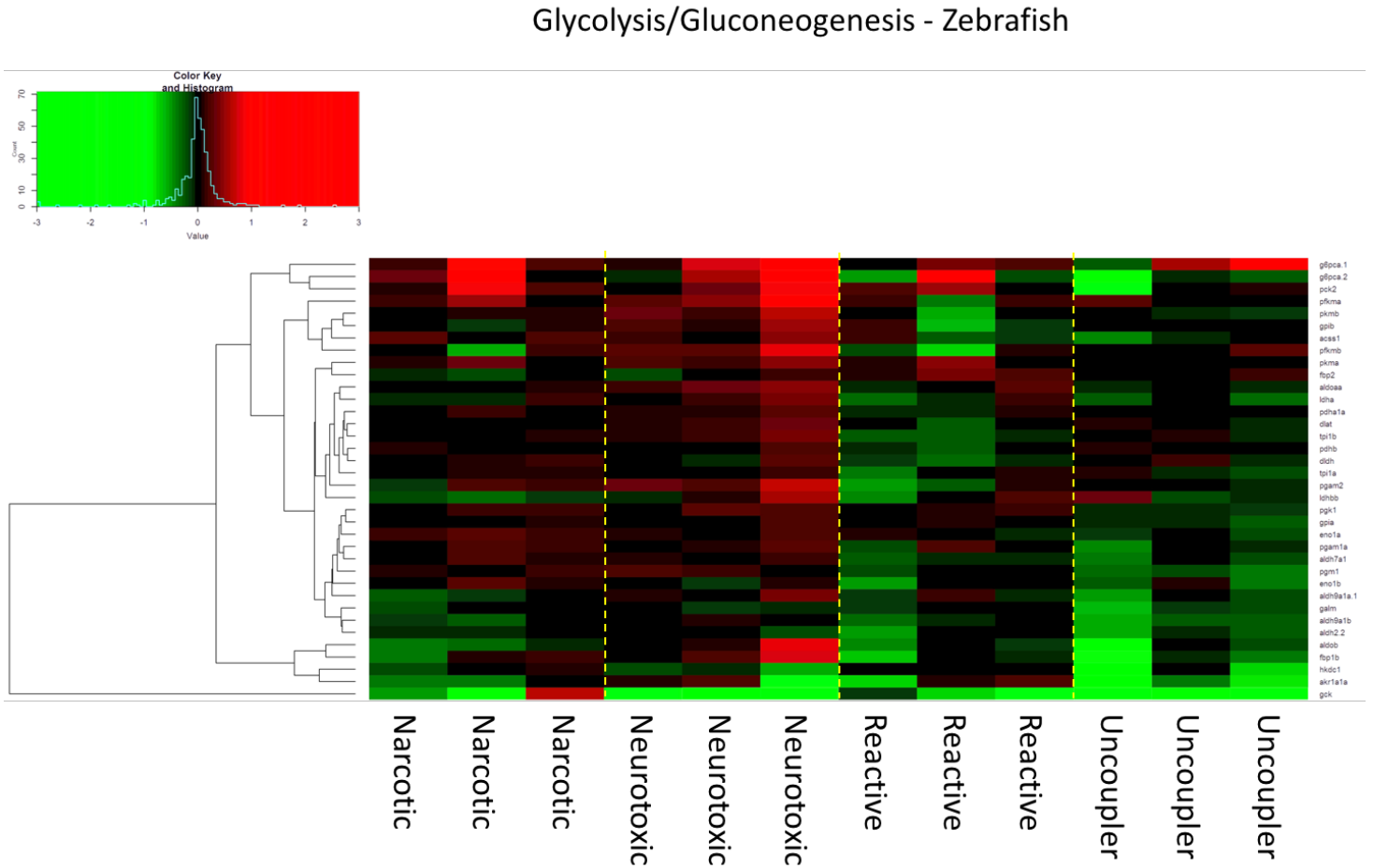


Figure 2.12: Gene expression level of the genes involved in glycolysis/gluconeogenesis in each of the different chemical classes. Red and green refer to up and down regulation, respectively. The heatmap reflects the results obtained by our gene expression and functional analysis with uncouplers showing a down-regulation compared to all the other classes. Heat-map has been performed using Spearman correlation as distance method and average as the clustering method for the genes.

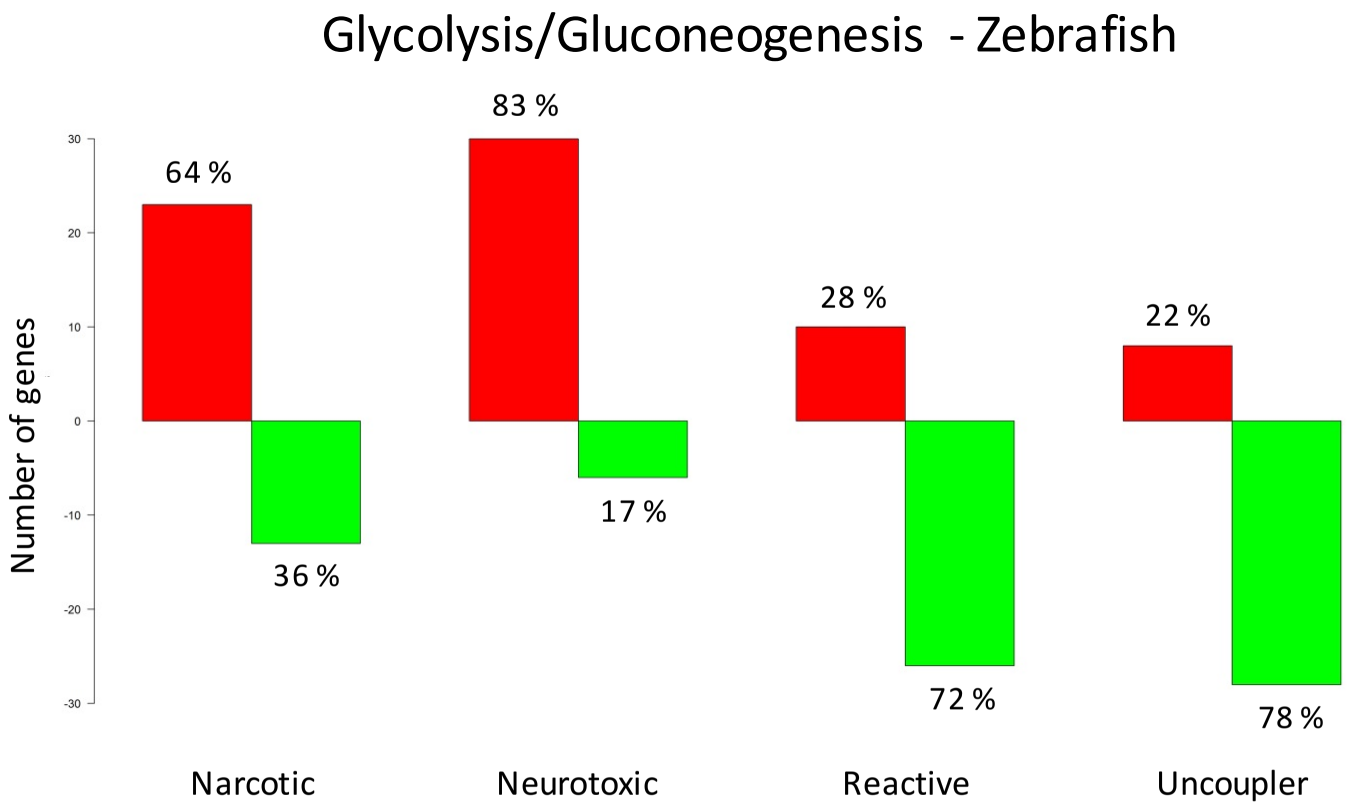


Figure 2.13: Degree of differentially expressed genes involved in glycolysis/gluconeogenesis in each of the different chemical classes. Red and green indicate up and down regulated genes, respectively. Most of the genes have a positive expression in narcotics and neurotoxics and a negative expression in reactives and uncouplers. Y-axis shows the number of genes differentially expressed (negative numbers to be considered as absolute).

Steroid biosynthesis, within the dataset of the zebrafish embryo system was found to be down-regulated in uncouplers compared to all the other classes as confirmed by the heat-map (fig. 2.14). Moreover, I found 96% of the genes involved in this pathway to be down-regulated in the uncoupler while in narcotics, neurotoxics and reactives 65%, 53% and 53% of the genes were found to be up-regulated, respectively (fig. 2.15).

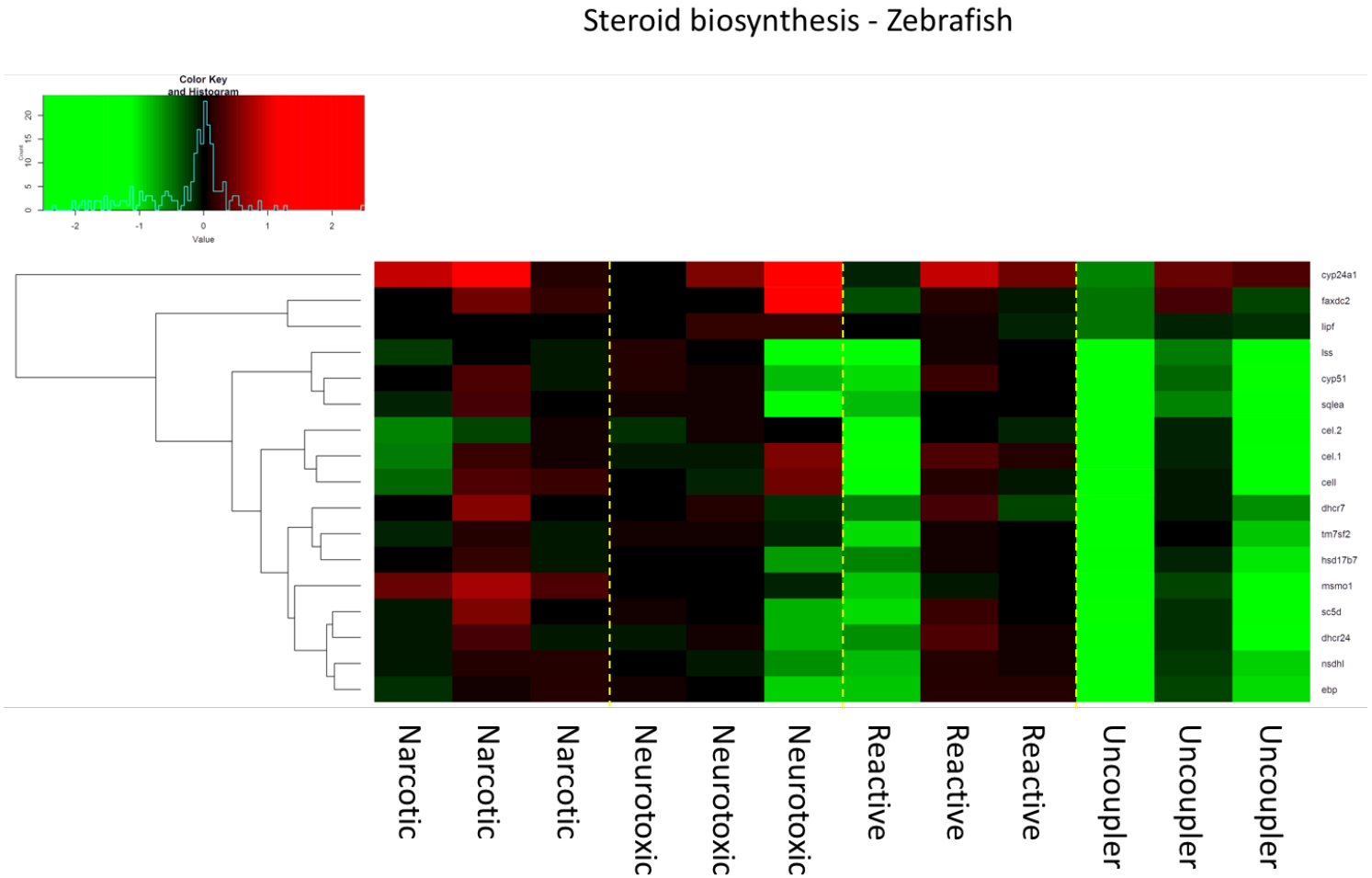


Figure 2.14: Gene expression level of genes involved in steroid biosynthesis in each of the different chemical classes. Red and green refer to up and down regulation, respectively. The heatmap reflect the results obtained by our gene expression and functional analysis with uncouplers showing a down-regulation compared to all the other classes. Heat-map has been performed using Spearman correlation as distance method and average as the clustering method for the genes.

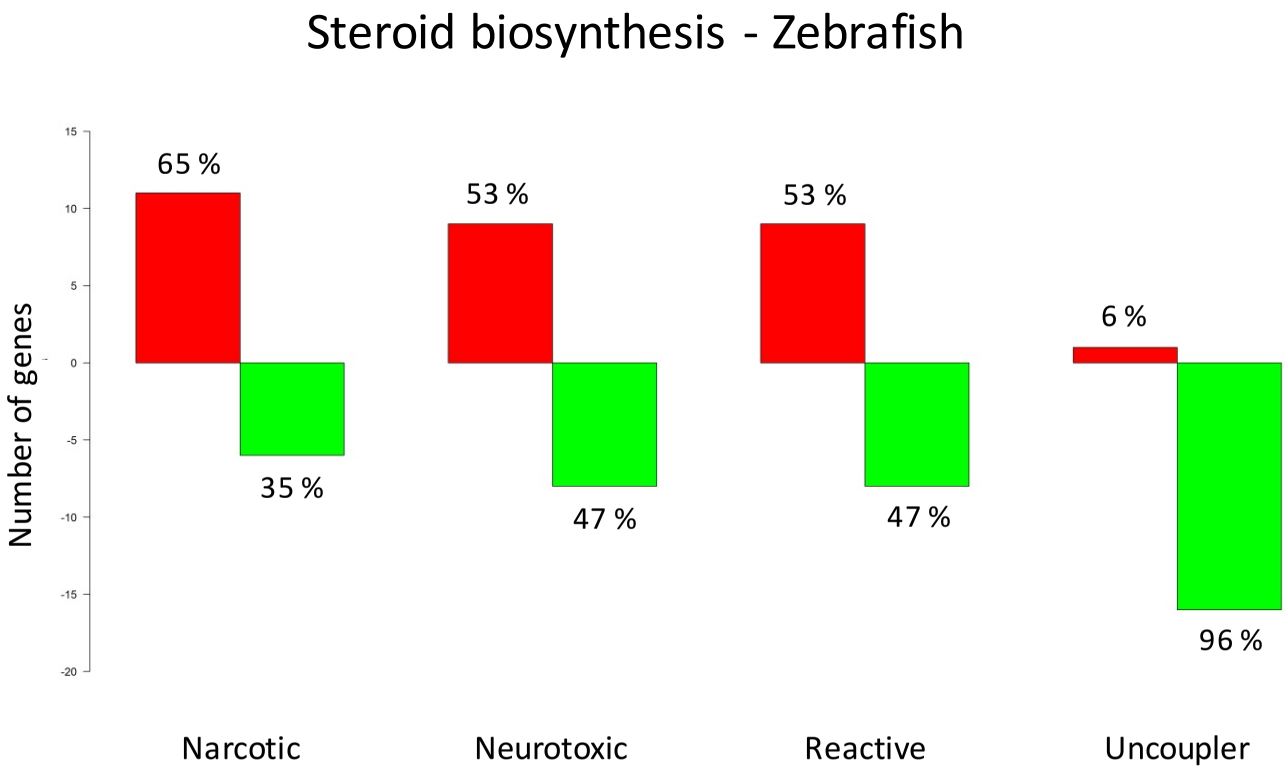


Figure 2.15: Degree of differentially expressed genes involved in steroid biosynthesis in each of the different chemical classes. Red and green indicate up and down regulated genes, respectively. Most of the genes have a negative expression in the uncouplers and a positive expression in narcotics while there is a good ratio within neurotoxics and reactives. Y-axis shows the number of genes differentially expressed (negative numbers to be considered as absolute).

Ribosome biogenesis, within the dataset of the zebrafish embryo system, was found to be down-regulated in neurotoxic compared to all the other classes as confirmed by the heat-map (Fig. 2.16). I showed that 93% of the genes involved in this pathway to be down-regulated in neurotoxics while narcotics, uncouplers and reactives were found to have 41%, 55%, and 41% of the genes, respectively, up-regulated (Fig. 2.17).

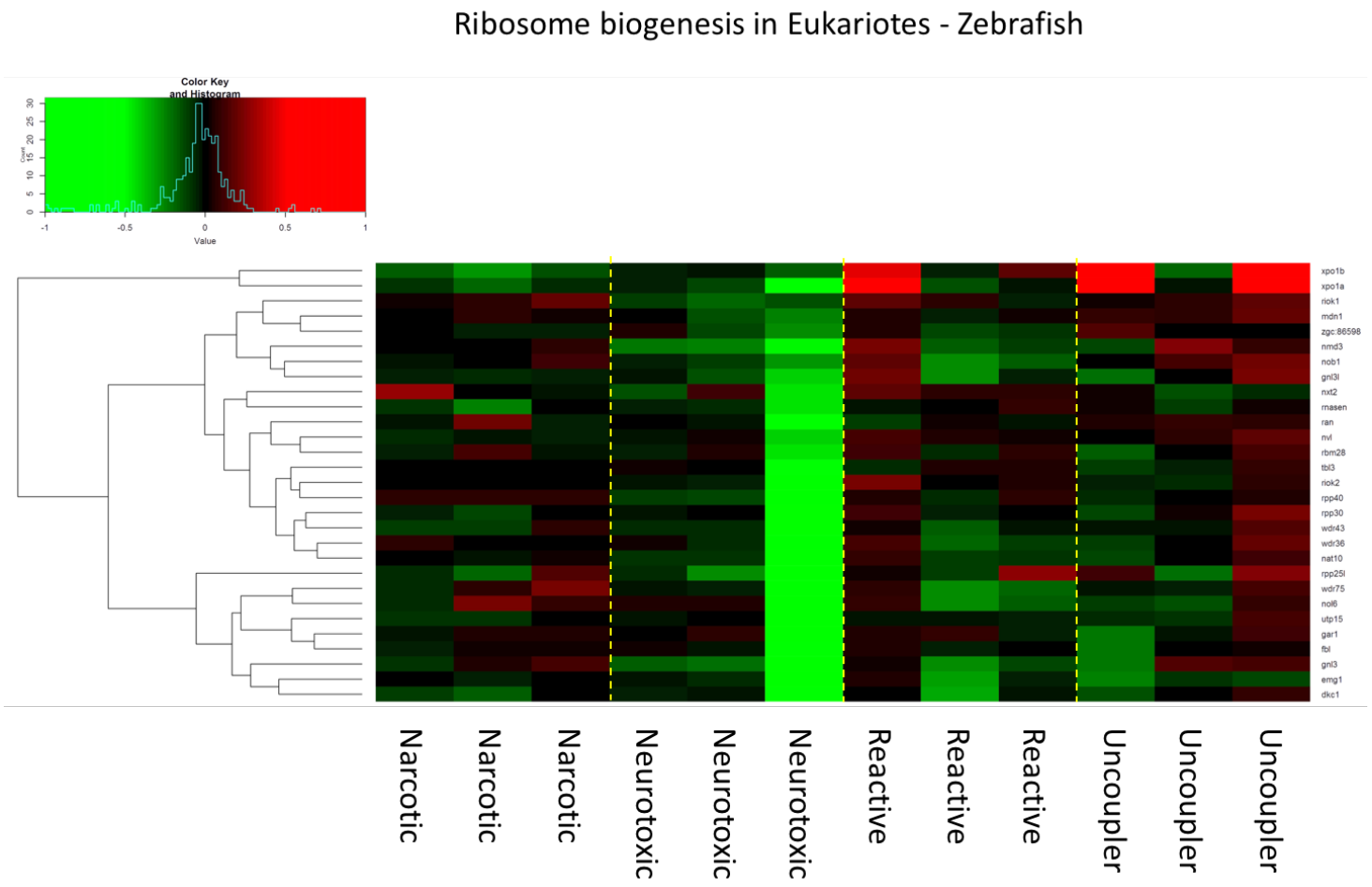


Figure 2.16: Gene expression level of genes involved in in ribosome biogenesis in each of the different chemical classes. Red and green refer to up and down regulation, respectively. The heatmap reflect the results obtained by our gene expression and functional analysis with neurotoxics showing a down-regulation compared to all the other classes. Heat-map has been performed using Spearman correlation as distance method and average as the clustering method for the genes.

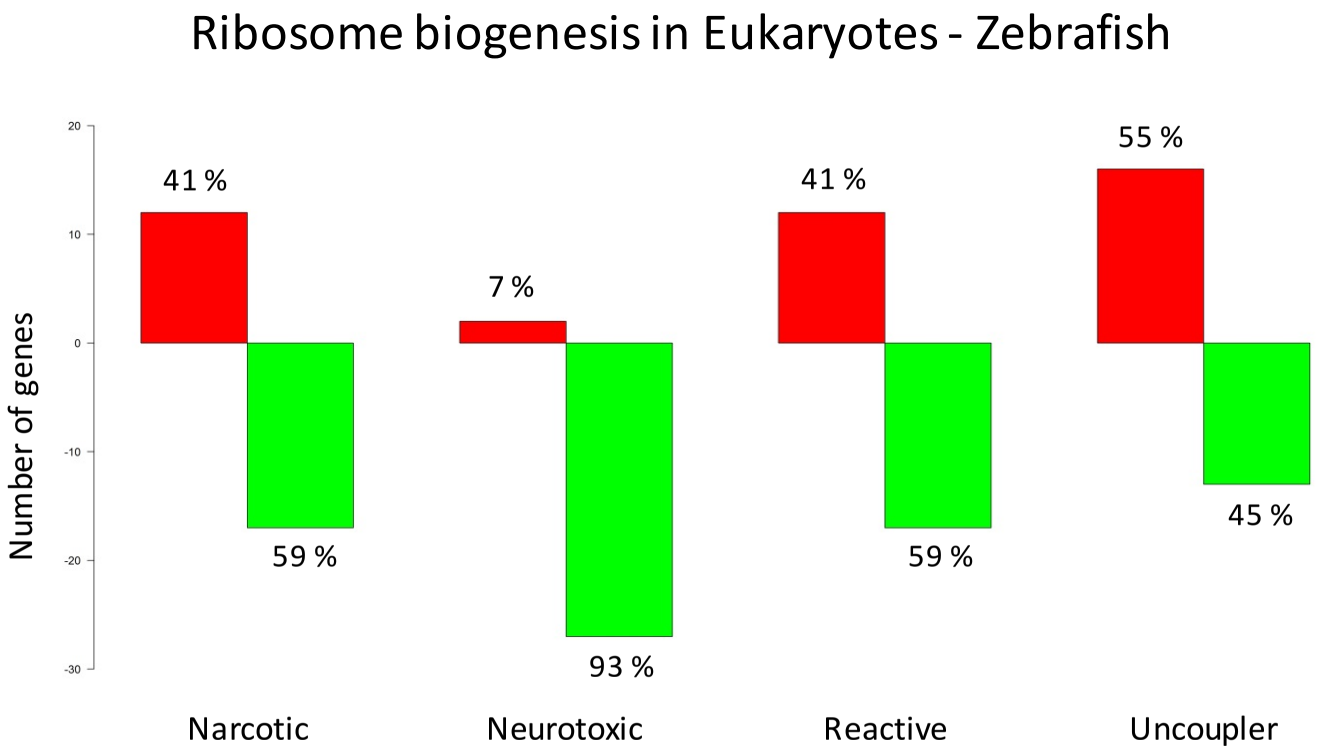


Figure 2.17: Degree of differentially expressed genes involved in ribosome biogenesis in each of the different chemical classes. Red and green indicate up- and down-regulated genes, respectively. Most of the genes have a negative expression in neurotoxics while in the other classes there is a balanced ratio of positively and negatively expressed genes. Y-axis shows the number of genes differentially expressed (negative numbers to be considered as absolute).

On the other hand, ribosome biogenesis within the dataset of the trout gill cell line system was found to be down-regulated in reactives compared to narcotics and uncouplers as confirmed by the heat-map (Fig. 2.18). Moreover, I found 92% of the genes involved in this pathway to be down-regulated in the reactives while in narcotics and uncouplers 87% and 97% of the genes, respectively, were up-regulated (fig. 2.19).

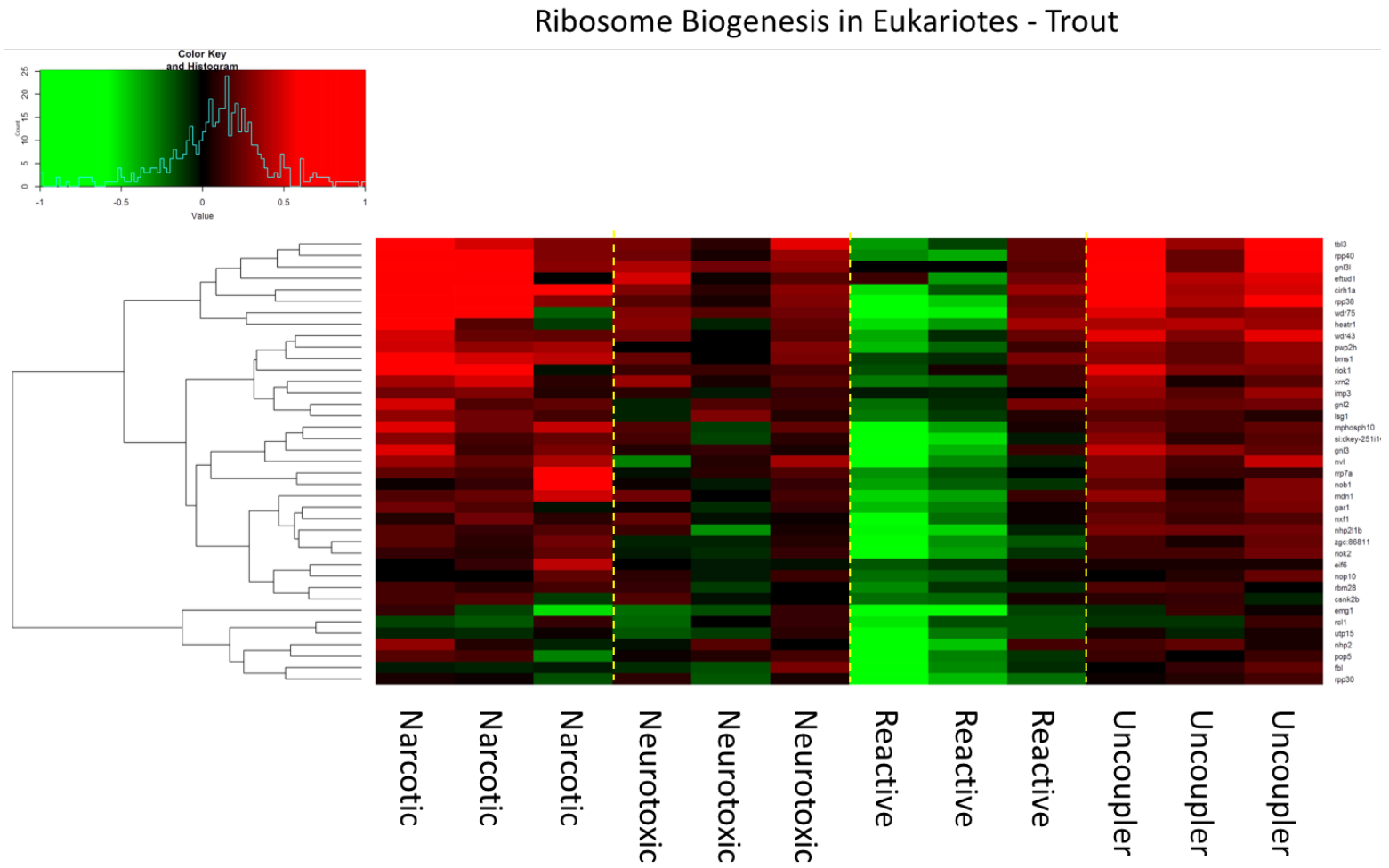


Figure 2.18: Gene expression level of genes involved in ribosome biogenesis in each of the different chemical classes. Red and green refer to up and down regulation, respectively. The heatmap reflect the results obtained by our gene expression and functional analysis with reactives showing a down-regulation compared to narcotics and uncouplers. Heat-map has been performed using Spearman correlation as distance method and average as the clustering method for the genes.

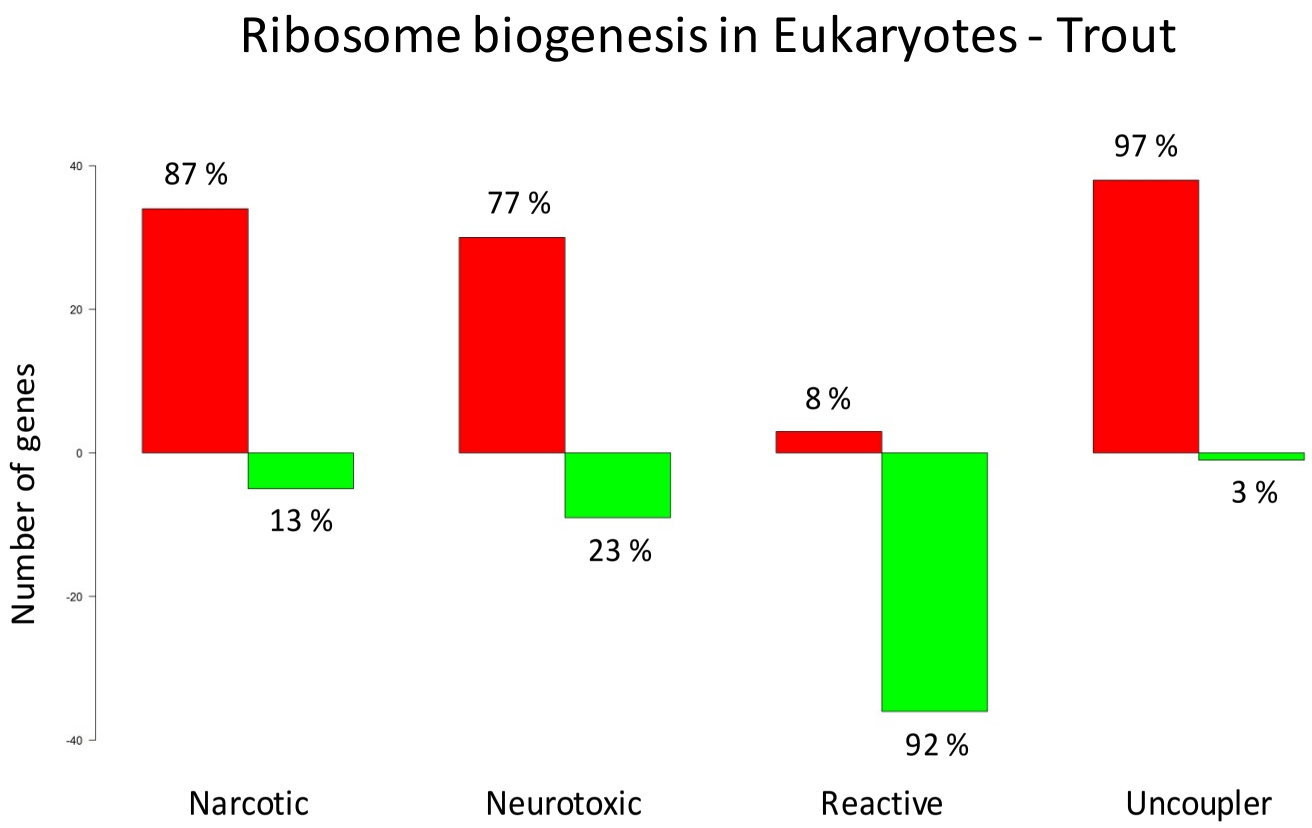


Figure 2.19: Degree of differentially expressed genes involved in ribosome biogenesis in each of the different chemical classes. Red and green indicate up and down regulated genes, respectively. Most of the genes have a negative expression in the reactives and a positive expression in all the other classes. Y-axis shows the number of genes differentially expressed (negative numbers to be considered as absolute).

2.5 Discussion

My work has led to the identification of biomarkers of chemical MoA in two biological systems commonly used for toxicity testing. The relevance of these findings is to promote the use of early markers in zebrafish embryos and a fish gill cell line to identify toxic chemicals. I believe that ultimately, if these markers will be validated in a larger set of chemicals our work would have contributed to reduce or replace the use of adult fish for toxicity testing.

My first analysis sought to characterise the functional profile of *in vitro* and *in vivo* exposures to single environmentally relevant chemicals. Results show that we were able to identify transcriptional signatures underlying chemical mechanisms of action for most of the high dose chemicals. The inability of low dose chemicals to elicit many meaningful transcriptional responses even when considering high FDR threshold, may be due either to the dose chosen or the fraction of bioavailable chemical being too low. Since we used a modified culture medium to increase chemical bioavailability, as described by Schirmer, Tanneberger and collaborators[74,165,166], I reasoned that low dose chemicals were ineffective because of the low concentration chosen. Remarkably, I identified most of the single chemical exposures to affect biological functions associated with embryo developmental pathways. This is consistent with the finding that some of these chemicals (i.e. DEP, PCP and Permethrin) have already been shown to induce developmental abnormalities in zebrafish embryos. Kim et al, showed that exposure of zebrafish embryos to diethylphthalate (DEP) caused early embryonic death and impairment of skeletal development[188]. PCP was shown to induce developmental delay of zebrafish embryos in a concentration-dependent manner[189] and DeMicco et al. showed that craniofacial abnormalities were generated by permethrin at doses approaching the LC_{50}[190]. One of the biggest challenges in defining the molecular mechanisms underlying chemical toxicity is given by the ability of the chemical itself to affect multiple pathways. In this context, it is of paramount importance to

discriminate between direct and indirect effects because of two main reasons. First, chemicals may have not specific targets leading to what is called “systemic toxicity” which may results in different biological functions to be affected simultaneously[191]. The second reason is associated with pathway crosstalk where deregulation of a given biological pathway may arise because it shares components of the signal transduction with another pathway[192]. Despite we didn’t investigate about potential crosstalk between the pathways identified to be deregulated by the single chemical exposure, our results are consistent with the findings of previous studies confirming we are able to characterize molecular mechanisms for this panel of chemicals.

The most important discovery that emerged from this work however, is the demonstration that the transcriptional response in both *in vitro* and *in vivo* systems is diagnostic of the chemical class. As chemicals share similar mechanism of action I tried to address whether we could characterise functional profile of a given MoA. I identified pathways to be diagnostic of chemical class. Glycolisys/gluconeogenesis and steroid biosynthesis pathways were found to be diagnostic of uncoupler chemicals in the zebrafish embryo system; these pathways were strongly down-regulated in relation to the other chemical classes and 78% and 96% of the genes involved in these pathways were found to be down-regulated. Ma et al., demonstrated that 2,4-DCP modulate transcription of steroidogenic genes in adult female zebrafish resulting in lower levels of 17β-estradiol (E2) and the down-regulation of aromatase (CYP19A) genes[193]. The finding that Glycolisys/gluconeogenesis is so widely affected is fully consistant with the fact that uncouplers do severely affect respiration and specifically the electron transport chain in the oxidative phosphorylation pathway.

Most of the genes in the ribosome biogenesis pathway (93%) were found to be down-regulated in zebrafish embryos exposed to neurotoxic chemicals. This finding is consistent with the

observation by Jeffries et al., that ribosome biogenesis is affected by permethrin in Delta smelt (*Hypomesus transpacificus*)[194] and by Connon et al., that identified genes encoding for ribosomal protein to be modulated by the exposure to Esfenvalerate in Delta smelt[195]. In the trout cell line, on the other hand, ribosome biogenesis was found to be down-regulated in reactives compounds with 92% of down-regulated genes. I can hypotesise that this transcriptional signature may have a functional effect on ribosome biogenesis. In fact, Wang et al., demonstrated that concentrations of acrolein ranging between 0 – 100 uM were able to impair ribosome biogenesis and polysome formation in human cancer cells[196]. This effect may be conserved across a wide range of species since Golla et al., identified several genes involved in ribosome biogenesis to be down-regulated in budding yeast after exposure to 0.4 mM of AA[197]. However, the concentration they used are around 2 orders of magnitude greater than the ones we used in our study. It is interesting to note that the ribosome biogenesis pathway is modulated by a different set of chemical classes in the two different system suggesting the presence of different compensatory mechanisms in the two biological systems.

One of the challenges associated with the classification of chemicals based on transcriptional signatures is linked with the presence of chemical mixture. As humans and animals are exposed to more than one chemical in the natural environment, characterizing molecular mechanisms of chemicals belonging to mixture can be very difficult because of the presence of antagonistic or synergic effects that may influence the toxicity of each of the compounds in the chemical mixture. Garcia-Reyero et al., assessed chemical mixture effects on Daphnia transcriptome by comparing molecular responses in the mixtures with those of the single chemical exposure[198]. Interestingly, they identified shared gene function between mixtures and single chemicals but unique gene functions arose from the mixture suggesting additive and nonadditive mixture effects has a key role in mixture toxicity. In this context, the molecular mechanisms I identified

for the single chemicals may differ when considering these compounds as part of environmental mixture. However, to better understand chemical mixture toxicity, knowledge of gene behaviour in both single chemicals and mixture should be investigated in parallel.

These results suggest the zebrafish embryo and the fish gill cell line in combination with a global method of quantifying transcripts offers a useful approach to generating mechanistic information of specific chemical exposures. This can identify suitable biomarkers that can assist the improvement of ERAs.

Chapter 3: Predicting *in vivo* toxicity from *in vitro* transcriptional responses following chemical exposure

The analysis strategy and all the in silico analysis in this chapter have been performed by myself

3.1 Abstract

The lack of detailed toxicology data for thousands of chemical compounds currently being released in the environment is in part due to the challenge of generating these data using traditional animal testing. High-throughput *in vitro* testing offers a valid alternative to animal testing. *In vitro* to *in vivo* extrapolation (IVIVE) refers to the use of *in vitro* testing to infer the effects of chemicals on whole organisms. In this chapter, I precisely address whether the transcriptional state of a gill cell line exposed to a given chemical can be used as a biosensor to predict toxicity in a zebrafish embryo. This is achieved by building an approach able to identify signatures associated either with toxicity or lipophilicity. By developing a regression model linking *in vitro* gene signatures that are independent of compound lipophilicity to whole embryo toxicity I have been able to identify gene signatures with the ability to discriminate between the different MoA. These results support the view that cell lines have the potential to inform about toxicity in whole organism and have the potential for the development of a Mode of Action assignment framework.

3.2 Introduction

Around 10 million tons of toxic chemicals are released into the environment by industries each year[199]. Due to the time and expense of generating data using conventional animal tests, health data for thousands of compounds currently in use are missing. Moreover, current legislation has become stricter regarding the use of whole organisms for toxicity testing since the primary endpoint is mortality and fish suffer distress and pain[41]. For these reasons, more effort was addressed towards the development of alternative methods for toxicity testing. High-throughput screening assays (HTS) have the ability to test many chemicals simultaneously offering a potential tool for prioritization applications[200], to investigate mechanisms of toxicity[201] as well as reducing the number of animals used. The ToxCast and Tox21 programs, established by the US EPA, have screened more than 1800 environmentally relevant chemicals carrying out more than 700 HTS assays covering around 300 cell signalling pathways and represent the best example of how HTS can be used to improve ecotoxicological assessment[202,203]. However, HTS is still at an early stage in the field of environmental toxicology compared with its application for drug development[202]. Most of the studies have focused on cellular and mechanistic pathways in model organisms as *D. rerio*, a suitable model for risk assessment[204]. Within the ToxCast program 309 chemicals have been screened to assess general phenotypic endpoints linked to embryo development and viability and results showed 62% of the compounds to be toxic to the developing zebrafish[205]. HTS have also been applied to cell cultures. George et al., used a gill cell line (RTgill-W1) in parallel with zebrafish embryos to assess how the shape of silver (Ag) nanoparticles influenced the toxicity profiles of cell viability and superoxide generation[206]. While the zebrafish embryos represent a well-established and accepted alternative method for toxicity testing, the RTgill-W1 cell line has not been fully validated yet. Advantages, limits and applications of cell lines and early stage embryos have been already extensively discussed in the previous chapter[164,207–211].

Two major challenges hamper the use of *in vitro* toxicity data for hazard risk assessment: a toxycodinamic problem, where interactions and mechanism happening between tissues and organs cannot be modelled by *in vitro* systems, and a toxycokinetic problem, where *in vivo* efficiency of exposure and chemical bioavailability is determined by biokinetic processes which may be either lost or different *in vitro*[212]. Absorption, distribution, metabolism and excretion (ADME) properties, which determine *in vivo* the relationship between the administered toxic dose and the internal dose or the effective free concentration at the target site, are missing in the *in vitro* systems. It has been demonstrated *in vitro* that considerable amount of chemicals become unable to trigger a toxic action due to binding to serum or cells[213]. To fill this gap, two main solutions are available which are 1) to increase the complexity of *in vitro* system to account for interactions between tissues and 2) to apply modelling techniques to simulate the behaviour of the complex system by using *in vitro* data to look for the best parameter values[214]. For these reasons, *in vitro* to *in vivo* extrapolation (IVIVE) is particularly challenging[215,216].

The present study fits the vision of an ecotoxicology framework that minimise animal testing. The study was designed to assess the potential of a fish cell line as alternative method for toxicity testing on a panel of chemicals which are of environmental interest. More specifically, I successfully proved that the transcriptional response of a fish gill cell line following chemical exposure is informative of molecular and toxicity response in a whole organism such as the zebrafish embryo. These results further support the use of a fish gill cell line as an alternative for toxicity testing and provide a suitable *proof of concept* approach for IVIVE.

3.3 Materials and methods

Experimental design, data acquisition and data pre-processing and normalization have been extensively explained in chapter 2.

3.3.1 Identification of gene signatures lipophilicity dependent and independent

Statistical analyses were all performed within the statistical environment R. Gene signatures correlated with EC_{10} and K_{ow} have been identified using quantitative SAM[117] as it is implemented in the "samr" package. Quantitative SAM identifies gene signatures that correlates with a given response variable. Within each biological system, gene signatures were assessed for their correlation with EC_{10} and K_{ow} values which have been used as response variables. For IVIVE, *in vivo* EC_{10} has been used as response variable to identify *in vitro* transcriptional signatures correlated with *in vivo* toxicity. Significance was assigned for genes with an FDR < 5%.

3.3.2 Regression analysis

A Regression analysis between EC_{10} and K_{ow} within each biological system have been run to identify residuals that cannot be explained by lipophilicity. Regression analysis and residuals computation have been achieved using MS Excel. Gene signatures have been identified using quantitative SAM and a 5% FDR threshold was applied. For IVIVE, residuals of the correlation between *in vivo* toxicity with lipophilicity have been used as response variable to identify *in vitro* gene signatures correlating to the component of the transcriptional response which is free of lipophilicity effects.

3.3.3 Functional analysis

All the functional enrichment was attained using DAVID webservice[118]. Functional analysis of gene signatures correlated with EC_{10} and K_{ow} was achieved selecting GO terms and KEGG pathways from a functional clustering approach as it is implemented in DAVID and applying a 1% FDR cut-off. Functional clustering groups together similar annotations which makes the biology clearer. Non-redundant terms were then arranged into wider functional domains according to the KEGG functional levels. Functional analysis of gene signatures correlated with residuals was achieved retrieving KEGG pathways at 1% FDR with a gene count > 5.

3.3.4 Predictive analysis

To assess the ability of gene signatures correlated with EC_{10}, K_{ow} and residuals to predict chemical class MoA we employed a K-nearest neighbours (KNN) classification algorithm. KNN is a simple and intuitive method widely used in classification problems given its low error rate. Giving a sample whose category is unknown, the classification algorithm will find the K-nearest neighbours by computing similarities (Euclidean distance) between samples in the training data. I will then get the category of the new sample according to the K-nearest neighbours. KNN is the simplest and straightforward classification algorithm and has many advantages over other algorithms as for example it being a non-parametric method. Here we applied a KNN classification algorithm using the "class" package within the statistical environment R. The K value was set equal to 4 and a leave-one-out cross validation (LOOCV) method was used to measure the error rate. Results have been reported using bar-charts.

3.4 Results

3.4.1 Identification of gene expression signatures that are correlated to toxicity

In the previous chapter (chapter 2) I have identified transcriptional signatures in both *in vitro* and *in vivo* systems that are associated to chemical classes defining general MoA categories. Here I asked whether we can identify transcriptional signatures that are linked to compound toxicity, irrespective of the MoA.

I first set to identify gene signatures which correlate with chemical toxicity, expressed as EC_{10} (effective concentration at which the 10% of the population exhibit a response), in both *in vitro* and *in vivo* systems. Remarkably, I could identify a total of 2,477 genes in the gill cell line that correlated to cell line chemical toxicity (1,545 positively correlated and 932 negatively

correlated) (fig. 3.1). In Zebrafish embryos, I could identify 1,415 genes correlated to chemical toxicity (339 positively correlated and 1,076 negatively correlated) (fig. 3.2).

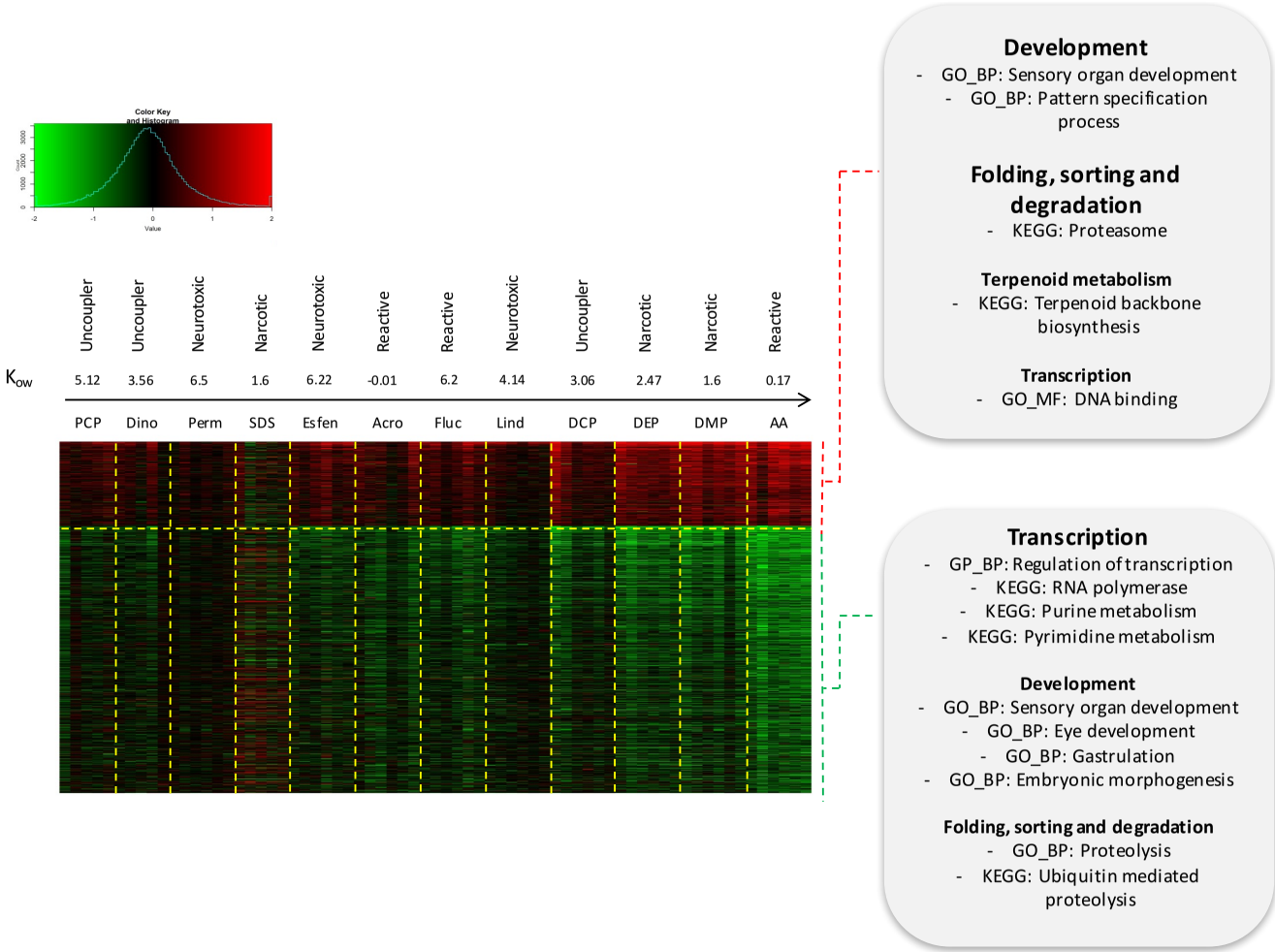


Figure 3.1: Gene signatures that correlates with chemical toxicity within the trout gill cell line system. Red and green refers to genes positively and negatively correlated with toxicity, respectively. A functional analysis of the genes positively and negatively correlated to the EC_{10} is reported. FDR values and genes included in each of the GO terms and Kegg pathways are provided in the supplementary material.

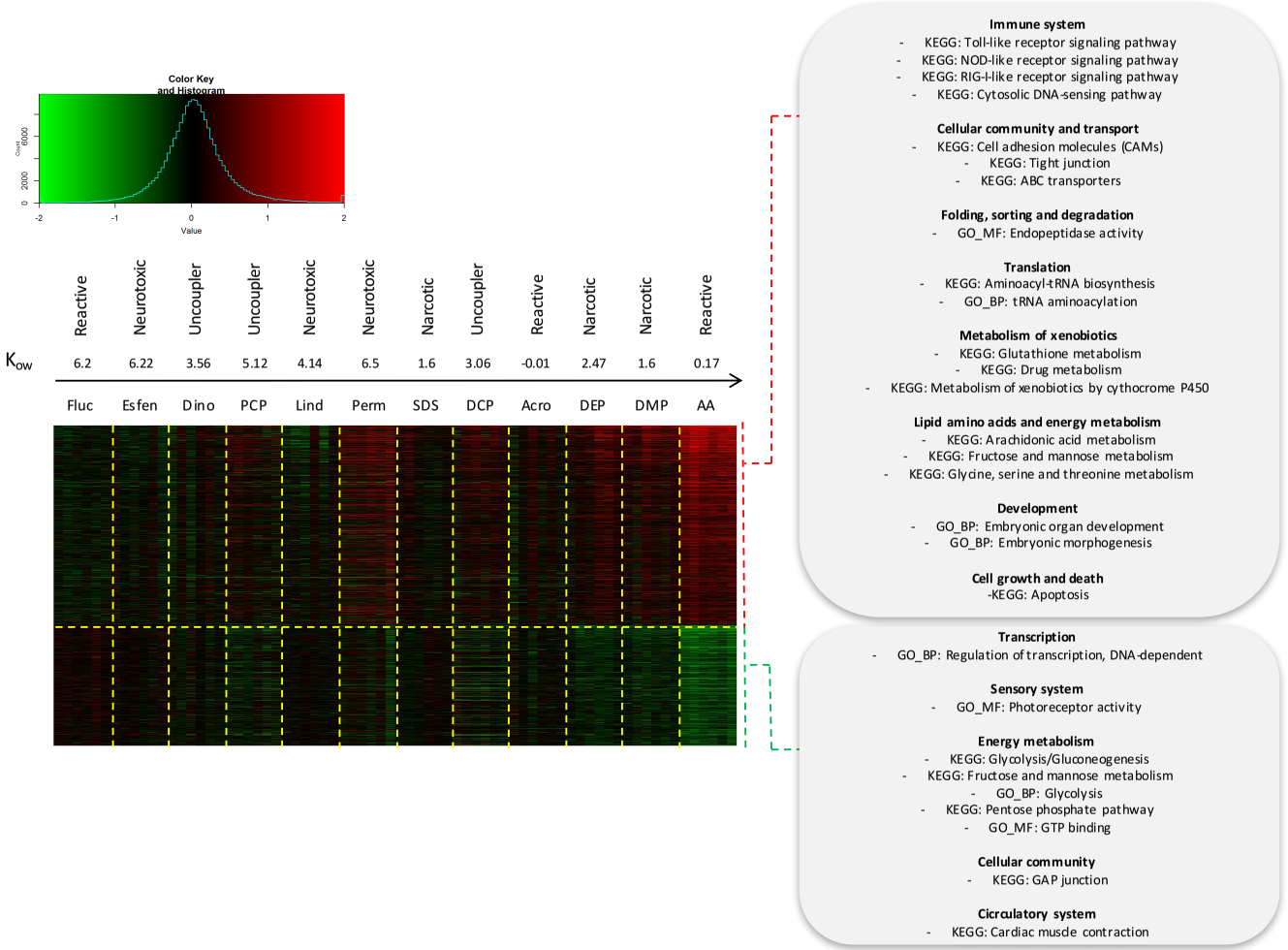


Figure 3.2: Gene signatures that correlates with chemical toxicity within the zebrafish embryo system. Red and green refers to genes positively and negatively correlated with toxicity, respectively. A functional analysis of the genes positively and negatively correlated to the EC_{10} is reported. FDR values and genes included in each of the GO terms and Kegg pathways are provided in the supplementary material.

I then asked whether I could identify transcriptional signatures in the *in vitro* system that correlate with *in vivo* toxicity. A total of 1383 genes (391 positively correlated and 992 negatively correlated) were found to be correlated (fig. 3.3). Interestingly, functional enrichment analysis of *in vitro* transcriptional signatures correlated to *in vivo* toxicity revealed only enrichment in the negatively correlated genes list. More precisely, I identified functions linked to whole organism development such as sensory organ development, embryonic morphogenesis and gastrulation. In addition, functions as transcription, cell growth and cell death (P53 signaling pathway and apoptosis), endocrine system (PPAR signalling pathway) and lipid metabolism (fatty acid metabolism) were also found. These results are consistent with the original hypothesis that a cell line may be informative of the complex transcriptional programs associated to whole organism response.

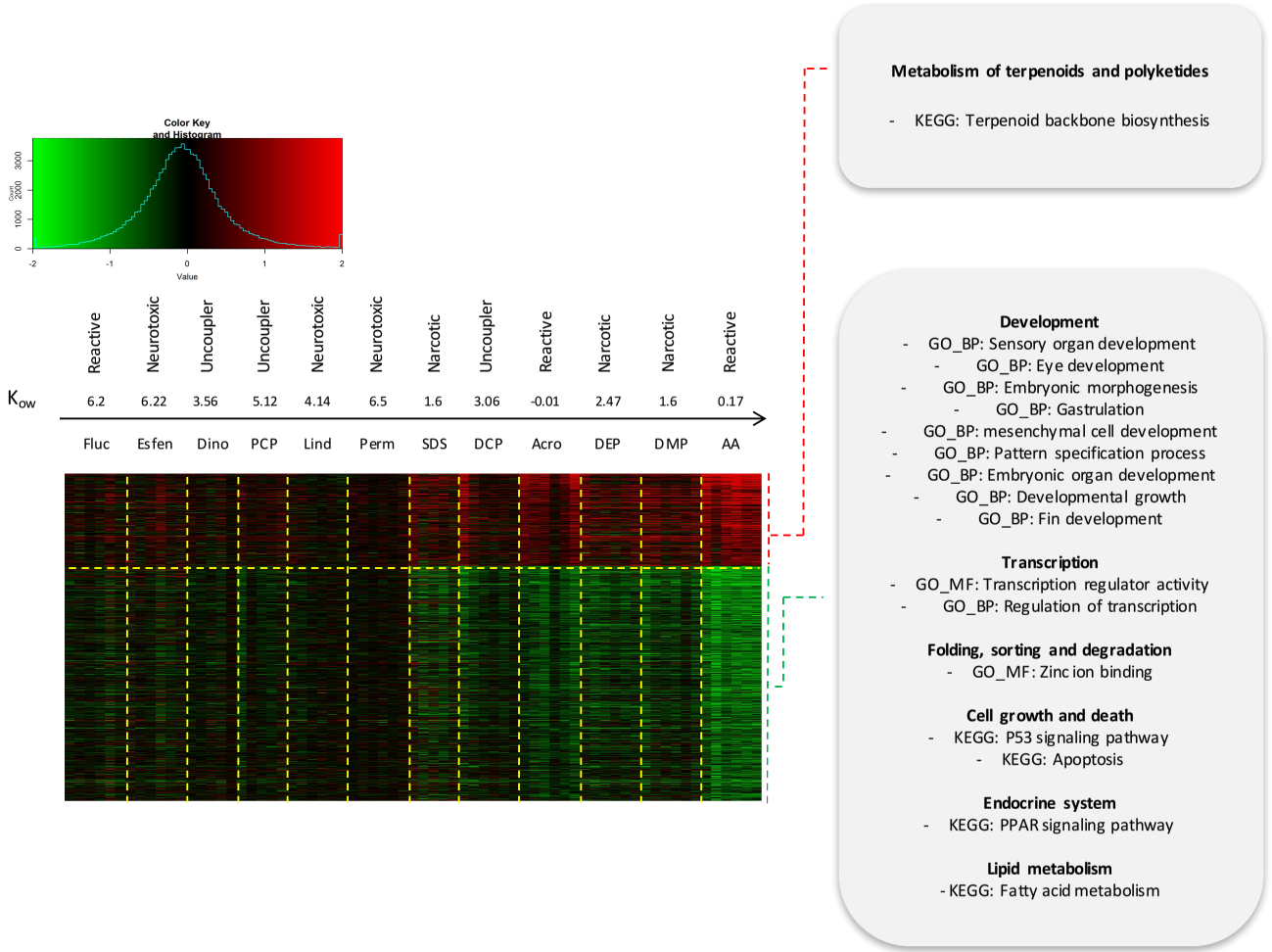


Figure 3.3: The *in vitro* gene signatures that correlate with *in vivo* chemical toxicity. Red and green refers to genes positively and negatively correlated with toxicity, respectively. A functional analysis of the genes positively and negatively correlated to the EC_{10} is reported. FDR values and genes included in each of the GO terms and Kegg pathways are provided in the supplementary material.

3.4.2 The relationship between toxicity and compound lipophilicity

The analysis described above revealed that in both systems molecular response can be informative of toxicity. While this is an interesting finding I reasoned it may simply reflect the important contribution of basal toxicity to the overall toxicity response. Since non-specific basal toxicity is proportional to compound lipophilicity there is a possibility that the transcriptional signatures I just identified reflected at least in part, compound lipophilicity. Deconvolution of specific and basal toxicity is therefore important.

I therefore hypothesise that a fraction of the genes correlated to toxicity may also be correlated with compound lipophilicity.

In order to test this hypothesis, I first identified genes whose expression correlates with logP, a measure of compound lipophilicity, in both *in vitro* and *in vivo* systems. I then computed the overlap between these signatures and the toxicity-related signatures we already developed. Consistent with my original hypothesis, I found a significant overlap between logP positively correlated genes and genes negatively correlated with toxicity and logP negatively correlated genes and genes positively correlated with toxicity (fig. 3.4).

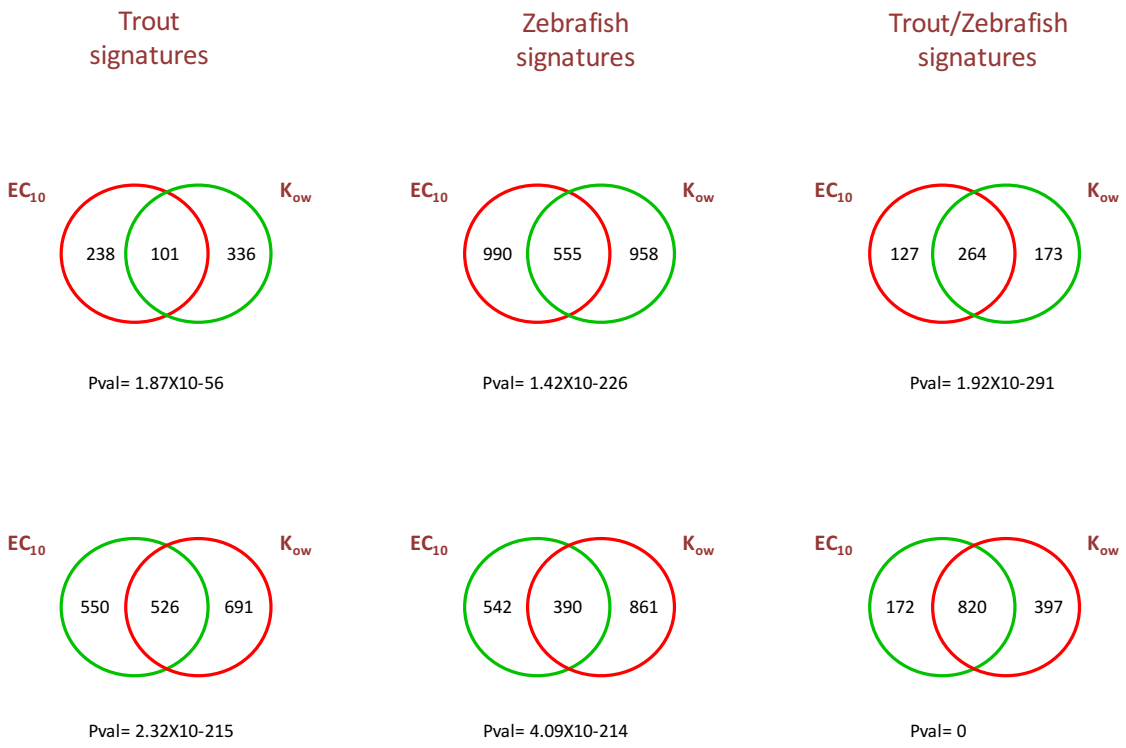


Figure 3.4: Venn diagrams showing gene overlaps between signatures associate with toxicity (EC_{10}) and with lipophilicity (K_{ow}) within each biological system. Gene overlaps are reported also for the cross-species analysis where we identified *in vitro* signatures linked to *in vivo* toxicity and lipophilicity. Red and green circles represent gene signatures positively and negatively correlated respectively.

3.4.3 Residual analysis to identify molecular signatures linked to specific toxicity effects

The analysis described above shows that a significant component of the transcriptional response to toxic chemicals can be explained as a non-specific interaction with biological membranes. The analysis of the overlap between EC_{10} and logP correlated signatures suggest that I may be able to identify excess toxicity signatures. In order to gain more confidence in our ability to identify excess toxicity signatures, I first developed a regression model linking toxicity with lipophilicity (basal toxicity model) (fig. 3.5-3.6) and then computed the residuals, which represent the toxicity that cannot be explained by lipophilicity (fig. 3.7-3.8).

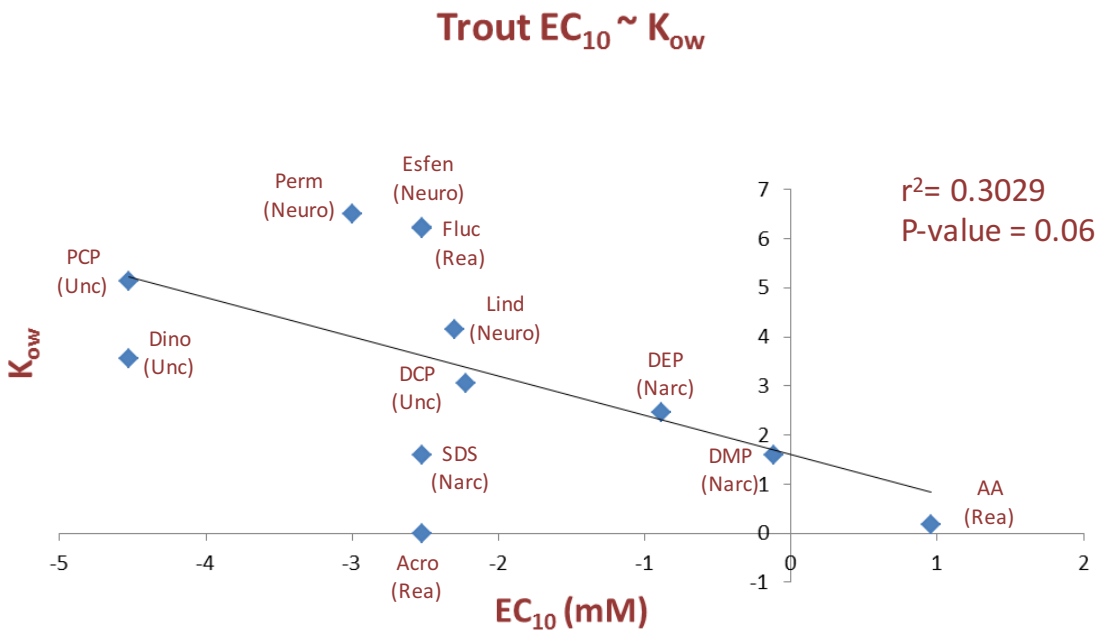


Figure 3.5: Correlation between chemical toxicity and chemical lipophilicity within the trout gill cell line system. The fitted line represents the toxicity that can be explained by the lipophilicity.

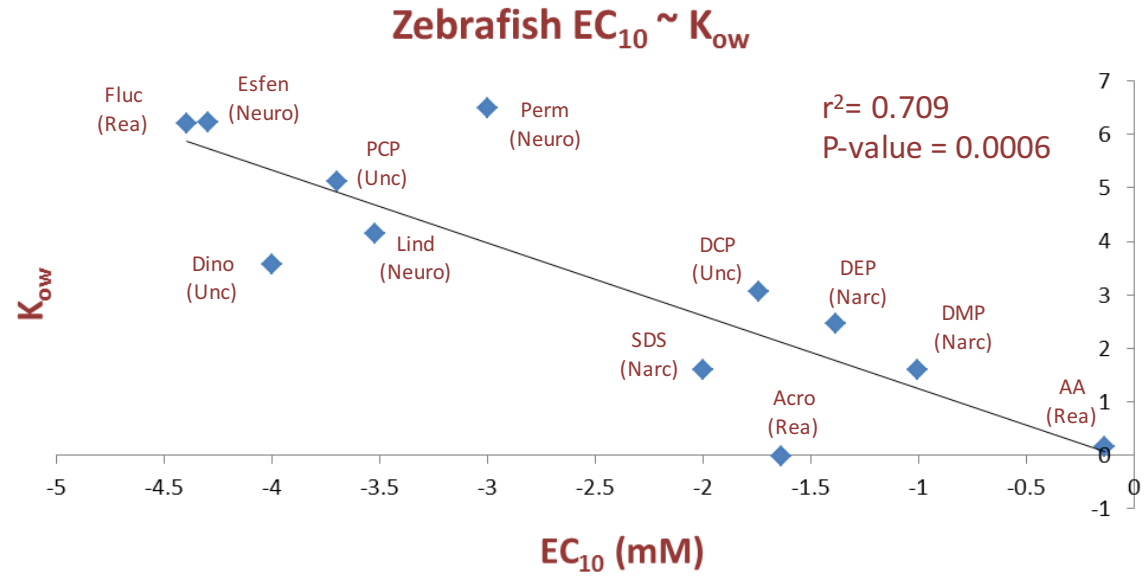


Figure 3.6: Correlation between chemical toxicity and chemical lipophilicity within the zebrafish embryo system. The fitted line represents the toxicity that can be explained by the lipophilicity.

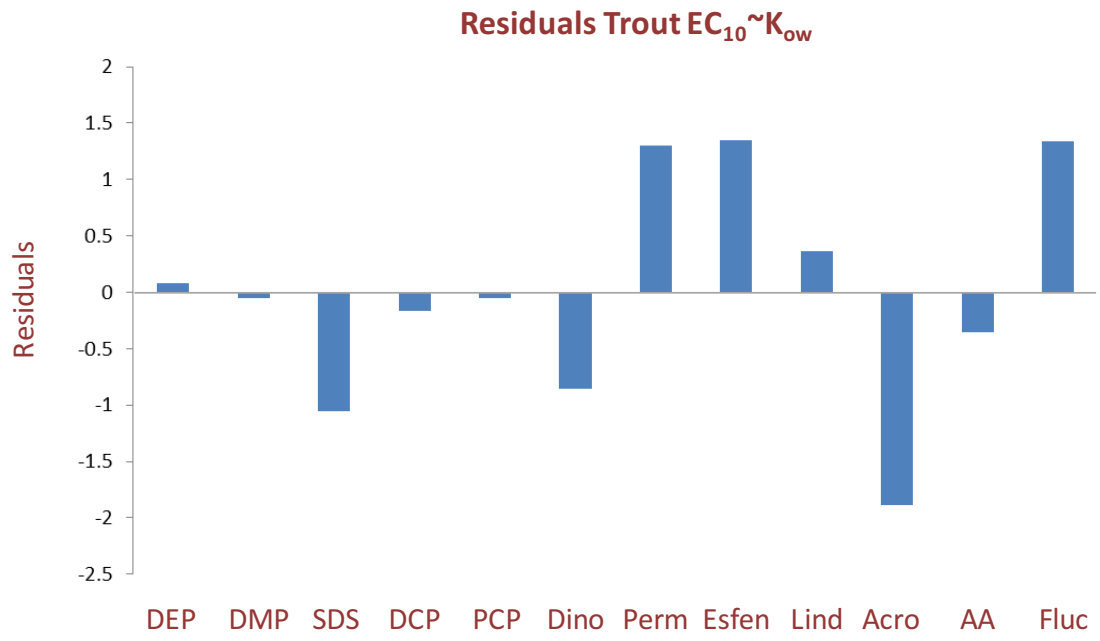


Figure 3.7: Residuals of the correlation between toxicity and lipophilicity within the trout gill cell line system which represent the component of the transcriptional response that cannot be explained by the lipophilicity.

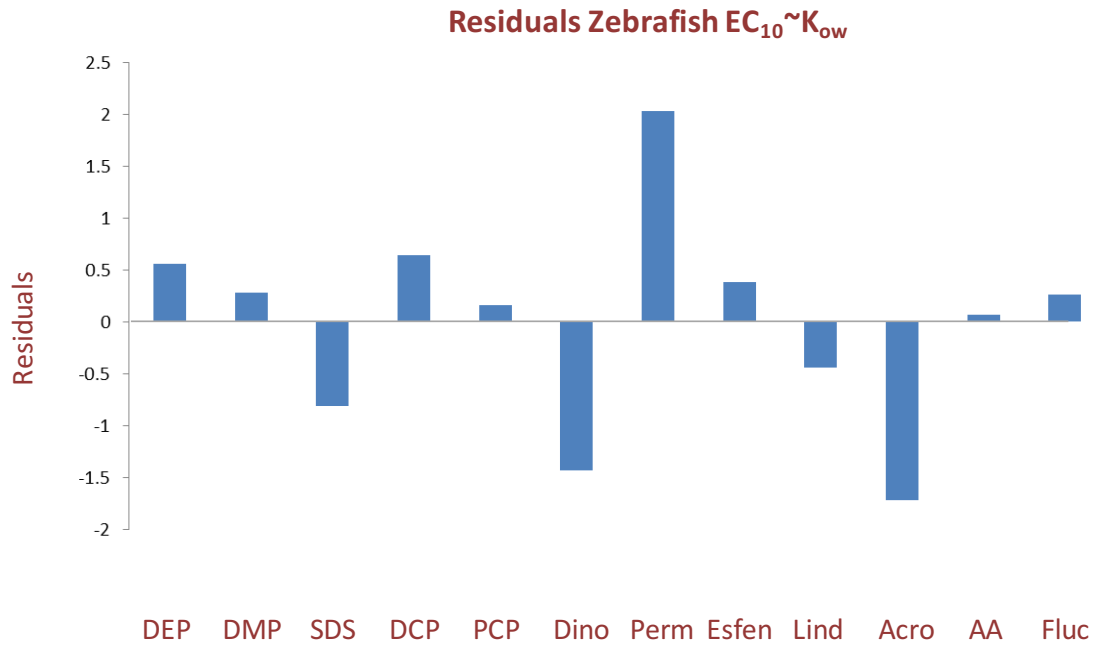


Figure 3.8: Residuals of the correlation between toxicity and lipophilicity within the zebrafish embryo system which represent the component of the transcriptional response that cannot be explained by the lipophilicity.

Finally, I set to identify genes whose expression correlate with the residuals in both *in vitro* and *in vivo* systems. A total of 1,091 genes (733 positively correlated and 358 negatively correlated) and 4,637 genes (2,065 positively correlated and 2,572 negatively correlated) were found to be differentially correlated at 5% FDR with residuals computed in the *in vitro* and *in vivo* models, respectively. 268 *in vitro* genes (208 positively correlated and 60 negatively correlated) were found to be differentially correlated to *in vivo* residuals. A functional enrichment analysis was then achieved to identify biological functions associate with each of the portion of the transcriptional response free from lipophilicity effects. Biological functions enriched of *in vitro* gene signatures correlated with *in vitro* residuals were mainly identified for those signatures positively regulated (table 3.1).

Domain	KEGG ID	Pathway Name	Genes
Nucletoide metabolism	dre00230	Purine metabolism	9
	dre00240	Pyrimidine metabolism	8
Transcription	dre03040	Spliceosome	6
Replication and repair	dre03030	DNA replication	15
	dre03410	Base excision repair	5
	dre03420	Nucleotide excision repair	9
	dre03430	Mismatch repair	5
Signal transduction	dre04010	MAPK signaling pathway	18
	dre04310	Wnt signaling pathway	6
	dre04630	Jak-STAT signaling pathway	5
	dre04350	TGF-beta signaling pathway	8
	dre04020	Calcium signaling pathway	5
Signaling molecules and interaction	dre04060	Cytokine-cytokine receptor interaction	5
	dre04512	ECM-receptor interaction	6
	dre04514	Cell adhesion molecules (CAMs)	5
Transport and catabolism	dre04142	Lysosome	5
Cell motility	dre04810	Regulation of actin cytoskeleton	11
Cell growth and death	dre04110	Cell cycle	9
	dre04210	Apoptosis	5
	dre04114	Oocyte meiosis	8
Cellular community	dre04510	Focal adhesion	10
	dre04520	Adherens junction	7
	dre04530	Tight junction	5
Immune system	dre04620	Toll-like receptor signaling pathway	7
Endocrine system	dre04910	Insulin signaling pathway	6
	dre04920	Adipocytokine signaling pathway	5
	dre04912	GnRH signaling pathway	6

Table 3.1: Functional enrichment of *in vitro* genes signatures found to be differentially correlated with *in vitro* residuals at 1% FDR. In red and green are reported KEGG pathways positively and negatively correlated with residuals, respectively while in brown those pathways enriched of genes both positively and negatively correlated with residuals. FDR values of each of the Kegg pathways are provided in the supplementary material.

Interestingly, most of the zebrafish embryo signatures negatively correlated with residuals were found to be involved in biological functions associated with transcription, translation, DNA replication and repair and folding, sorting and degradation (table 3.2). Functions positively

regulated are instead associated with pathways linked to energy and amino acids metabolism, signal transduction and with immune, endocrine and circulatory system.

Domain	KEGG ID	Pathway Name	Genes
Carbohydrate and energy metabolism	dre00010	Glycolysis / Gluconeogenesis	15
	dre00020	Citrate cycle (TCA cycle)	9
	dre00030	Pentose phosphate pathway	10
	dre00051	Fructose and mannose metabolism	8
	dre00052	Galactose metabolism	5
	dre00500	Starch and sucrose metabolism	6
	dre00520	Amino sugar and nucleotide sugar metabolism	10
	dre00620	Pyruvate metabolism	8
	dre00640	Propanoate metabolism	6
	dre00650	Butanoate metabolism	5
	dre00562	Inositol phosphate metabolism	6
	dre00190	Oxidative phosphorylation	7
	dre00910	Nitrogen metabolism	5
Lipid and nucleotide metabolism	dre00140	Steroid hormone biosynthesis	13
	dre00240	Pyrimidine metabolism	26
	dre00230	Purine metabolism	37
Amino acids metabolism	dre00250	Alanine, aspartate and glutamate metabolism	11
	dre00260	Glycine, serine and threonine metabolism	7
	dre00270	Cysteine and methionine metabolism	6
	dre00280	Valine, leucine and isoleucine degradation	12
	dre00330	Arginine and proline metabolism	6
	dre00310	Lysine degradation	8
	dre00350	Tyrosine metabolism	7
Glycan metabolism	dre00510	N-Glycan biosynthesis	10
Metabolism of cofactors and terpenoids	dre00860	Porphyrin and chlorophyll metabolism	5
	dre00900	Terpenoid backbone biosynthesis	5
Transcription	dre03020	RNA polymerase	6
	dre03022	Basal transcription factors	7
	dre03040	Spliceosome	56
Translation	dre03010	Ribosome	8
	dre00970	Aminoacyl-tRNA biosynthesis	6
Folding, sorting and degradation	dre04120	Ubiquitin mediated proteolysis	22
	dre03050	Proteasome	18
	dre03018	RNA degradation	19

(Continue next page)

(Continued)

Domain	KEGG ID	Pathway Name	Genes
DNA replication and repair	dre03030	DNA replication	27
	dre03410	Base excision repair	13
	dre03420	Nucleotide excision repair	16
	dre03430	Mismatch repair	14
	dre03440	Homologous recombination	14
	dre03450	Non-homologous end-joining	7
Membrane transport	dre02010	ABC transporters	6
Signal transduction	dre04010	MAPK signaling pathway	30
	dre04350	TGF-beta signaling pathway	8
	dre04370	VEGF signaling pathway	6
	dre04630	Jak-STAT signaling pathway	9
	dre04020	Calcium signaling pathway	24
	dre04310	Wnt signaling pathway	8
	dre04330	Notch signaling pathway	7
Signaling molecules and interaction	dre04080	Neuroactive ligand-receptor interaction	9
	dre04060	Cytokine-cytokine receptor interaction	8
	dre04512	ECM-receptor interaction	11
Transport and catabolism	dre04144	Endocytosis	9
	dre04142	Lysosome	15
Cell motility	dre04810	Regulation of actin cytoskeleton	24
Cell growth and death	dre04210	Apoptosis	8
	dre04110	Cell cycle	41
	dre04114	Oocyte meiosis	20
	dre04115	p53 signaling pathway	22
Cellular community	dre04510	Focal adhesion	21
	dre04520	Adherens junction	6
	dre04540	Gap junction	16
Immune system	dre04620	Toll-like receptor signaling pathway	12
	dre04621	NOD-like receptor signaling pathway	6
	dre04622	RIG-I-like receptor signaling pathway	5
Endocrine system	dre04910	Insulin signaling pathway	16
	dre00150	Androgen and estrogen metabolism	5
	dre04920	Adipocytokine signaling pathway	11
	dre04916	Melanogenesis	7
	dre04912	GnRH signaling pathway	10
	dre04914	Progesterone-mediated oocyte maturation	20
Circulatory system	dre04260	Cardiac muscle contraction	10
	dre04270	Vascular smooth muscle contraction	10
Environmnetal adaptation	dre04710	Circadian rhythm	8

Table 3.2: Functional enrichment of *in vivo* genes signatures found to be differentially correlated with *in vivo* residuals at 1% FDR. In red and green are reported KEGG pathways positively and negatively correlated with residuals, respectively while in brown those pathways enriched of genes both positively and negatively correlated with residuals. FDR values of each of the Kegg pathways are provided in the supplementary material.

Biological functions enriched of *in vitro* gene signatures correlated with *in vivo* residuals were just identified for those signatures positively correlated (table 3.3). These functions were associated with nucleotide metabolism and with DNA replication and repair.

Domain	KEGG ID	Pathway Name	Number of genes
Nucleotide metabolism	dre00230	Purine metabolism	5
	dre00240	Pyrimidine metabolism	6
Replication and repair	dre03030	DNA replication	6
	dre03420	Nucleotide excision repair	5

Table 3.3: Functional enrichment of *in vitro* genes signatures found to be differentially correlated with *in vivo* residuals at 1% FDR. In red are reported KEGG pathways positively correlated with residuals. FDR values of each of the Kegg pathways are provided in the supplementary material.

In the initial approach I used, I first identified signatures correlated with lipophilicity and I then removed them from our dataset. However, some of the signatures may also correlate with toxicity as showed in figure 3.9. Residuals analysis on the other hand, have the potential to truly identify signatures which are independent of the lipophilicity.

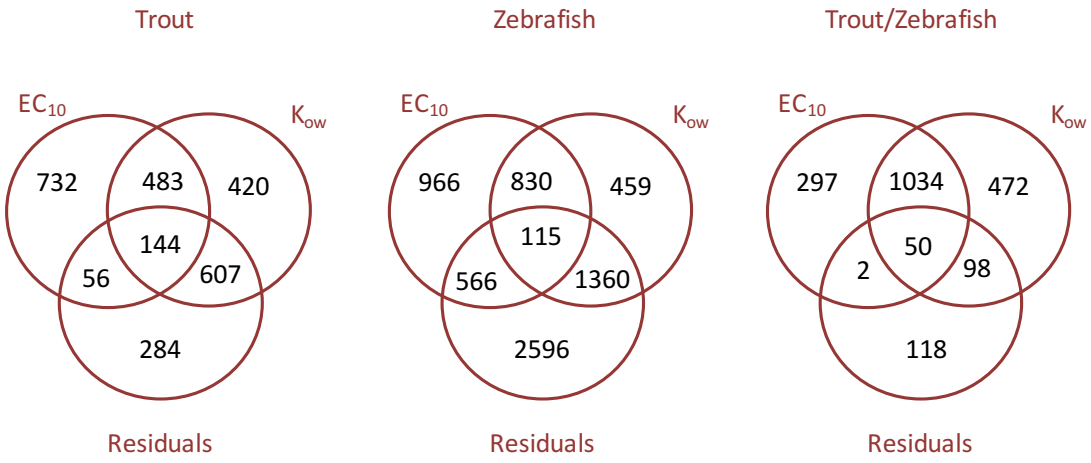


Figure 3.9: Venn diagrams showing gene overlaps between signatures associate with toxicity (EC_{10}) and with lipophilicity (K_{ow}) as previously showed plus the overlap with the signatures associated with the residuals within each biological system.

3.4.4 Identifying good predictors of chemical class compound

Since I have been able to identify gene expression signatures that are potentially representing excess toxicity, I reasoned that these should be good predictors of compound MoA class. I employed KNN classification algorithm to address this hypothesis. Interestingly, I discovered that *in vitro* gene expression signatures linked to either K_{ow}, EC_{10} or excess toxicity are good predictors of compound MoA, whereas the whole genome transcriptional signature tend to be poor predictor for all classes of chemicals except for neurotoxic, which as expected are poorly predicted (fig. 3.10). The *in vivo* system is a better predictor of compound MoA but the selection of excess toxicity genes does not seem to increase dramatically the prediction (fig. 3.11).

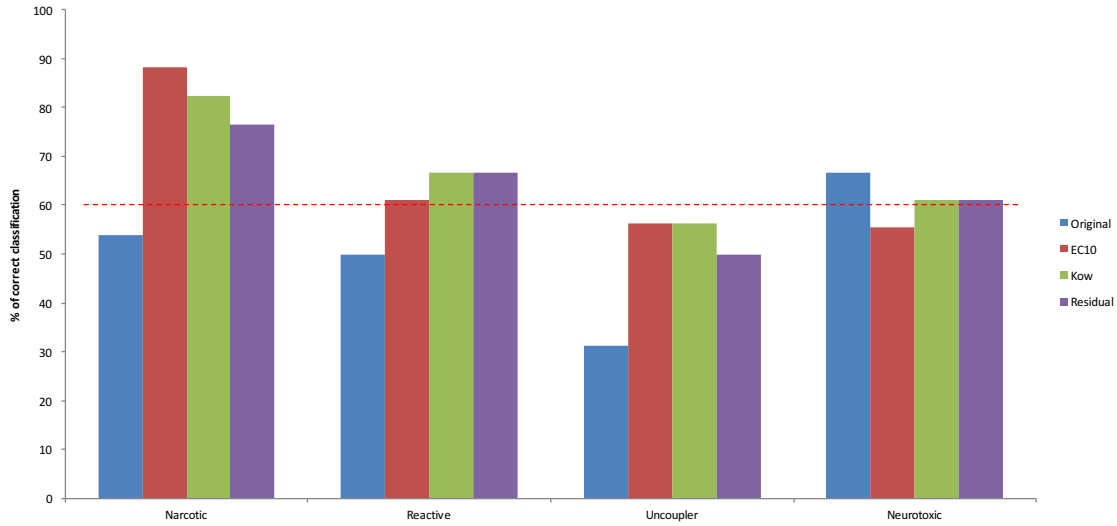


Figure 3.10: Bar-charts report results of the KNN classification for the *in vitro* model when using gene signatures coming from the original dataset and those correlating with *in vitro* toxicity, lipophilicity and residuals. The y axis reports the percentage of chemical classes correctly classified.

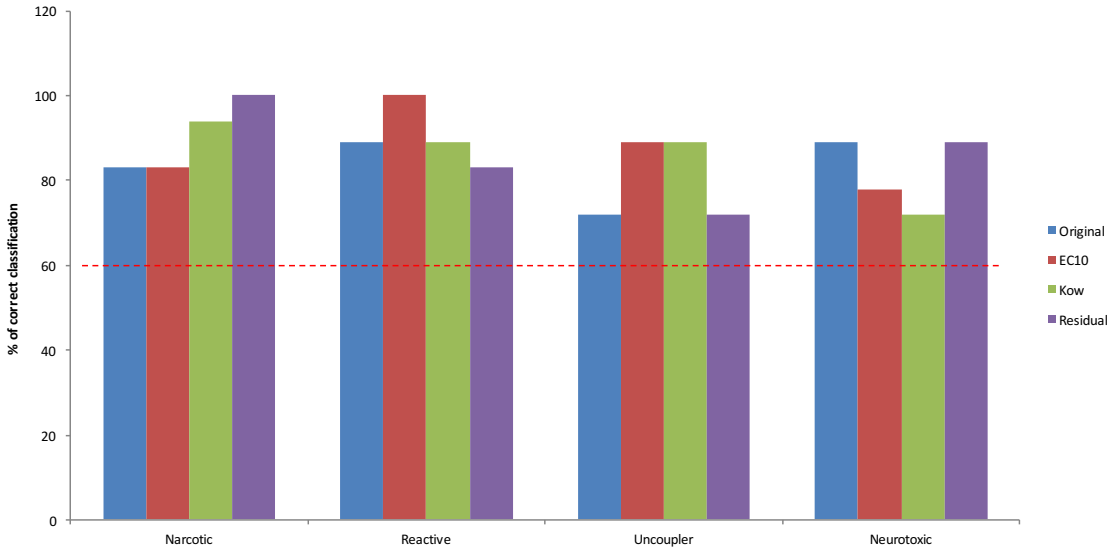


Figure 3.11: Bar-charts report results of the KNN classification for the *in vivo* model when using gene signatures coming from the original dataset and those correlating with *in vivo* toxicity, lipophilicity and residuals. The y axis reports the percentage of chemical classes correctly classified

Most important, I discovered that *in vitro* gene signatures linked to either K_{ow}, EC_{10} and excess toxicity are good predictors of compound MoA in the *in vivo* system (fig. 3.12). More specifically, excess toxicity is better predictor for all classes of chemicals.

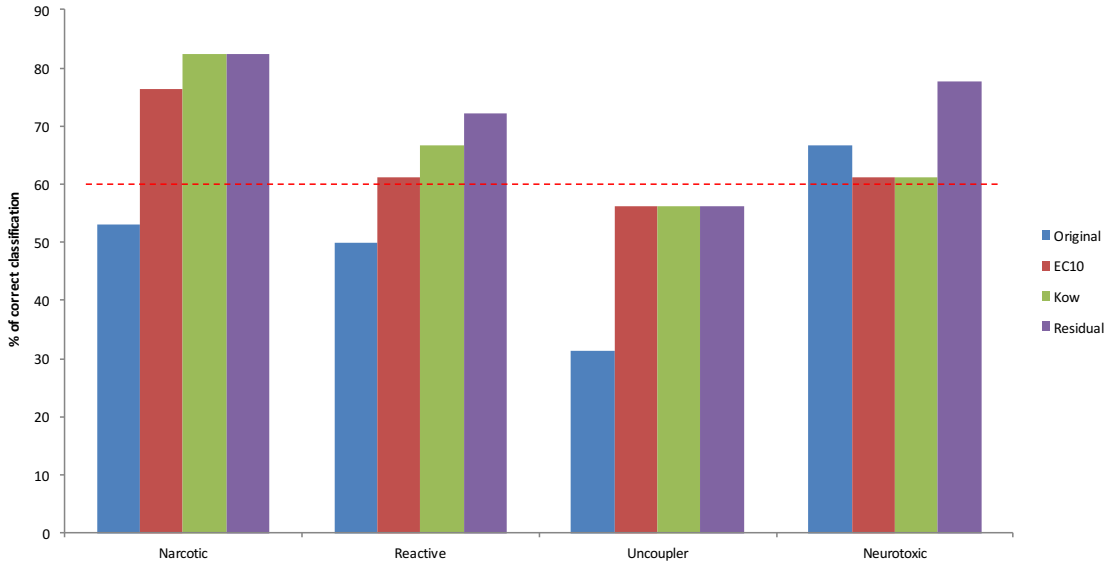


Figure 3.12: Bar-charts report results of the KNN classification for the *in vitro* gene signatures when using gene signatures coming from the original dataset and those correlating with *in vivo* toxicity, lipophilicity and residuals. The y axis reports the percentage of chemical classes correctly classified.

3.5 Discussion

In chapter 2 I did show that the transcriptional response of the RTgill-W1 cell line can be used to predict a chemical MoA. Although with its limitations, this proof of concept study has shown that an *in vitro* system can be used to infer mechanisms of action at the organism level. In this chapter, I have shown that the transcriptional response of the same cell line can be used to predict compound toxicity in the same cell line and, importantly, can be predictive of the development for embryonic abnormalities in Zebrafish embryos.

The functional analysis of the signatures I have discovered is linked to biological terms with high statistical significance. Interestingly, the diagnostic cell live transcriptional signature represents biological pathways that are active during embryo development. This is by far the strongest molecular evidence published so far supporting the view that this *in vitro* system is not just a biosensor but that also has the potential to inform about biological processes specific of embryos.

The approach I have developed has some limitations, mainly due to the relatively small number of chemicals in the dataset. Therefore, the most important issue is whether these findings can be extrapolated to a larger chemical space that include more chemicals within each MoA category and possibly a broader range of specific MoAs. With regards to the MoA categories, in the present study only 4 categories are considered: narcotics, reactives, neurotoxics and uncouplers. These MoA are in a partial agreement with the Verhaar scheme for the classification of chemical compounds[217–219]. In the Verhaar scheme, pollutants can be classified into 4 different categories being: class 1 or inert chemicals, which are nonpolar narcotics and represent baseline toxicity; class 2 or less inert which include polar narcotics; class 3 or reactive chemicals, with a non-selectively enhanced as compared to baseline toxicity; class 4 or specifically-acting compounds, which react with specific receptors and class 5 or unclassified chemicals. Chemicals are classified into any of these classes depending on their chemical structure with the exception of the last category representing compounds acting by specific MoA, which classification has to be based on specific knowledge. With the exception of neurotoxic compounds which represent a broader MoA category and narcotics compounds whose distinction between polar and nonpolar has not been considered, the other two MoA categories in our study follow the Verhaar classification. Another widely adopted MoA classification scheme is represented by the OASIS profiler, developed by the laboratory of mathematical chemistry at the University “Prof. As.

Zlatarov" in Bulgaria. This profiler classifies chemicals into one out of seven categories and its major advantage, compared with the Verhaar schema which is based on chemical structure, is the ability to assign categories based on MoA which provide a clear mechanistic foundation with the ability to improves transparency and acceptability.

By considering more MoA categories as the AChE inhibition and the iono/osmoregulatory/circulatory impairment or by discriminating among the different specific neutotoxic MoA, the robustness my approach could be improved towards a more reliable predictive model. This can be achieved by leveraging the knowledge of public databases as the MOAtox developed by Barron et al.[220].

The second issue is a statistical one. The regression model I have used to define the residuals is effectively meant to represent a traditional basal toxicity model. The fact that I have built the model by using a small number of chemicals, of which only a few are bona fide narcotics raise the issue that the regression line may not represent the true relationship between lipophilicity and toxicity. A larger number of chemicals is necessary to develop an accurate model. Moreover, the model I have developed does not discriminate between polar and non-polar narcotics. These are both interacting with biological membranes in a lipophilicity dependant manner but they are thought to do so with different chemical physical mechanisms, resulting in significantly different biological effects [221,222]. Ideally, with a sufficiently larger number of chemicals belonging to both classes I may be in the position to develop this proof of concept study to a fully applicable predictive model.

Having identified transcriptional responses associated with excess toxicity and baseline toxicity I addressed the question of whether *in vitro* gene signatures were able to inform about chemical MoA in an *in vivo* system. Results revealed that transcriptional signatures linked to toxicity are

able to discriminate between chemical MoAs and that this is also true when extrapolating from an *in vitro* system to whole animal. This proof of concept study offers a valuable approach that, once further developed fulfilling the limitations previously highlighted, has the potential to be used for IVIVE and may help reducing the use of animal testing as well as providing fast, cost-effective and high-throughput methods for conducting chemical risk assessment.

Chapter 4: *In silico* computational transcriptomics reveals novel endocrine disruptors in Largemouth bass (*Micropterus salmoides*)

I have developed data analysis strategies and performed all the in silico analysis with the exception of the microarray annotation which was achieved by John Herbert. Laboratory validation have been performed by our collaborating lab

4.1 Abstract

In recent years, decreases in fish populations have been attributed, in part, to the effect of environmental chemicals on ovarian development. To understand the underlying molecular events, I developed a dynamic model of ovary development linking gene transcription to key physiological endpoints, such as gonadosomatic index (GSI), plasma levels of estradiol (E2) and vitellogenin (VTG), in largemouth bass (*Micropterus salmoides*). I was able to identify specific clusters of genes, which are affected at different stages of ovarian development. A sub-network was identified that closely linked gene expression and physiological endpoints and by interrogating the Comparative Toxicogenomic Database (CTD), Quercetin and Tretinoin (ATRA) were identified as two potential candidates that may perturb this system. Predictions were validated by investigation of reproductive associated transcripts using qPCR in ovary and in the liver of both male and female largemouth bass treated after a single injection of Quercetin and Tretinoin (10 and 100 µg/Kg). Both compounds were found to significantly alter the expression of some of these genes. These findings support the use of omics and online repositories for

identification of novel, yet untested, compounds. This is the first study of a dynamic model that links gene expression patterns across stages of ovarian development.

4.2 Introduction

The increasing amount of pollutants released into the environment is a major issue for the development of a sustainable economy. Growing numbers of anthropogenic pollutants affect freshwater and marine environments with profound impact on species of economic importance. This leads to an increase in the need for additional ecosystem maintenance to secure a constant, minimally burdened, food supply[223,224]. Moreover, due to their position in the food chain, higher level organisms (including humans) can be negatively affected through biomagnification of these toxic substances[225,226]. Endocrine disruptors (EDs), in particular, have the ability to associate with adverse effects, such as reproduction[227,228]. EDs are exogenous agents that interfere with the synthesis, transport, binding action, metabolism, secretion or elimination of natural endogenous hormones responsible for reproduction, homeostasis and developmental processes; the disturbance of which may lead to adverse outcomes[229–231]. These compounds are commonly found in daily use products such as detergents, cosmetics, processed food and products containing flame retardants[232]. Many taxa exhibit reproductive and developmental abnormalities following exposure to EDs, including humans, reptiles, mammals, amphibians, birds, fish, and invertebrate organisms[233]. For example, one of the most common abnormalities in animals in the aquatic environment that is caused by exposure to EDs is intersex, defined as males that have both sperm and oocytes in their testis[234]. EDs elicit tissue-specific responses[235] and have been shown to be toxic even at very low concentrations[236]. Moreover, time of exposure has been shown to be a crucial factor that determines the potency of EDs[229].

Largemouth bass (LMB) (*Micropterus salmoides*) is an important economic fish species widely distributed throughout the USA. LMB are popular as a sports fish and they represent a keystone species in freshwater ecosystems due to their trophic position as an apex predator. LMB reproduction is typically synchronous as they develop their gonads over a spawning season,

which is controlled by both environmental and physiological factors (e.g. temperature, photoperiod and endogenous hormonal triggers[237]). Oocyte growth can be divided into two main stages of development, classified as primary growth or pre-vitellogenic and secondary growth or vitellogenesis[238,239]. These developmental stages can be further divided into discrete reproductive stages depending on ovarian morphology and oocyte maturation as defined by Martyniuk et al.[8,240]. The first stage of development is named perinuclear stage (PN) and it is characterized by the formation of the follicle which is made up of granulosa cells surrounding the oocyte, a basal lamina and the theca cells. Moreover, meiosis is arrested at the diplotene stage of prophase I and an intensive transcriptional activity is taking place[241]. The name of the stage is given by the production of multiple nucleoli following nucleolar amplification, which becomes oriented in a perinuclear position[242]. The oocyte keeps increasing its size entering in the cortical alveoli stage (CA). This stage is characterized by the formation of large vesicles organized in a multi-layered structure at the oocyte periphery that keeps growing and fuses together[243]. The next stage is characterized by the oocyte uptake of nutritional resources as the egg yolk protein vitellogenin (Vtg) and is named vitellogenesis stage (VTG). The Vtg is synthesized by hepatocytes, carried to the oocyte by the bloodstream and incorporated through a receptor-mediated process[239,241]. Once its uptake is completed the oocyte enters the maturation stage (OM) where the nucleus, also called germinal vesicle, starts to migrate to the animal pole of the oocyte. At the ovulation stage (OV) the oocyte emerges from the follicle becoming an egg. Quantifying the molecular events underlying ovary development dynamics in response to pollutants facilitates improved understanding of the mechanisms and hence improves our ability to design and manufacture safer products.

In this study, I used transcriptome profiling data coupled with computational approaches to model the effects of chemicals in the LMB ovary. Using omics datasets, I first constructed a

dynamic model representing the development of healthy ovaries from unexposed fish. By mapping the responses of a transcriptome from LMB collected from a polluted site, I was then able to identify modules (clusters of genes), which are perturbed at different stages of ovarian development. By utilizing the Comparative Toxicogenomic Database (CTD)[244], a robust database providing information about chemical interaction with genes, proteins and disease, I identified Tretinoin and Quercetin as potential chemicals that were associated to the observed molecular response in LMB ovary following reproductive disruption. Both compounds were subsequently tested for their potential as reproductive endocrine disruptors using exposure experiments in the laboratory with LMB. This study demonstrates that by utilizing computational approaches and online knowledge bases to understand the underlying molecular response of organisms, it is possible to identify putative chemical candidates that may impact reproductive health. This approach is highly relevant for classifying chemicals prior to conducting risk assessments, and I propose that this is a viable approach for chemical prioritization, reducing animal numbers, and developing safer chemicals in the public domain.

4.3 Materials and methods

4.3.1 Experimental design

For a detailed description of the experimental design and the data processing workflow see[8]. Briefly, wild largemouth bass (LMB) were collected from the St. Johns River in Florida from October 2005 to April 2007 at Welaka (29.48° N, 81.67° W) located approximately 20 miles south of Palatka, FL. This area is considered to be relatively free from the influence of industrial effluent and agricultural runoff[245]. Water temperature varies dramatically over the year and is reported by a month to month basis by Martyniuk et al[240]. At the sampling time, ovaries were dissected, and fish were categorized based on histology into 7 different stages of reproductive

development: perinucleolar (PN), cortical alveoli (CA), early vitellogenin (eVTG), late vitellogenin (lVTG), early ovarian maturation (eOM), late ovarian maturation (lOM) and ovulation (OV). Physiological endpoints as Plasma vitellogenin (VTG) levels, 17β-estradiol (E2) levels and gonadosomatic index (GSI) were also measured. Ovaries were then processed for gene expression profiling, with four biological replicates for each ovarian stage , using a custom LMB microarray platform[246]. These data were used to characterize the molecular events underlying oocyte maturation in the LMB ovary, identifying potential biomarkers of atresia[8]. In a second study, a mesocosm experiment was set up by placing wild adult LMB (+ 3 years) at different stages of reproduction (late CA or eVTG), sampled from DeLeon Springs in Florida, in the Apopka ponds in October (collected from the ponds for samples in January) and in January (collected from the ponds for samples in April). LMB were in the ponds for 4 months. This site is well-known to be impacted by anthropogenic sources[62]. The contaminant load consisted of high levels of organochlorine pesticides (DDT, dieldrin, toxaphene, and others)[62]. These organochlorine pesticides, are known to disrupt LMB reproduction due to their estrogenic and anti-androgenic properties[247–249]. Ovaries were then collected and processed for gene expression profiling using four biological replicates[62].

4.3.2 Annotation

The LMB microarray (Agilent ID: GPL 13229; Santa Clara, CA, USA) consists of 15,950 sequences. To improve on the previous multi-species annotation, the microarray was re-annotated using the most recent annotated genomes available. Two approaches, leveraging the power of the NCBI blast tool, were used: 1) Using blastn (search a nucleotide database using a nucleotide query), we aligned all of the array-design sequences against the most closely related and most completely annotated genome, which was the three-spined stickleback (*Gasterosteus aculeatus*) and then, using the same method, the sequences were aligned to RefSeq zebrafish (*Danio rerio*)

cDNA; 2) The array-design sequences were compared against the three-spined stickleback protein with blastx (search a protein database using a translated nucleotide query) and then with blastp (search a protein database using a protein query) against the RefSeq zebrafish protein. In both approaches, the e-value threshold was set at $1e^{-6}$. The two approaches led to the annotation of 6,373 and 5,522 sequences, respectively. A total of 7,772 genes (4,926 unique genes) were successfully associated to an official gene symbol of which 4,338 were associated to the zebrafish genome (98%) providing improved coverage over the previous annotation (1,031 genes, 15%).

4.3.3 Differential gene expression and clustering

Differentially expressed genes were identified using the Significance Analysis of Microarray (SAM)[117] within the statistical environment R ("samr" package). To identify expression changes across the different stages of development, I applied a time-course SAM. To identify genes whose expression was differentially expressed in at least one stage, a multiclass SAM, which does not consider the time and is not constrained by a specific response function, was applied. Significant genes were defined by a threshold of 10% FDR (False Discovery Rate) to maximize the number of differentially expressed genes. To visualize the relationship between samples the differentially expressed genes were used as input to a principal component analysis using the "prcomp" package within the statistical environment R. In order to simplify the complexity of the dataset I set to identify clusters of genes whose expression was correlated across the different stages of development. I employed SOTA (Self-Organizing Tree Algorithm)[110] using a Pearson correlation distance measure, an unsupervised neural network with a binary tree topology that can be easily scaled to large datasets. Each cluster was functionally annotated using the web-based software tool DAVID[118]. Biological gene ontology and KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways with an FDR < 5% were considered.

4.3.4 Dynamic modelling and chemical mapping

Modules identified by the SOTA approach using the two different SAM methods were then merged into a single dataset. I chose to represent each module with a 4th degree polynomial interpolation of its genes. The approach interpolated the 7 ovarian stages to 100 pseudo-timepoints. The resulting dataset and the physiological measurements (VTG, GSI and E2) were then used as input to a TimeDelay-ARACNE (TDA) algorithm[130]. This method extracts dependencies between two genes by incrementally delaying the expression profile of one gene against another; this results in a comparison where the timecourse of gene$_1$ "t_0 t_1 t_2 … $t_{(n-1)}$" is then compared to a delayed timecourse of gene$_2$ "t_1 t_2 t_3 … t_n". The amount by which one profile is delayed with respect to another is defined as the time-delay. By testing multiple time-delays and identifying the highest dependency between two genes, a directionality can be inferred and represented graphically in a network format. TDA has been successfully compared to dynamic Bayesian networks and ordinary differential equations and it has been shown to have a good accuracy for network reconstruction[130]. To identify whether any of the identified clusters were up and down regulated as a result of pollution, I utilized a gene set enrichment analysis (GSEA)[121]. In addition to the standard GSEA procedure, a pre-ranked list of genes can be provided to the algorithm. To define the pre-ranked gene list, I conducted a "Two-class unpaired" SAM analysis between the polluted and clean sites and recorded the d-statistic for each gene. The d-statistics were then provided to the GSEA approach (to rank each gene). I then imported the previously identified clusters as gene-sets. GSEA then tested whether any of the identified clusters were generally up (positive d-statistic) or down (negative d-statistic) regulated in respect to pollution. Clusters identified as enriched were recorded and visually represented on the network view. Functional enrichment of the sub-network of interest was achieved using DAVID with a 1% FDR threshold applied.

4.3.5 CTD enrichment

To identify chemicals with the ability to affect the sub-network of interest, I utilised the curated data collection in the Comparative Toxicogenomics Database (CTD)[244]. The CTD provides manually-curated information about interactions between chemical and gene/protein as well as chemical-disease relationships to aid in the development of hypotheses about mechanisms underlying chemical and environmentally influenced disease. To identify which potential compounds might be involved in generating the transcriptome response that I observed, I first downloaded the CTD (Download-date: 14 January 2016) and identified the species with the broadest number of chemical interactions (*Homo sapiens*). As the zebrafish has been extensively used as a model system for human disease, human gene ortholog information is better defined than other species covered in the CTD. I selected the human subset of data and although this is a needed compromise, it provides me with the opportunity to explore potentially conserved mechanisms. I first converted my zebrafish genes to human genes using ZebrafishMine[250] (3,977 genes).To identify which chemical may interact with the sub-network of interest I calculated the EASE score (Expression Analysis Systematic Explorer; as defined in the DAVID web-based tool[118]) against the CTD database filtered for those compounds having associated more than 5 genes. The thresholding is necessary to reduce the potential of identifying spurious compound hits where a low gene-set size results in a significant p-value. Retrieved p-values were adjusted using a Benjamini and Hochberg correction (analogous to a FDR). This resulted in a list of compounds that have more genes in common than expected by random chance. I ordered the compounds by their respective p-value (10% FDR threshold) to identify the top identified toxicants.

4.3.6 Prediction validation

Prediction validation was carried out by our collaborating lab. Twenty reproductive largemouth bass (10 males and 10 females for each exposure group) were injected intraperitoneally with 10

or 100 µg/Kg of Quercetin or Tretinoin dissolved in DMSO. The controls were injected with the carrier solution DMSO. Following a forty-eight hours exposure, the fish were anesthetized, weighed, bled, and dissected. For real-time PCR, sample sizes for female ovary and liver were as follows: Control (n=6), Quer 10 (n=6), Quer 100 (n=7), Tret 10 (n=6), and Tret 100 (n=6). Sample sizes for male liver were as follows: Control (n=6), Quer 10 (n=6), Quer 100 (n=6), Tret 10 (n=6), and Tret 100 (n=6). Gonad and liver tissues were collected and flash frozen in liquid nitrogen for RNA purification and extraction. RNA was assessed for quality using the Agilent 2100 Bioanalyzer and all samples showed a RIN > 7.0. RNA was subjected to a DNAse treatment using DNAse Turbo as per manufacturer's protocol (Ambion). The cDNA synthesis was performed using 1 µg total RNA (using the iScript BioRad protocol). Primer sets for target genes were collected from the literature for LMB. The genes investigated in this study included androgen receptor (*ar*), estrogen receptor alpha, betaa and betab (*erα*, *erβa*, *erβb*), aromatase (*cyp19a*), steroidogenic acute regulatory protein (*star*), vitellogenin (*vtg*) and vitellogenin receptor (*vtgr*). We chose these reproductive transcripts because (1) the computational analysis was done in reproductive tissues (liver and ovary) over a breeding season and these transcripts are sensitive to maturation, (2) we have observed that these genes are perturbed by chemicals in LMB and (3) these gene assays are widely used and reliable in our laboratory. 18S rRNA was used to normalize gene targets. Real-time PCR was performed using the CFX Connect™ Real-Time PCR Detection System (BioRad) with SSoFast™ EvaGreen® Supermix (BioRad, Hercules, CA, USA), 200 nM of each forward and reverse primer, and 3.33 µL of diluted cDNA. The two-step thermal cycling parameters were as follows: initial 1-cycle Taq activation at 95 °C for 30 s, followed by 95 °C for 5 s, and primer annealing for 5 s. After 40 cycles, a dissociation curve was generated, starting at 65.0 and ending at 95.0°C, with increments of 0.5 °C every 5 s. Normalized gene expression was extracted using CFX Manager™ software with the relative ΔΔCq method (baseline subtracted). All primers used

in the qPCR analysis amplified one product, indicated by a single melt curve. Details about the primers are provided in Table 4.1. Significant differences between group means were analysed using analysis of variance (ANOVA) followed by Dunn's post-hoc test. A value of $p < 0.05$ was used to indicate significant differences.

Gene	Forward (5'-3')	Reverse (5'-3')	T (°C)	R2	Efficiency (%)
[251]**rps18**	CGG CTA CCA CAT CCA AGG AA	CCT GTA TTG TTA TTT TTC GTC ACT ACC T	58	0.996	88.1
[62]**rps23**	CAG AAA TGG CAC GAT AAG CA	GAC CTT TAC GCC CAA ATC C	58	0.979	104.3
[62]**ef1α**	GGA CAA ACT GAA GGC AGA GC	ACA CCA GCA GCA ACA ATC AG	58	0.992	92.1
[252]**ar**	CAC CAC AGA GAA TGT GCC TGA	CAG GTG AGT GCG CCG TAA	58	0.997	93.2
[253]**esr1**	CGA CGT GCT GGA ACC AAT GAC AGA G	ACC TCC GGT CAC TGA TGA TTT TCC TCC T	58	0.999	103.9
[253]**esr2a**	CCG ACA CCG CCG TGG TGG ACT C	AGC GGG GCA AGG GGA GCC TCA A	58	0.991	101.4
[253]**esr2b**	GTG ACC CGT CTG TCC ACA CA	TCT GGG GTC AGT GCA GGA GA	58	0.985	98.8
[252]**cyp19**	TGG ATC AAG TGG ATG TCC TCA GT	CCA GGA AGA GTC TGT TGG AGA TG	58	0.999	98.6
[252]**star**	ACC CCT CTG CTC AGG CAT TT	GGG CTC CAC CTG CTT CTT G	58	0.980	99.8

[61]**vtg receptor**	ATG CAG ACT TTA ACC AGC AGA TGA T	TGT CCA GCA AAG TGC TAT AAA TGG	58	0.998	96.1
[254]**vtg**	CAG AGT GAG ATG GGC GTT G	CAG GCG TTT GTT GGG TGT	58	0.994	104.6

Table 4.1: The table provide the list of primers, along with their sequences and parameters, employed for the experimental validation by qPCR of tretinoin and Quercetin effects on LMB ovary development.

4.3.7 Chemical-Set Enrichment Analysis

To test whether my approach preferentially identifies endocrine disruptors as by random chance I first downloaded a defined list of endocrine disruptors from the CTD. Secondly, I extracted the estimates from the fisher exact test performed during identification of chemicals. I then used these two datasets as input to a standard Pre-Ranked GSEA.

4.3.8 Analysis strategy

The overarching objective of my analysis was to identify chemicals that were most likely to disrupt ovarian development in largemouth bass. This was achieved by developing a dynamic network model, linking changes in the transcriptional state of different stages of normal LMB ovaries and measured key physiological endpoints (GSI, VTG and E2). The first step in developing this model was to reduce the overall complexity of the gene expression profiles. I first identified differentially expressed genes during ovary development (Fig 4.1, step 1a-1b) and then clustered the transcripts, via self-organizing trees, based on similarity of gene expression profiles (Fig. 4.1, step 1c). This reduced the total number of variables in the dynamic network model and provided means for better biological interpretation. The linkages between the key physiological indicators and the clusters of gene expression profiles were inferred using a time-delay mutual information algorithm (Fig. 4.1, step 2) and the response of the transcriptome to the polluted environment

was then mapped via gene set enrichment analysis onto the network (Fig. 4.1, step 3). By identifying which gene clusters are 1) directly connected to key physiological endpoints and 2) significantly enriched in the polluted site, I identified a sub-network of interest. Finally, by interrogating the CTD, and matching gene expression profiles to the clusters perturbed in an environment under chemical stress, I was able to identify chemical candidates that were predicted to interfere with ovary development (Fig. 4.1, step 4). Resulting chemical stressors were then ordered and the likely candidates tested on their ability to perturb ovary related systems (Fig. 4.1, step 5).

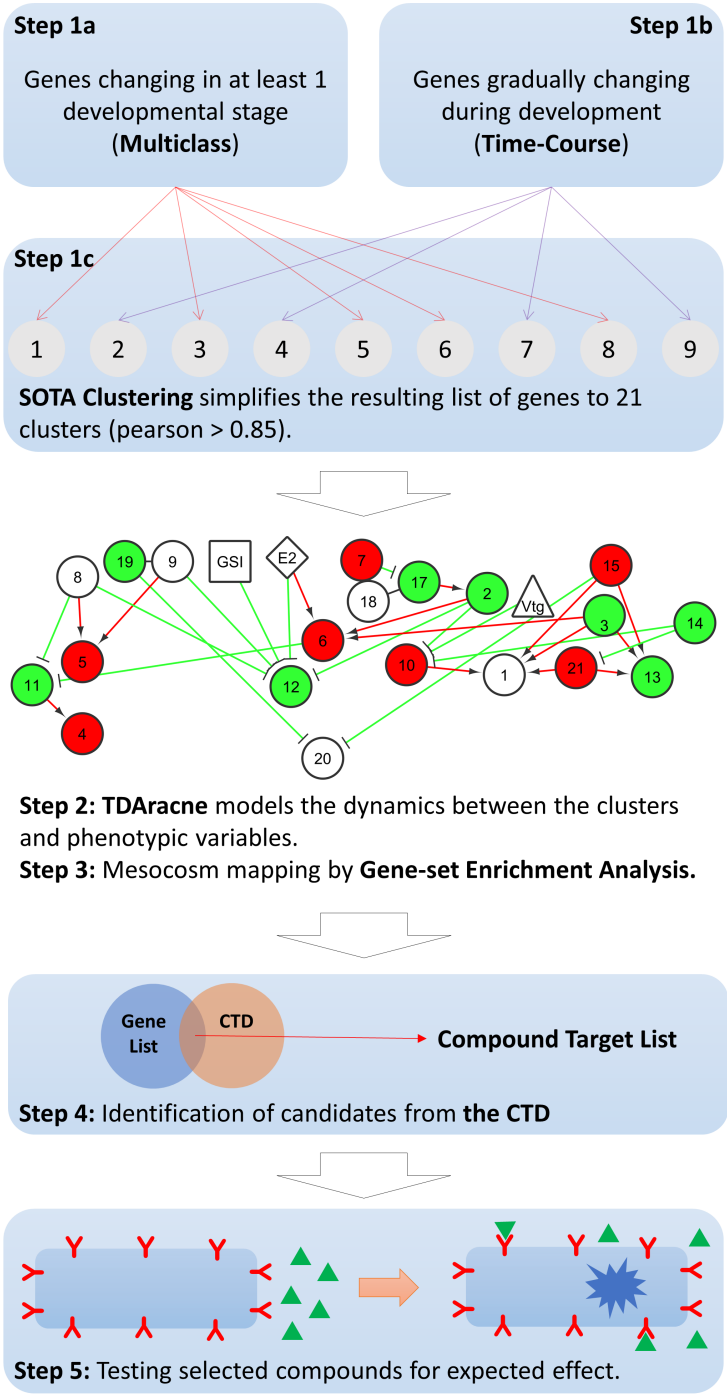


Figure 4.1: Schematic representation of the analysis pipeline. The data acquired by expression profiling underwent analysis to identify differentially expressed genes (step 1 a and b), in addition to clustering genes sharing similar expression profiles (step 1c). The linkage between gene clusters and key physiological indicators such as VTG, GSI and E2 across the different stages of development were inferred using a mutual information-based algorithm and response of the transcriptome in fish inhabiting a polluted environment were mapped via GSEA (step 2 and 3); the CTD database was interrogated to identify potential chemical candidates with the ability to affect endocrine functions driving ovary development (step 4); finally, candidate chemicals were then tested experimentally (step 5).

4.4 Results

4.4.1 Differential gene expression analysis identifies biological functions involved in ovary development

In order to understand what biological processes are linked to ovary development I first set to identify genes differentially expressed across the seven stages of development. This was achieved by two criteria: 1) genes which changed expression over the different stages of ovary development (One class SAM time-course), and 2) genes which were significantly different in at least one developmental stage (multiclass). I identified 3057 genes associated with oocyte development and 5047 whose expression was significantly different in at least one stage at 10% FDR (Fig. 4.2). The two gene-lists were then combined to give the best possible understanding of the molecular response to normal ovary development.

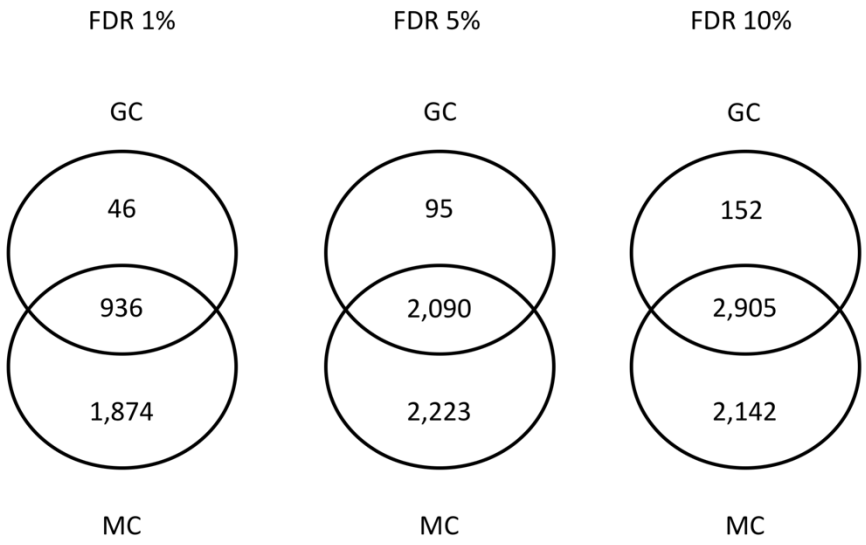


Figure 4.2: Venn diagrams showing number of differentially expressed genes obtained using either of the two methods (MC and GC) and their overlap across different FDR thresholds.

In order to visually represent the dynamics of change in the transcriptional state of healthy ovaries, the resulting list of 5,199 genes was used as input to a principal component analysis (PCA) (Fig. 4.3). I could project the state of ovary samples in two dimensions while retaining 55% of variance. This visual representation was consistent with a high reproducibility of replicated samples as well as an expected progression from early stages such as perinuclear (PN) to ovary maturation (OV).

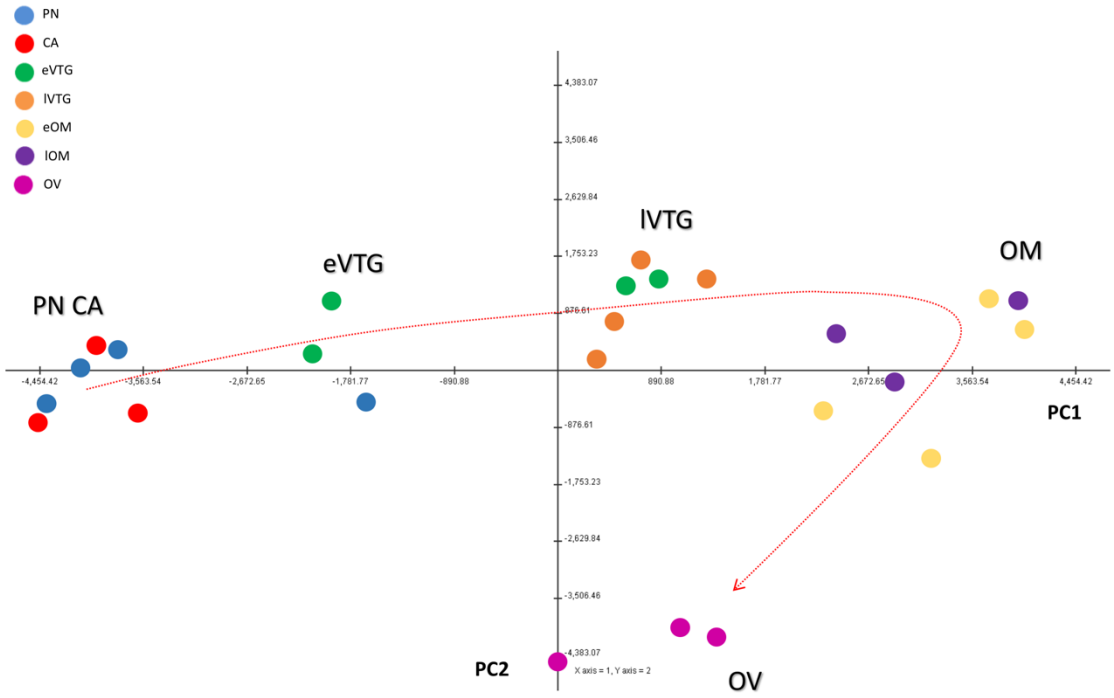


Figure 4.3: Principal component analysis (PCA) shows a clear progression of ovary development from the first to the terminal stage of ovarian development. Stages are defined as PN (perinuclear), CA (cortical alveoli), eVTG (early vitellogenesis), lVTG (late vitellogenesis), eOM (early ovarian maturation), lOM (late ovarian maturation) and OV (ovulation).

For each of the gene-sets, a self-organising tree algorithm (SOTA) was applied and for each cluster, functional annotation was retrieved. This approach yielded 12 and 9 clusters for the time-course and multiclass gene-sets, respectively. Heat-maps of the clusters showed a variation in the responses that were identified by either of the differential gene expression approaches (Fig.

4.4). The identified functions represented were mainly linked to transcriptional and translational activity, energy metabolism and cell growth activity, for example as mTOR signaling pathway or cell cycle, respectively. Interestingly, mTOR is regulating both cell cycle and energy metabolism by controlling the selecting translation of growth factor induced genes[255].

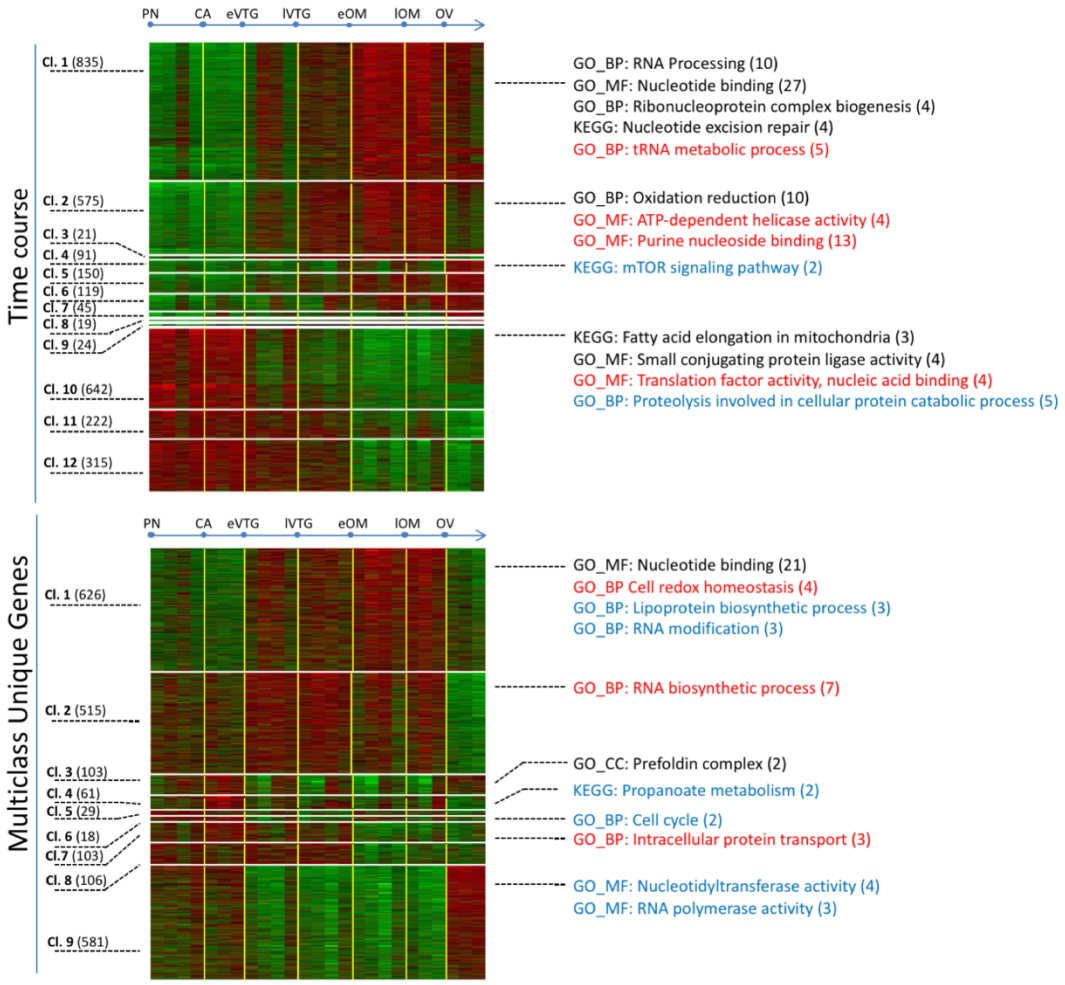


Figure 4.4: Expression profiles of transcripts identified by two different differential gene expression approaches. Red and green shows up and down regulation, respectively. Functional annotation is reported with black, red and blue terms representing gene ontology terms at 5%, 10% and 20% FDR, respectively. FDR values of each of the GO terms are provided in the supplementary material.

4.4.2 A dynamic model of ovary development

To develop a dynamic model of ovary development, the identified clusters and key physiological endpoints were used as a input into TimeDelay-ARACNE (TDA). This resulted in a directed network where edges, linkages between nodes (clusters of genes and endpoints), indicate a positive or negative direction of effect based on their temporal profiles. The resulting network represented 20 modules linked through 30 edges that are coloured red and green to distinguish positive and negative effects (Fig. 4.5).

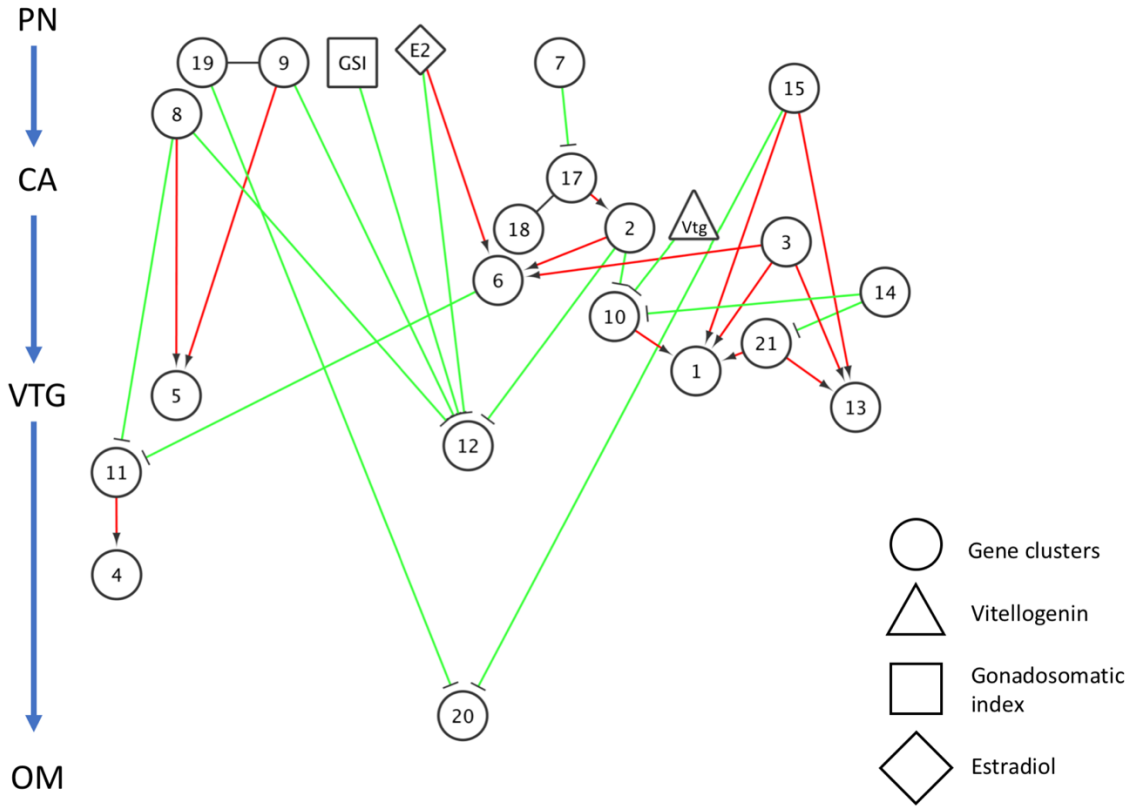


Figure 4.5: The figure reports the model of ovary development inferred employing TD-ARACNE algorithm (details in the materials and methods section of the paper). The model links gene clusters with physiological measurements (Vtg, E2 and GSI) across the different stage of ovary development in a time-dependent manner. Ability of a cluster at an early stage of development to enhance or inhibit the expression of a cluster at a later stage is displayed in red and green, respectively.

Noteworthy was that several clusters were directly associated to the key physiological endpoints (Table 4.2). Cluster 12 was linked to GSI and E2 and contained genes that had representative functions including transduction (ribosome and spliceosome), nucleotide metabolism (purine and pyrimidine metabolism) and energy metabolism (oxidative phosphorylation). Cluster 6 linked to E2 and contained genes that had representative functions that were associated with glycolysis/gluconeogenesis pathway. Functions associated with embryo development (eye development), DNA replication (DNA replication, nucleotide excision repair and mismatch repair), translational activity (ribosome) and energy metabolism (oxidative phosphorylation) were found to be represented in cluster 10 which was linked to VTG.

Clusters	KEGG Pathway	FDR
1	DNA replication	$6.5X10^{-7}$
	Lysosome	$6.9X10^{-7}$
	Cell cycle	$2.6X10^{-6}$
	Apoptosis	$5.4X10^{-6}$
	Glycolisis/Gluconeogenesis	$8.7X10^{-6}$
2	Sphingolipid metabolism	$1.4X10^{-3}$
	Glutathione metabolism	$1.4X10^{-3}$
	MAPK signaling pathway	$1.5X10^{-3}$
	Focal adhesion	$1.6X10^{-3}$
	Apoptosis	$1.8X10^{-3}$
4	Endocytosis	$2.6X10^{-4}$
	Basal transcription factors	$1.7X10^{-3}$
5	Cell cycle	0.02
6	Glycolysis/Gluconeogenesis	$9.5X10^{-3}$
10	Oxidative phosphorylation	$3.9X10^{-9}$
	Nucleotide excision repair	$2.4X10^{-4}$
	Gap junctions	$3.2X10^{-4}$
	Oocyte meiosis	$7.4X10^{-3}$
	Progesterone-mediated oocyte maturation	0.02
11	Lysosome	$5.7X10^{-5}$
12	Spliceosome	$7.5X10^{-5}$
	Ribosome	$6.1X10^{-4}$
13	Purine metabolism	$6.2X10^{-8}$
	Cell cycle	$1.2X10^{-6}$
	DNA replication	$1.9X10^{-4}$
14	Purine metabolism	$1.9X10^{-5}$
	Oocyte meiosis	$7.4X10^{-5}$
	DNA replication	$1.5X10^{-4}$
	Lysosome	$7.6X10^{-4}$
	Cell cycle	$1.5X10^{-3}$
15	Spliceosome	$1.6X10^{-3}$
	Ribosome	0.02
17	Ribosome	0.02
19	Ribosome	$4.7X10^{-4}$
21	Ribosome	$1.4X10^{-8}$
	Tight junction	$5.7X10^{-6}$
	Regulation of actin cytoskeleton	$1.9X10^{-5}$

Table 4.2: The table shows the functional enrichment for each of the gene clusters belonging to the dynamical model of ovary development. Functional enrichment has been achieved using DAVID webservice and retrieving KEGG pathways. A 5% FDR threshold was used. Number of genes within each of the Kegg pathways are provided in the supplementary material.

4.4.3 Mapping the effects of pollutant exposure on the model for healthy ovary development

Having developed the network representing normal ovary development, I hypothesized that this would provide a platform for quantifying the effects of pollution. To test this hypothesis, I utilized an additional dataset, which was developed as part of the same initial study, which compared ovary transcriptomes of LMB collected from a heavily polluted and a pristine site. All fish in this experiment were histologically classified as late vitellogenesis stage and so the molecular differences were expected to affect clusters around the level of eVTG-lVTG. By using gene set enrichment analysis, I was able to determine whether any of the clusters were either up or down-regulated as a result of the polluted environment. Interestingly, the effect of pollution extended well beyond the expected stage of development, showing significant effects even in clusters placed in PN and CA stages suggesting a regression, developmental inhibition or a stronger overlap of different ovary development stages appearing simultaneously during the process (Fig. 4.6).

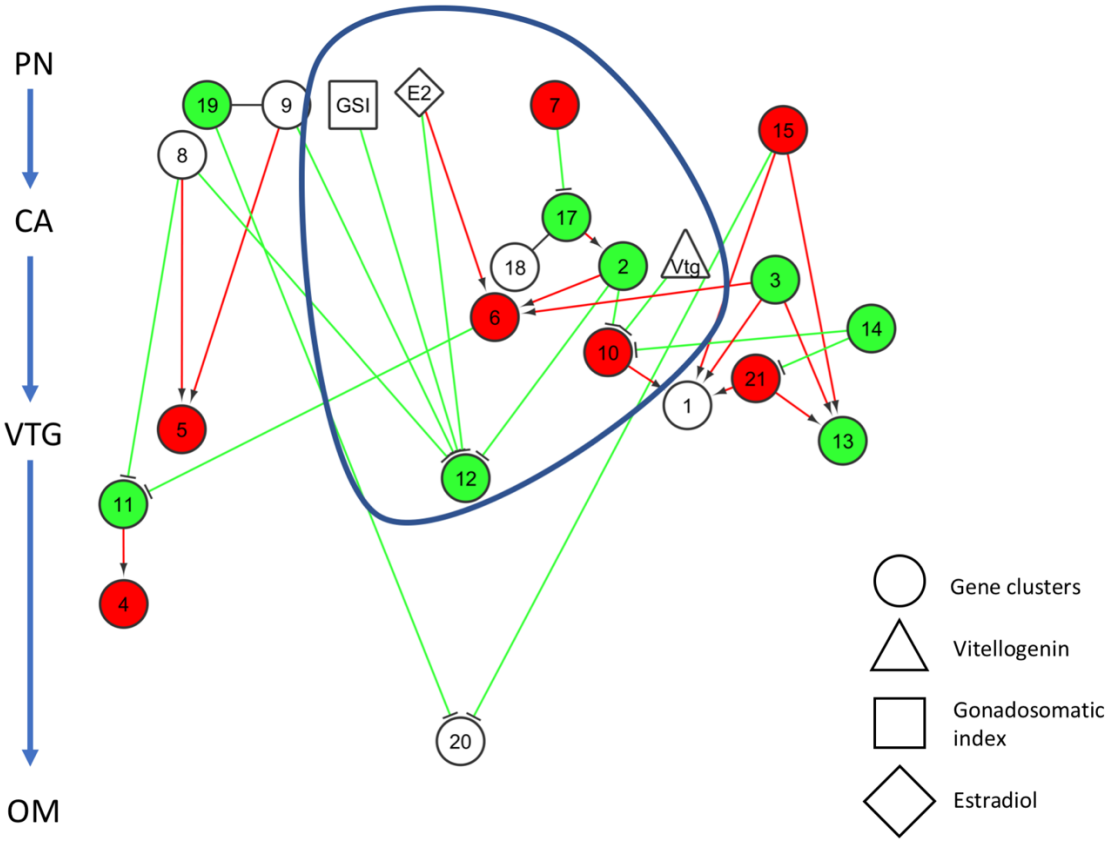


Figure 4.6: The transcriptome responses in the ovary due to a polluted site were mapped onto the developed dynamic network and a sub-network of interest, which included all three physiological measurements, was identified. Clusters positively or negatively enriched with pollution-related genes are displayed in red and green, respectively.

This led to the identification of a sub-network, which connected to physiological endpoints and clusters 2, 6, 7, 10, 12 and 17. Functional characterization of this sub-network revealed evidence of E2-dependent functions, which are well known to drive ovary development (Table 4.3).

KEGG pathway	Gene count	Bonferroni
Spliceosome	17	$3.69X10^{-13}$
Oxidative phosphorylation	13	$3.33X10^{-10}$
Nucleotide excision repair	10	$8.79X10^{-9}$
Ubiquitin mediated proteolysis	10	$4.54X10^{-7}$
Focal adhesion	10	$5.87X10^{-7}$
Ribosome	11	$1.83X10^{-6}$
MAPK signaling pathway	9	$5.74X10^{-6}$
Basal transcription factors	7	$1.05X10^{-5}$
Pyrimidine metabolism	9	$1.42X10^{-5}$
Calcium signaling pathway	7	$2.38X10^{-5}$
P53 signaling pathway	7	$6.68X10^{-5}$
Purine metabolism	8	$7.31X10^{-4}$
Inositol phosphate metabolism	5	0.001
Gap junction	5	0.001
TGF-beta signaling pathway	6	0.001
Endocytosis	7	0.002
Lysosome	7	0.002
Regulation of actin cytoskeleton	7	0.003
Cysteine and methionine metabolism	5	0.004
Glutathione metabolism	5	0.007
GnRH signaling pathway	5	0.009
Adherens junction	5	0.009

Table 4.3: The table reports the functional enrichment of the sub-network identified. More specifically, all the genes within the clusters belonging to the sub-network have been tested together for enrichment using DAVID webservice and retrieving KEGG pathways. A 1% FDR threshold was applied.

4.4.4 Identification of chemicals with the potential to disrupt ovarian development

Having identified a sub-network with evidence for ovary development perturbation, I next asked the question of whether it was possible to identify chemical compounds that have the ability to perturb these functions. I therefore collapsed the clusters (2, 6, 7, 10, 12 and 17) within our sub-network to a single entity and interrogated the CTD database for potential entries of interest that have been shown to perturb this set of genes. This resulted in a list of 10 entries, Cyclosporine, Valproic acid, Copper Sulphate, Methyl Methanesulfonate, Cobalthous Chloride, Acetaminophen, Atrazine, Formaldheyde, Tretinoin and Quercetin, that showed a significant enrichment with the potential to perturb networks associated with ovary development (Table 4.4). Some of the identified entries are widely used as drugs (Valproic acid, Cyclosporine and Tretinoin) or food additives (Quercetin). Interestingly, three of the identified compounds (Valproic Acid, Cyclosporine and Quercetin) have already been shown to have estrogenic activity which increased confidence that my approach identified likely candidates for endocrine disruption activity.

CHEMICAL	DESCRIPTION	ENDOCRINE REFERENCE	FDR
Valproic Acid	Drug used to treat epilepsy and bipolar disorders. It acts as histone deacetylase, by blocking voltage-gated sodium channels or affecting GABA levels.	Estrogenic activity[256] Steroidogenic effect[257]	$8.9x10^{-34}$
Cyclosporine	Immunosuppressant drug used to reduce the activity of the immune system by interfering with the activity and growth of T cells.	Estrogenic effect[258]	$9.9x10^{-31}$
Copper sulphate	Inorganic compound with a wide range of application. It is mainly used by industries or as analytical reagent and is environmentally relevant.	Steroidogenic inhibition[259]	$3.3x10^{-21}$
Methyl methanesulfonate	Alkylating agent used in cancer treatment	Reproductive toxicity[260]	$1.0x10^{-13}$

Cobaltous Chloride	Inorganic compound with a wide range of application.	Indirect anti-estrogenic[261]	$1.7x10^{-12}$
Acetaminophen	Drug commonly known as Paracetamol, it is used to treat pain and fever	Anti-estrogenic activity[262] Anti-androgenic activity[263]	$1.7x10^{-11}$
Atrazine	Herbicide used to prevent broadleaf weed in crops	Anti-androgenic and anti-estrogenic[264]	$1.2x10^{-11}$
Quercetin	Flavonoid found in many fruits, vegetables, leaves and grains used as dietary supplement.	Estrogenic activity[265]	$1.6x10^{-11}$
Formaldehyde	Organic compound used for the production of resins and it is known to be carcinogen.	Reproductive toxicity[266]	$2.3x10^{-11}$
Tretinoin	A retinoic acid used to treat acne and leukaemia. It acts by forcing APL cells to differentiate and stops them from proliferating.		$6.6x10^{-10}$

Table 4.4: List of chemical compounds identified using the CTD database. These compounds are predicted to affect biological functions underlying the gene regulatory network of ovarian development. Compound description and endocrine references are reported along with p-values of enrichment.

To identify the best possible candidates for further testing, I compared each of the 6 compounds effects to well-known key endocrine-related genes (*esr1, esr2α, star*, *ctsD*, *ctsB*, *fst*, *cyp19a*, *cyp3a*, *nr0b1*, *zp3*) (Table 4.5). This identified cyclosporine, Acetaminophen, Atrazine, Tretinoin, Quercetin and valproic acid as chemicals likely to have the potential to disrupt endocrine functions. As all of these chemicals except Tretinoin have been previously shown to exhibit endocrine or reproductive toxicity-related effects, and Quercetin has been shown to impact mammalian ovary development[265], we then experimentally tested these two compounds as endocrine disruptors in LMB.

CHEMICAL	*esr1*	*esr2a*	*star*	*ctsD*	*ctsB*	*fst*	*cyp19a*	*cyp3a*	*nr0b1*	*zp3*
Valproic Acid	◎	◎	◎	◎	◎	◎	◎		◎	
Cyclosporine			◎	◎	◎	◎	◎			
Copper Sulfate			◎		◎	◎	◎			
Methyl Methanesulfonate										
Cobaltus Chloride										
Acetaminophen	◎				◎	◎				
Atrazine	◎	◎	◎		◎	◎	◎			
Quercetin	◎	◎		◎	◎		◎			
Formaldehyde										
Tretinoin	◎		◎	◎	◎	◎	◎		◎	

Table 4.5: The table reports some of the criteria considered for selecting chemical compounds to be experimentally validated for their ability to affect ovary development. Testing compounds were chosen on the basis of their ability to affect key-endocrine genes well-known drivers of ovary development.

4.4.5 Experimental validation of predicted compounds and their effects on endocrine related genes

To assess the potential of the two selected compounds, Quercetin and Tretinoin, to perturb ovary development, fish were exposed at 10 μg/Kg and 100 μg/Kg of each compound for 48 h. Doses were selected based on what we have found in literature as these compounds have already been tested in mice and toxicity doses have been extrapolated to fish. The expression of key endocrine genes, androgen receptor (*ar*), the three oestrogen receptors (*esr1*, *esr2a* and *esr2b*), aromatase (*cyp19a*), the steroidogenic acute regulatory protein (*star*) and the vitellogenin receptor (*vtgr*) were tested by qPCR in ovary and *esr1* and *vtg* in liver tissue. Neither of the compounds significantly affected the expression of androgen receptor in ovary tissue (Fig. 4.7). Quercetin significantly perturbed all three ERs at the highest dose examined (*esr2a* was significant also at the lowest dose) while Tretinoin perturbed *esr2a* only at the high dose. *Esr2a*,

in particular, appeared to be more responsive to the treatments than the other two receptors. The high dose of Quercetin also perturbed *cyp19a* and *vtgr* expression while Tretinoin affected *star* expression at 10 μg/Kg and *vtgr* expression at 100 μg/Kg. In the liver, the low dose of Tretinoin only significantly affected the expression levels of the *esr1* in males while *vtg* was not affected by either of the two compounds (Fig. 4.8).

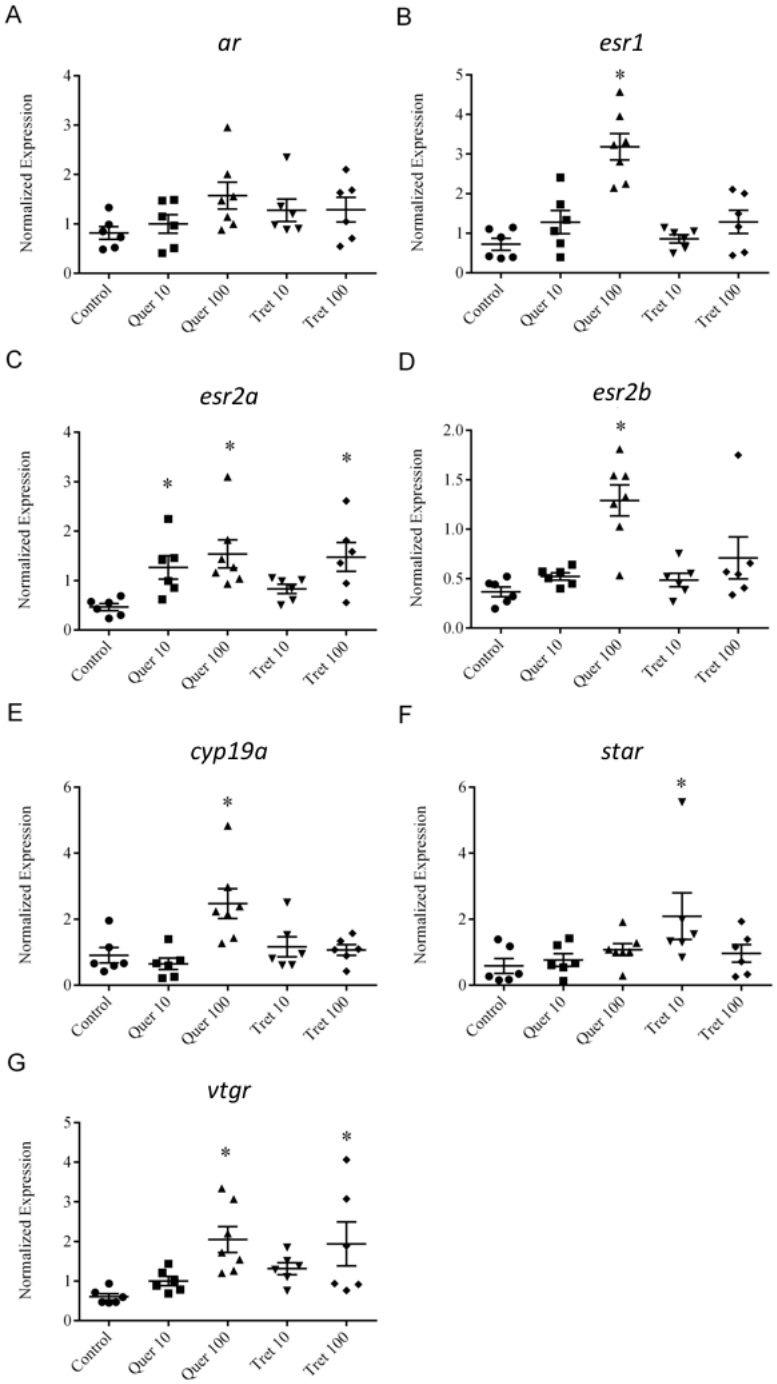


Figure 4.7: Expression levels of reproductive-related transcripts in ovary tissue after exposure to candidate chemicals.

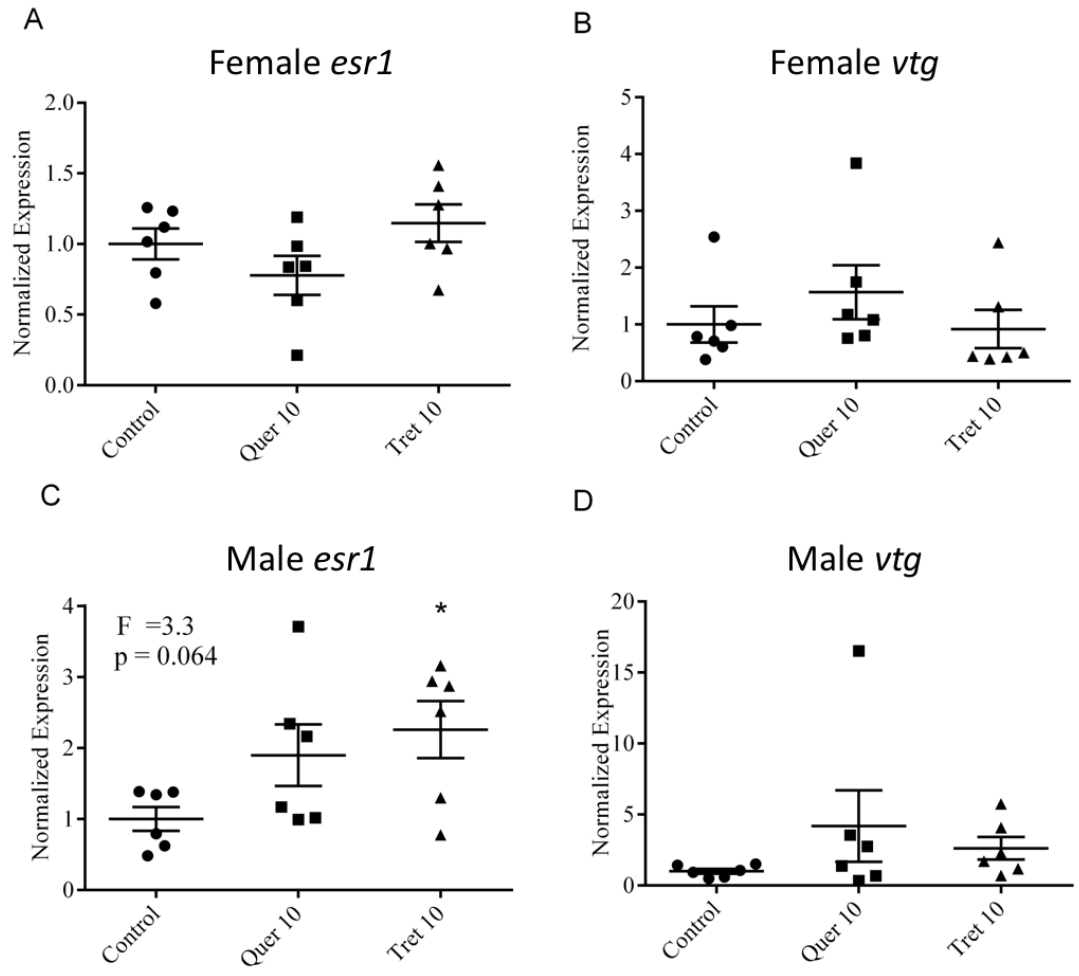


Figure 4.8: Expression levels of reproductive-related transcripts in liver tissue after exposure to candidate chemicals in both males and females.

To verify that my method preferentially identifies endocrine disruptors I applied a GSEA algorithm with the EDCs, as defined by the CTD, as the set. The EDC chemical set was found to be positively enriched (FDR < 5%) adding more confidence to our approach (Figure 4.9).

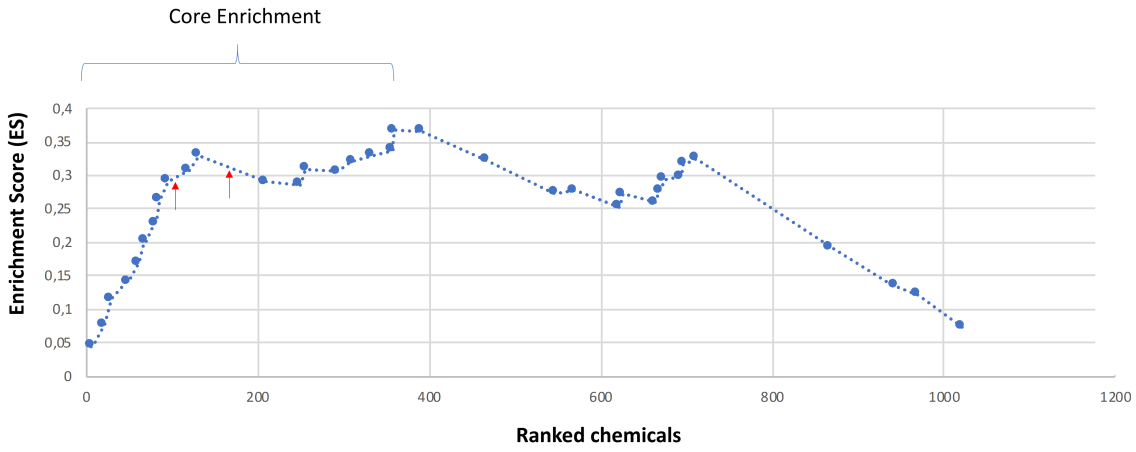


Figure 4.9: GSEA analysis of endocrine disruptor selection by our methodology. GSEA reported an enrichment of known endocrine disruptors in our list with an FDR < 5% suggesting that our methodology preferentially selects endocrine disruptors.

4.5 Discussion

4.5.1 Gene regulatory network of ovary development

Previously, we have characterized molecular pathways and temporal gene expression patterns in female largemouth bass across the different stages of ovarian development[8]. Here, I identified relationships between gene expression patterns in the different stages of ovarian development. This is the first study of its kind where a dynamic model is able to link gene expression patterns at early stages of development with those at later stages. Remarkably, I demonstrated that this approach can be used to identify chemicals that alter endocrine-related functions driving ovary development.

My first goal was to develop a dynamic model of healthy ovary development that functionally describes the dynamics undergoing ovarian maturation and oocyte growth. The approach I chose is entirely data driven. Unlike the currently available models that are based on the description of the pharmacodynamics underlying hormones and VTG activity along the hypothalamic-pituitary-

gonadal (HPG) axis[267–270], my model is able to capture completely novel regulatory interactions. I successfully identified clusters of genes whose expression was related to a particular stage of ovarian development. Previous studies aiming to characterize gene expression profiles underlying ovarian development have already been performed in different fish species and these are consistent with our model. Guzman et al.[271] characterized expression profiles of ovarian genes regulated by the follicle-stimulating hormone (*fsh*) in Coho salmon (*Oncorhynchus kisutch*). They determined that the expression of the transcript of *cyp19a* (aromatase) peaked at eVTG-lVTG stage and showed a positive correlation with the transcript of the *fsh* receptor (*fshr*), supporting the idea that *fsh* stimulates the production of E2 in the ovary via upregulation of *cyp19a*. In my network model, the *cyp19a* belongs to cluster 10, which has the initial change of expression at the eVTG-lVTG stage, supporting similar relationships as identified by Guzman et al. and well-established since decades[272]. Gene expression profiles during vitellogenesis were also determined in Atlantic cod (*Gadus morhua*) by Breton et al.[273]. They identified *cyp19a* as over-expressed (4-fold) during vitellogenesis in the ovary, which drives the synthesis of E2 that ultimately regulates *vtg* synthesis in the liver. In my network model, aromatase expression is included in cluster 10, which is connected with VTG, supporting the data presented by Breton and collaborators. Here I also inferred a gene regulatory network linking these profiles with key physiological endpoints. Moreover, the ability of either a physiological endpoint or a cluster at an early stage of development to trigger the activation or inhibition of features at a later stage of development was identified by using the temporal profile of ovary progression. Gene regulatory networks are particularly useful in the field of environmental toxicology for identifying chemical mode of action[274], deriving toxicity thresholds[275] and for inferring gene targets of drugs and chemical compounds[276]. The gene regulatory network I developed precisely aids in the

prediction of gene targets of chemical compounds that can be used potentially to develop a new set of biomarkers of ovarian toxicity.

One limitation of my model is that it does not consider all of the components of the HPG axis, I therefore may miss components/interactions of the system critical to ovary development. However, my simplification does consider time delays associated with signal transduction processes, leading to the development of a model where E2 and VTG play key roles during ovarian development as they show time sensitivity in reproduction. Development of more detailed and more biologically accurate mathematical models is required to better understand the full scope of effects that stressors are able to trigger[277,278].

4.5.2 Tretinoin and Quercetin as endocrine disruptors

Chemicals often used as drugs or food additives are widespread in the environment and many of them have already been characterized for their ability to disrupt fundamental biological processes in environmentally relevant species[279,280]. My computational approach led to the prediction of Tretinoin and Quercetin as potential endocrine disruptors (EDs) with the ability to alter key biological processes involved in the ovary development of the LMB.

My results suggest Quercetin has the ability to disrupt reproductive targets such as the estrogen receptors, the aromatase (*cyp19a*) and the *vtgr* in the ovary. Quercetin is a natural occurring flavonoid found in many fruits, vegetables, leaves and grains. Flavonoids, in plants, have many different roles as they improve growth and seedlings development, attract pollinators helping seed germination and are responsible for the aroma and the colours of flowers[281]. Moreover, they serve as a barriers against many environmental stresses such as UV radiation[282]. Quercetin has been demonstrated to be beneficial to health in mammals because of its antioxidative, anticancer, free radical scavenging and antiviral activities[283,284]. Its presence in the environment

can be associated with industrial effluents from the pulp and paper industry processing plant material[285,286]. Quercetin is relevant for aquaculture as it has been investigated as a potential fish food supplement for its beneficial properties[287]. Moreover, it also has potential effects in lowering levels of lipids[288–290] whose blood presence in fish has been associated with declining health conditions[291]. Few studies have also demonstrated that Quercetin may improve follicular development and oocyte quality *in vitro* and *in vivo*[292,293]. My findings suggest that low amounts of Quercetin only affect the expression of *esr2a*. However, Quercetin may also have negative effects on fish as shown by Weber et al., which demonstrated that Quercetin exposure (100ppb) in female Japanese medaka (*Oryzias latipes*) promoted follicular atresia[294]. Further studies have also shown that Quercetin has estrogenic-like effects on ovary development[265].

Further experimental evidence for the potential of Tretinoin to act on key endocrine genes was conducted and demonstrated its effect on *esr2a*, *vtgr* and *star*. Tretinoin, also called all-trans retinoic acid (ATRA), is. one of the metabolites of vitamin A (retinol). Retinoic acid is a biologically active metabolite of vitamin A (retinol) that, through the binding with retinoic acid receptors (RARs and RXRs), is involved in a wide range of biological functions including embryo development[295,296], immune system[297], reproduction[298,299] and vision system[300]. Although Tretinoin mechanism of action is unknown, it may elicit its molecular action through the activation of retinoid receptors[301] as well as PPAR[302]. Tretinoin is a drug used worldwide for the treatment of acne vulgaris and photodamage[301]. Despite its common use, few studies have been conducted to address any potential environmental toxicity, and most of these studies reveal developmental toxicity[303–305]. The only study which investigated potential effects of Tretinoin on ovarian developmental processes was carried out by Pu et al., which showed that ATRA improved *in vitro* oocyte nuclear maturation in goat after a 22h exposure at concentrations below the ones tested in this publication[306]. Despite the fact that this compound is not classified as a chemical

of concern for the environment, it nevertheless demonstrates that my approach can predict chemicals *de novo* as potential reproductive disruptors.

The additional experimental test performed have revealed new insight into the toxicity of Tretinoin and Quercetin, supporting predictions that these compounds can act as EDs based upon changes in mRNA levels for estrogen receptors, aromatase, *star* and *vtgr*. Interestingly, the two compounds seemed to act differently. Expression levels of all estrogen receptors were affected by Quercetin. It is interesting to notice the potential non-monotonic dose-response behavior of Tretinoin for *star* expression levels which has been previously observed in ED compounds[307]. However, further experimental validation is required to understand the relationship between Tretinoin and this endocrine disruption potential and more doses are needed to confirm non-monotonic dose-response behaviours. For the first time, I provide evidence that Tretinoin can affect transcripts related to steroidogenesis and *vtgr* mRNA levels. Evidence of Quercetin ovarian toxicity was associated also with aromatase activity. This demonstrates that omics analyses of target-organ specific perturbation can identify highly relevant toxicants that have yet to be tested. Validation of my predictions further increase our confidence that my findings have the potential to improve environmental risk assessment as well as providing a new tool for screening chemical compounds.

Chapter 5: Daphnia as test organism for endocrine disruption

The whole experimental design was performed by our collaborating lab while I have developed data analysis strategies and performed all the in silico analysis.

5.1 Abstract

Recently, there has been increasing evidence that chemicals that act as endocrine disruptors may have an effect in invertebrates, mainly insects and crustaceans. *Daphnia magna* has been proposed as a model organism for environmental ecotoxicology being a good biosensor for endocrine disruption because these chemicals may act by the juvenile and moulting hormone, endocrine hormones controlling development and growth. However, so far there is no direct evidence that all chemicals that are known endocrine disruptors in vertebrates truly act on this *Daphnia magna* pathways. Here I address this question by first constructing gene expression signatures representing Daphnia response to both Juvenile and moulting hormones and then by comparing a panel of chemicals to the reference expression profiles of these two hormones. Interestingly, I identified Diazinon and Atrazine on one side and λ-cyhalothrin and bifenthrin on the other one, to have a similar expression profile to either juvenile and the moulting hormone, respectively. This finding is consistent with the working hypothesis that ED's act affecting these hormones signalling pathways. My findings suggest that the approach I have developed could be used to identify novel endocrine disrupting compounds hence supporting the use of *Daphnia magna* for rapid screening in risk assessment.

5.2 Introduction

Environmental contaminants have been increasingly investigated in the last decades for their ability to interfere with fundamental biological functions of living organisms. Because of their effect on the maturation, growth and reproduction, chemicals that are able to affect the endocrine system are of particular importance in ecotoxicology. Many of these endocrine disrupting chemicals (EDCs) elicit their toxic effects by mimicking the action of endogenous hormones, for example by binding to their nuclear receptors[308,309]. Even if invertebrates account for around 95% of all animals[310], no much effort has been invested in understanding their potential in signalling environmental endocrine disruption and most of the studies of EDCs effects have been performed on vertebrates.

The endocrine system of both vertebrate and invertebrate species controls homeostasis, development, growth and reproduction[311,312]. However, vertebrate and invertebrate endocrine systems present significant differences. First, the vertebrate endocrine system involves a hierarchical array of organs where hormones are produced and released by the same endocrine gland and they reach their target organs by the circulatory system. The invertebrate endocrine system, on the other hand, is more centralized with hormones produced and stored in different organs and hemolymph is responsible for delivering them to their target sites. Also, the vertebrate endocrine system is composed of both neurosecretory cells and endocrine glands while the invertebrate endocrine systems present fewer true glands. The arthropods, in particular insects and crustaceans, are the only group presenting true endocrine glands similar to their vertebrate counterparts. Most of the knowledge we currently have about the invertebrate endocrine system is coming from studies on insects as the insect class is the richest

within the invertebrates[313]. Moreover, members of the arthropod subphylum Crustacea are the one dominating the aquatic environment. Differences in endocrine system components are present in the different species but the general functioning is conserved. Crustaceans endocrine system is made up of different organs of which the X-organ-sinus gland complex is the main neuroendocrine component and strikingly resemble the brain-corpora cardiaca complex of insects[314]. Hormones are produced by the X-organ which is connected by axonal terminals with the sinus gland in which they are stored. The main hormones secreted by the sinus gland include the MIH (molt inhibiting hormone)[315], the GIH (gonad inhibiting hormone)[316], the MOIH (mandibular organ inhibiting hormone)[317], the NDH (neurodepressing hormone)[318] and the CHH (crustacean hyperglycaemic hormone)[319]. Crustacean moulting is controlled by the MIH which exert an inhibitory effect on the Y-organs (structurally homologous to the insect prothoracic organs) responsible for the production of the moulting hormone, MH (20-hydroxyecdysone)[320,321]. Prior to moulting, MIH levels decrease triggering increasing levels of ecdysteroids that lead to moulting[322]. MH acts through a dimeric nuclear receptor that binds DNA and activates de novo gene transcription. Interestingly, one of its monomers, namely ultraspiracle (USP) is structurally similar to the vertebrate retinoid-X-receptor (RXR)[323]. It has been shown that vertebrate endocrine disruptors can mimic the effect of the MH and are able to induce male offspring by controlling sex determination. Ecdysteroids have been shown to play an important role during embryonic development and genes responsible for their synthesis have been identified[324]. Hannas et al. showed ecdysteroids are able to affect mRNA levels of vitellogenin genes[325]. However, moulting is not under the sole control of MH. Juvenile hormones, JHs, are terpenoids which control reproduction and maturation[326]. Among the several forms of juvenile hormone identified, the JH III is the one occurring in most insects. In crustaceans, methyl farnesoate is the principal juvenile hormone synthesized[327]. Methyl farnesoate is produced by

the mandibular organ (homologous to the insect corpora allata) and is under the negative control of MOIH. The molecular basis of juvenile hormone signalling has been recently described by Jindra et al[328]. High levels of JH are necessary during the pre-moulting period to maintain the juvenile form. Decreasing level of JH will trigger the development of mature features and the moulting process. The role of methyl farnesoate in the regulation of moulting and reproduction has been investigated by Reddy et al.[329]. Moreover, the interplay between juvenile hormones and ecdysteroids has been demonstrated[330–333]. JH plays an important role also in sex determination. Daphnids produce female offspring by parthenogenesis under favourable environmental conditions, but in response to various unfavourable external stimuli, it produces male offspring. This process is regulated by methyl farnesoate[334]. Good reviews of invertebrates and crustaceans endocrinology and potential effects endocrine disrupting compounds are able to elicit are given by LaFont, LeBlanc, Soin and Rodriguez, respectively[312,335–337]. Therefore, the dominant signaling pathways driving moulting, maturation and reproduction in crustaceans and, more in general in insects, are either ecdysteroids or methyl farnesoate related.

The crustacean *Daphnia magna* is an established model species to study aquatic environmental toxicity[50]. Most of these studies have applied standard ecotoxicological assays using established endpoints to determining hormone agonist effects on daphnids and just a few studies have applied gene expression profiling approaches[338]. In the recent decades, studies investigating effects of endocrine disruptors on *D. magna* have been constantly increasing and most of them identified a good number of compounds to act as ecdysteroids or methyl farnesoate analogs. Ginjupalli and Baldwin identified Pyriproxyfen, a juvenile hormone analog, to affect D.magna reproduction by increasing male production and decreasing overall fecundity[339]. Giraudo et al., recently demonstrated the ability of three different benzotriazoles, ubiquitous aquatic contaminants due to industrial and domestic activities, to regulate endocrine molecular

processes linked to ecdysteroids signalling[340]. Recently, Song et al. proposed the first invertebrate ED AOP linking MoA of EDs affecting moulting process with adverse effects according to OECD guidelines[28].

In this study I tested the ability of a wide range of compounds, including a number of known EDs, to affect biological pathways linked to the activity of juvenile and moulting hormones defined as a global transcriptional response. Interestingly, I show that some of these compounds elicit a similar transcriptional response to either the JH or MH. Moreover, by comparing the expression signature of *Daphnia magna* exposed to JH and MH with a collection of test compounds I have been able to identify novel chemicals that may potentially affect *D. magna* endocrine system. More specifically, my analysis revealed that exposure to Diazinon and Atrazine elicit a transcriptional response that represents biological functions associated with steroid metabolic process and regulation of hormone levels. This is remarkably similar to the transcriptional response elicited by exposure to JH. Interestingly, exposure to λ-cyhalothrin and bifenthrin induce a transcriptional response that is similar to exposure to the moulting hormone. These findings support the use of *Daphnia magna* for rapid screening of endocrine disrupting compounds in risk assessment.

5.3 Materials and methods

5.3.1 The dataset

Transcriptional data from exposure of *Daphnia magna* to a panel of environmentally relevant compounds were obtained from a previous study[341]. Among the 24 compounds used in the previous study, we added back Aroclor1242 given the objective of our study. Briefly, the dataset represents gene expression profiles of Daphnia magna adults exposed for 24 h to nLC_{50} of a panel of compounds (table 5.1). Total RNA extraction was performed and arrayed using a custom

Agilent microarray (AMAID: 023710, GPL15139). Nine out of the twenty-five compounds were classified as Endocrine Disruptors (ED's) (fig. 5.1). Among them, two were the natural moulting and juvenile hormone, 20-hydroxyecdysone and methyl farnesoate respectively. Moreover, Piriproxyfen is an analog of the MF while Ponasterone A is the analogue of the moulting hormone.

CHEMICAL	LIPOPHILICITY
2-chloroethylvinyl ether	1.29
20-hydroxyecdysone	1.3
Acrylonitrile	0.25
Aroclor1242	6.3
Atrazine	2.61
Bifenthrin	6
Chloroform	1.97
Chlorpyrifos	4.96
Diazinon	3.81
Dichlorobenzene	3.38
Methoxychlor	4.95
Methylfarnesoate	5.61
MTBE	0.94
Nonylphenol	5.76
Parathion	3.83
Permethrin	6.5
Phenanthrene	4.52
Phenol	1.5
Ponasterone A	2.14
Pyriproxyfen	4.89
Toluene	2.73
Toxaphene	5.28
Trichloroethylene	2.42
β-estradiol	4.01
λ-cyhalothrin	7

Table 5.1: Panel of the 25 chemicals used in our project. Compound selected encompass a wide range of lipophilicity values.

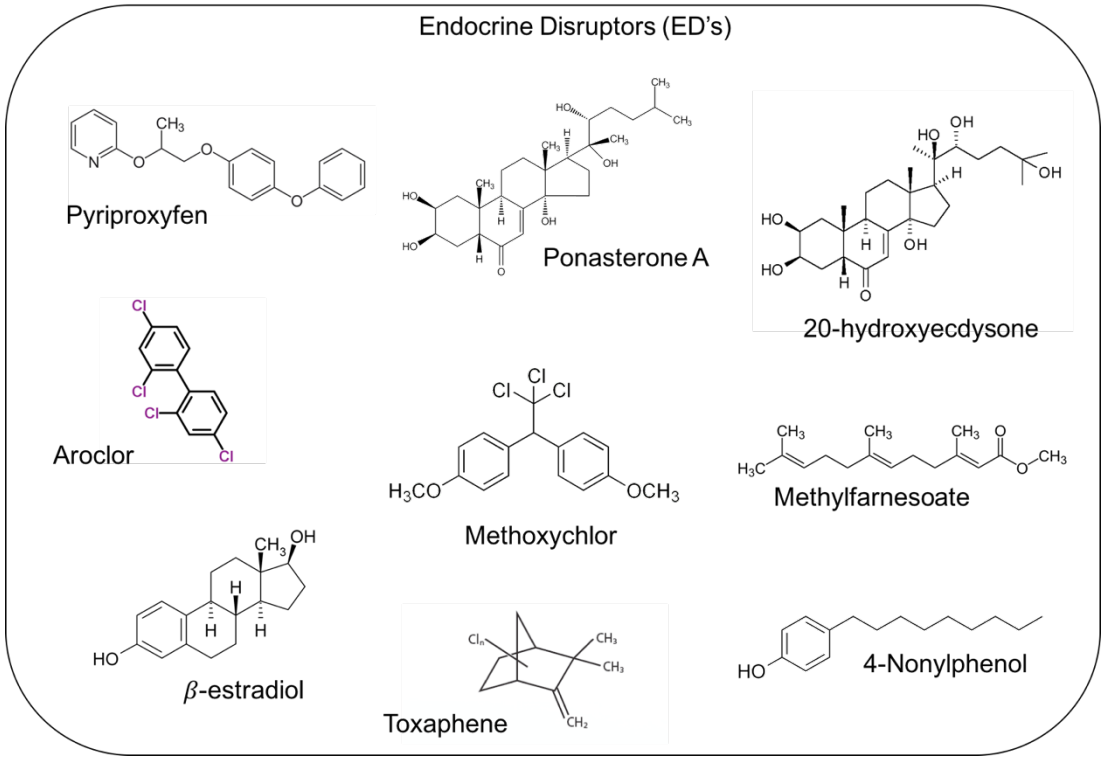


Figure 5.1: The graph shows the 9 ED's we included in our study.

Before running downstream analysis, I set to re-annotate the array. I used annotation coming from a personal communication with Bruno Campos where the annotation is based on genome information. The daphnia dataset was made of 14338 probes (13481 unique probes). Only 4988 out of the 13481 probes were successfully annotated (3656 non-redundant genes). I successfully increased the number of annotated genes from the initial 1425.

A schematic representation of the strategy analysis we followed is in fig. 5.2. I first carried out an exploratory analysis to look for lipophilicity signatures. I then set to characterize the components of the transcriptional responses that were either linked or not to compound lipophilicity. After removing K_{ow} associated signatures from the initial dataset, I set to define both 20-hydroxyecdysone and methyl farnesoate functional profile. Afterwards, I investigated the ability

of ED's as well as other non-organic compounds to affect biological pathways triggered by the action of these two hormones. Finally, I apply a statistical modelling procedure to identify a predictor of endocrine disruption.

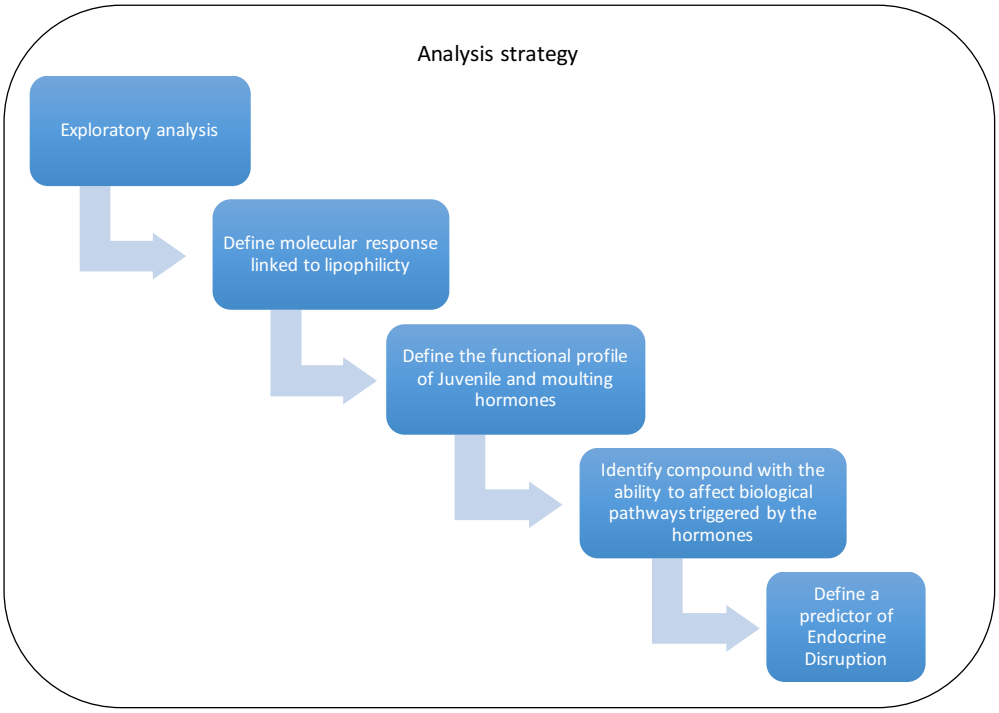


Figure 5.2: The flow chart shows the strategy analysis we followed. I first run an exploratory analysis by leveraging the PCA analysis. I then Discriminated between basal and excess toxicity signatures. Next, we defined the functional profile of the two hormones regulating Daphnia endocrine system and we successively identify compounds whose transcriptional profile was comparable to the hormones. Finally, a predictive model of endocrine disruptors was computed.

5.3.2 Differential gene expression analysis

To identify genes correlated to compound lipophilicity I applied a Quantitative Significance Analysis of Microarray[117] (Q-SAM) as implemented in the "samr" package developed in the statistical environment R. Chemical compounds were first ordered by increasing K_{ow} value and we then run the quantitative SAM analysis using the lipophilicity values as response variables. Quantitative SAM identifies genes that correlate with a given response variable, K_{ow} in this case.

A 1% FDR threshold was applied. Heat-maps were obtained using Spearman correlation as distance method and average as the clustering method. To identify genes differentially expressed as a result of exposure to 20-hydroxyecdysone and methyl farnesoate as well as to the other compounds I used a one-class sam. Genes up and down-regulated were selected applying an FDR threshold of 5%.

5.3.3 Functional analysis

Functional analysis of lipophilicity dependent and independent signatures as well as the one of juvenile and moulting hormone signatures was carried out at a pathway level. KEGG pathways were retrieved using DAVID web service[118]. Pathways enriched at 1% FDR and represented by at least 5 genes were selected and classified within a broader functional domain according to the KEGG levels of organization. Functional analysis of each of the compound in the panel, used to run a comparative analysis between the hormones and all the other compounds, was achieved at both gene and pathway level. Biological GO terms and KEGG pathways associated with each of the chemical were retrieved using DAVID and applying an FDR threshold of 5% and 1% respectively. The comparative functional analysis between K_{ow} and K_{ow}-free molecular signatures as well as between the hormones and the other compounds in the panel was run within the statistical environment R using the full set of differentially expressed genes (up and down-regulated genes), simply looking at the functional overlap. Redundant GO terms were summarized using REVIGO web service[122] using default parameters (allowed similarity of 0.7 and simRel as functional similarity measure) and selecting *Homo sapiens* database to identify the size of each GO term.

5.3.4 Defining distances between hormones and other compounds

To identify chemical compounds that could potentially affect pathways underlying moulting or juvenile activity I set to use Euclidean distance coupled with a principal component approach. I

first summarized sample replicates by average. I then retrieve principal components able to explain at least 80% of variance. A total of 8 PC's were retrieved (80.6% of variance) and Euclidean distances between hormones of interest and the other compounds was computed. The whole analysis was achieved using the "prcomp" function within the statistical environment R.

5.3.5 Statistical modelling

To find a predictor of endocrine disruption I employed a variable selection strategy that uses a genetic algorithm methodology coupled with the KNN classification method as implemented in the GALGO package developed in the statistical environment R. KNN is a simple classification algorithm with the ability to do non-linear classification and identify quite complex patterns. KNN was set-up using default parameters: 3 as the number of neighbours to consider, 1 as the minimum number of neighbours of the same class needed to predict the sample in that class and "Euclidean" as distance method. The modelling procedure was run using the default settings with a model size of 5. Model accuracy was estimated splitting the data into a training set (2/3) and a test set (1/3). At this stage the training set is used to build statistical models and their accuracy is computed using the test set by applying a leave-one-out cross validation procedure. A 1000 models were computed to reach the fitness goal of 95%. Out of the 1000 models, one representative model was developed using a forward selection strategy. This approach ranks the model variables (genes) based on their frequency in the population and the top 50 most present variables are incrementally tested, by adding each variable one by one. The final representative model is chosen as the one with the higher accuracy and the smaller number of variables.

5.4 Results

5.4.1 Transcriptional response to chemical exposure is influenced by compound lipophilicity

Our group has previously demonstrated that a considerable fraction of the transcriptional response following exposure to a wide range of chemicals (see method section for a description of the chemical exposure set) is dependent on the lipophilicity of compounds[341]. Interestingly, chemicals with K_{ow}>1.8 induce a dramatically different transcriptional response when compared to lower K_{ow} compounds. I, therefore, reasoned that in order to identify specific transcriptional signatures associated with hormone activity I may need to remove genes whose expression correlate with compound lipophilicity. I first performed an exploratory analysis using principal component analysis to verify that the largest source of variation is linked to lipophilicity (fig. 5.3).

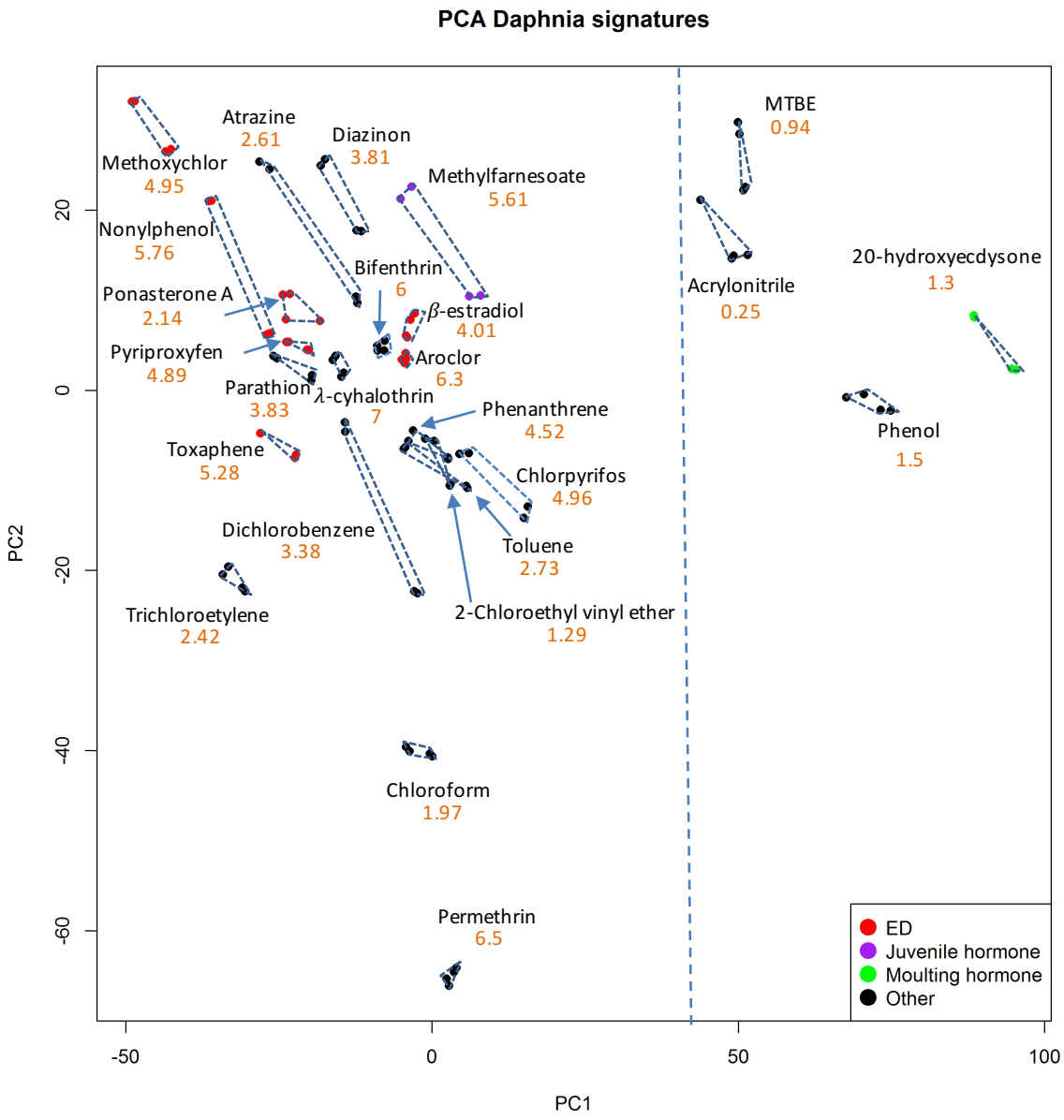


Figure 5.3: PCA clearly shows the separation based on K_{ow} (values reported). PC1 (x axis) and PC2 (y axis) explain 29.4% and 11.3% of variance respectively.

Consistent with previous observations, the first principal component (PC1) separates chemicals according to their K_{ow}. Moreover, ED's (marked in red in Figure 5.3) group together suggesting

they may have a similar expression profile. This suggests that they may act on a similar mechanism.

Having verified that lipophilicity is an important factor, I then set to identify genes that significantly correlated to compound lipophilicity. I identified 2174 genes to be correlated with K_{ow} (1159 positively and 1015 negatively) at 1% FDR (fig. 5.4).

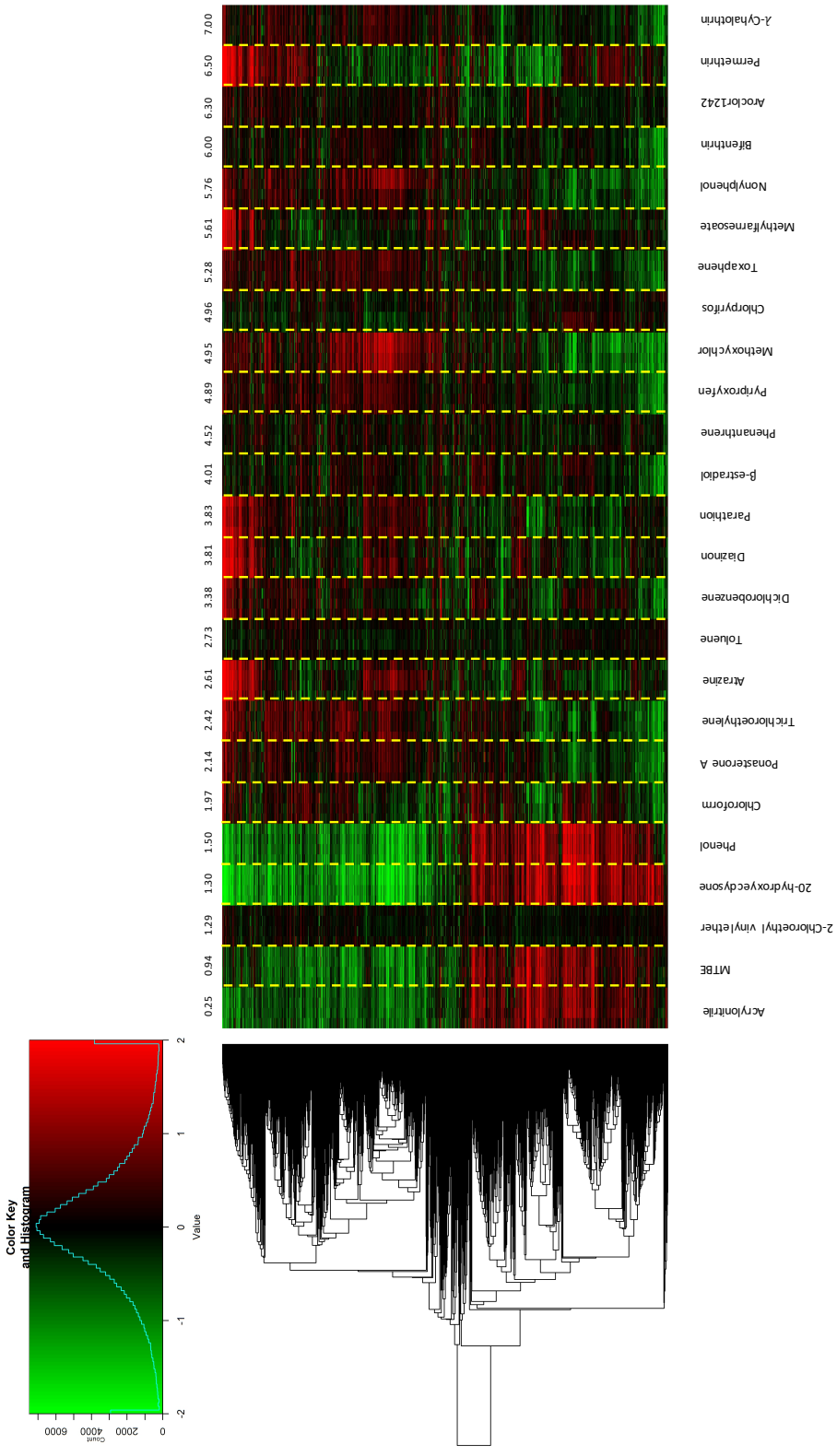


Figure 5.4: Genes found to be either positively (red) or negatively (green) correlated with K_{ow}. Chemicals are ordered by increasing K_{ow}. A transcriptional switch at K_{ow} value of around 1.8 is visible. Heat-map has been performed using Spearman correlation as distance method and average as the clustering method for the genes.

As previously described in Antczak et al.[341], I can observe a transcriptional switch at K_{ow} =~ 1.8. Interestingly, 20-hydroxyecdysone and methyl farnesoate have very different K_{ow} values and the heat-map reveals expression profile of these two hormones to potentially antagonize each other. To facilitate the interpretation of the lipophilicity signature I performed a pathway-level analysis. As genes do not work alone but in the context of complex biological networks, pathway level analysis offers a better overview of the biological functions ongoing into the cell as a result of a specific exposure. I identified 117 pathways enriched in genes correlated to K_{ow}. Specifically, 69 of them (59%) were positively correlated to K_{ow} while 48 (41%) were negatively correlated to K_{ow} (table 5.2 and 5.3). Moreover, 28 pathways (24%) included both positively or negatively K_{ow} correlated genes.

Domain	KEGG ID	Pathway Name	Genes
Carbohydrate and energy metabolism	hsa00010	Glycolysis / Gluconeogenesis	9
	hsa00030	Pentose phosphate pathway	5
	hsa00051	Fructose and mannose metabolism	7
	hsa00052	Galactose metabolism	10
	hsa00500	Starch and sucrose metabolism	13
	hsa00520	Amino sugar and nucleotide sugar metabolism	11
	hsa00620	Pyruvate metabolism	6
	hsa00562	Inositol phosphate metabolism	7
Lipid and nucleotide metabolism	hsa00071	Fatty acid metabolism	6
	hsa00564	Glycerophospholipid metabolism	7
	hsa00230	Purine metabolism	14
	hsa00240	Pyrimidine metabolism	8
Amino acids and glycan metabolism	hsa00250	Alanine, aspartate and glutamate metabolism	8
	hsa00270	Cysteine and methionine metabolism	7
	hsa00330	Arginine and proline metabolism	6
	hsa00340	Histidine metabolism	6
	hsa00350	Tyrosine metabolism	9
	hsa00410	beta-Alanine metabolism	5
	hsa00450	Selenoamino acid metabolism	5
	hsa00510	N-Glycan biosynthesis	12
	hsa00512	O-Glycan biosynthesis	5
Xenobiotics metabolism	hsa00982	Drug metabolism - cytochrome P450	5
	hsa00983	Drug metabolism - other enzymes	7
Transcription, folding, sorting and degradation	hsa03040	Spliceosome	22
	hsa04120	Ubiquitin mediated proteolysis	16
	hsa03018	RNA degradation	13
	hsa03030	DNA replication	9
Signal transduction	hsa04010	MAPK signaling pathway	17
	hsa04012	ErbB signaling pathway	9
	hsa04310	Wnt signaling pathway	19
	hsa04340	Hedgehog signaling pathway	5
	hsa04350	TGF-beta signaling pathway	9
	hsa04370	VEGF signaling pathway	7
	hsa04020	Calcium signaling pathway	12
	hsa04070	Phosphatidylinositol signaling system	7

(Continue next page)

(Continued)

Domain	KEGG ID	Pathway Name	Genes
Signaling molecules and interaction	hsa04080	Neuroactive ligand-receptor interaction	12
	hsa04060	Cytokine-cytokine receptor interaction	6
	hsa04512	ECM-receptor interaction	7
	hsa04514	Cell adhesion molecules (CAMs)	6
Transport and catabolism and cell motility	hsa04144	Endocytosis	12
	hsa04142	Lysosome	14
	hsa04810	Regulation of actin cytoskeleton	16
Cell growth and death	hsa04110	Cell cycle	18
	hsa04114	Oocyte meiosis	11
	hsa04115	p53 signaling pathway	5
Cellular community	hsa04510	Focal adhesion	21
	hsa04520	Adherens junction	12
	hsa04530	Tight junction	13
	hsa04540	Gap junction	9
Immune system	hsa04640	Hematopoietic cell lineage	5
	hsa04620	Toll-like receptor signaling pathway	5
	hsa04650	Natural killer cell mediated cytotoxicity	6
	hsa04660	T cell receptor signaling pathway	9
	hsa04664	Fc epsilon RI signaling pathway	7
	hsa04666	Fc gamma R-mediated phagocytosis	9
	hsa04670	Leukocyte transendothelial migration	12
	hsa04062	Chemokine signaling pathway	11
Endocrine system	hsa04910	Insulin signaling pathway	15
	hsa04920	Adipocytokine signaling pathway	8
	hsa03320	PPAR signaling pathway	9
	hsa04912	GnRH signaling pathway	8
	hsa04914	Progesterone-mediated oocyte maturation	14
	hsa04916	Melanogenesis	10
	hsa04614	Renin-angiotensin system	5
Circulatory and nervous system and development	hsa04270	Vascular smooth muscle contraction	11
	hsa04730	Long-term depression	9
	hsa04722	Neurotrophin signaling pathway	11
	hsa04320	Dorso-ventral axis formation	5
	hsa04360	Axon guidance	9

Table 5.2: List of significant pathways (1% FDR) represented by at least 5 genes found to be positively associated with lipophilicity. Pathways are arranged according to the KEGG levels of organization. FDR values of each of the Kegg pathways are provided in the supplementary material.

Domain	KEGG ID	Pathway Name	Genes
Carbohydrate and energy metabolism	hsa00020	Citrate cycle (TCA cycle)	7
	hsa00190	Oxidative phosphorylation	26
	hsa00910	Nitrogen metabolism	5
Lipid and nucleotide metabolism	hsa00564	Glycerophospholipid metabolism	7
	hsa00590	Arachidonic acid metabolism	8
	hsa00230	Purine metabolism	15
	hsa00240	Pyrimidine metabolism	11
Amino acids and xenobiotics metabolism	hsa00260	Glycine, serine and threonine metabolism	7
	hsa00280	Valine, leucine and isoleucine degradation	8
	hsa00330	Arginine and proline metabolism	8
	hsa00480	Glutathione metabolism	6
	hsa00983	Drug metabolism - other enzymes	6
Transcription, translation, folding, sorting and degradation, replication and repair	hsa03020	RNA polymerase	5
	hsa03022	Basal transcription factors	6
	hsa03040	Spliceosome	16
	hsa03010	Ribosome	31
	hsa00970	Aminoacyl-tRNA biosynthesis	6
	hsa04130	SNARE interactions in vesicular transport	6
	hsa04120	Ubiquitin mediated proteolysis	14
	hsa03050	Proteasome	8
	hsa03420	Nucleotide excision repair	8
Signal transduction and interaction	hsa04010	MAPK signaling pathway	15
	hsa04310	Wnt signaling pathway	15
	hsa04330	Notch signaling pathway	7
	hsa04350	TGF-beta signaling pathway	6
	hsa04020	Calcium signaling pathway	14
	hsa04150	mTOR signaling pathway	5
	hsa04080	Neuroactive ligand-receptor interaction	6
Transport and catabolism, cell motility and cell growth and death	hsa04144	Endocytosis	13
	hsa04810	Regulation of actin cytoskeleton	16
	hsa04110	Cell cycle	11
	hsa04114	Oocyte meiosis	14
	hsa04210	Apoptosis	6
Cellular community	hsa04510	Focal adhesion	12
	hsa04520	Adherens junction	8
	hsa04530	Tight junction	11
	hsa04540	Gap junction	6
Immune system	hsa04612	Antigen processing and presentation	9
Endocrine system	hsa04910	Insulin signaling pathway	11
	hsa03320	PPAR signaling pathway	6
	hsa04912	GnRH signaling pathway	9
	hsa04914	Progesterone-mediated oocyte maturation	6
	hsa04916	Melanogenesis	9
Circulatory and nervous system and development	hsa04260	Cardiac muscle contraction	9
	hsa04270	Vascular smooth muscle contraction	13
	hsa04720	Long-term potentiation	8
	hsa04730	Long-term depression	10
	hsa04722	Neurotrophin signaling pathway	7

Table 5.3: List of significant pathways (1% FDR) represented by at least 5 genes found to be negatively associated with lipophilicity. Pathways are arranged according to the KEGG levels of organization. FDR values of each of the Kegg pathways are provided in the supplementary material.

Biological pathways found to be associated with lipophilicity cover a wide spectrum of functions. Metabolism pathways (energy, lipid and amino acids) are the ones most represented. Carbohydrate and energy metabolism pathways found to be positively correlated with lipophilicity includes pyruvate metabolism, glycolysis/gluconeogenesis and inositol phosphate metabolism while oxidative phosphorylation and citrate cycle (TCA cycle) were found to be negatively correlated with lipophilicity. Lipid metabolism includes fatty acid metabolism and arachidonic acid metabolism which were positively and negatively correlated with lipophilicity respectively. Moreover, glycerophospholipid metabolism included both genes positively and negatively correlated to lipophilicity. Interestingly, signalling pathways are perturbed. For example, I saw that genes in the ErbB, hedgehog, VEGF signalling pathways were found to be positively correlated while notch signalling was negatively correlated to lipophilicity. Also, Wnt and TGF-beta signalling pathways were found to be represented by genes that were either positively or negatively correlated to lipophilicity. These genes play an important role during embryo development. Toll-like receptor and T-cell signalling pathways involved either in cell signalling and immunity were found to be positively correlated while calcium and insulin signalling pathways, involved in calcium and glucose homeostasis, includes genes either positively or negatively correlated to lipophilicity. Since they act through membrane receptor triggering signalling cascades, it is not surprising that transcriptional signatures are linked to a mechanism of toxicity that is based on a non-specific membrane interaction. Pathways involved in cell growth and death as cell cycle and apoptosis are also affected. Lastly, pathways underlying endocrine system were also differentially correlated. These results are consistent with the

hypothesis that highly lipophilic compounds are able to affect a wide range of biological pathways. From a practical perspective, the molecular signatures I have identified can be now eliminated from the data. The resulting new dataset could then be used to search for transcriptional signatures specifically linked to endocrine disruption.

5.4.2 Comparative analysis between lipophilicity dependent and independent signatures

Having characterized the functional profile of the lipophilicity gene signature, I set to identify the component of the chemicals transcriptional response that is lipophilicity independent. To retrieve transcriptional responses not affected by lipophilicity I removed signatures found to be correlated to the K_{ow} from the initial dataset. This new dataset included 1,482 gene expression profiles (fig. 5.5).

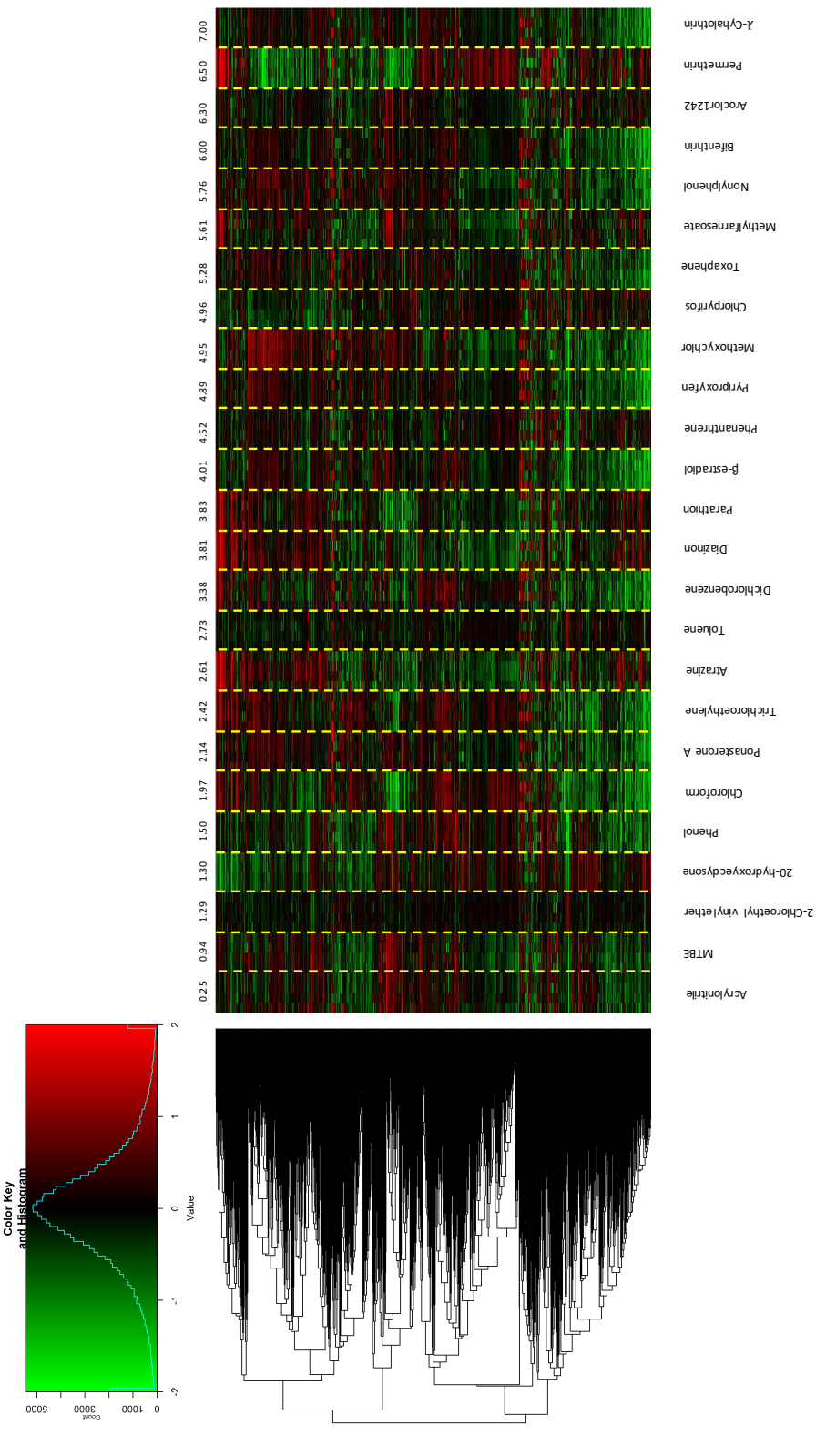


Figure 5.5: Genes which are not affected by compound lipophilicty. Chemicals are ordered, as in figure 2, by increasing K_{ow}. Heat-map has been performed using Spearman correlation as distance method and average as the clustering method for the genes.

The heat-map representing this gene signature shows that indeed the expression of these genes do not follow the paradigm of a lipophilicity dependent transcriptional switch.

Functional analysis of this dataset (table 5.4) also reveals a wide range of biological functions. Surprisingly, I observed a high degree of similarity at a functional level between lipophilicity dependent and independent signatures. 63 pathways (81% of the lipophilicity independent functional profile) were found to be affected by both the lipophilicity dependent and independent signature. Interestingly, signalling pathways affected by lipophilicity as ErbB, Wnt, Notch, Hedgehog, VEGF (Vascular Endothelial Growth Factor) and Jak-STAT were not found to be enriched by lipophilicity independent signatures.

Domain	KEGG ID	Pathway Name	Genes
Carbohydrate and energy metabolism	hsa00010	Glycolysis / Gluconeogenesis	12
	hsa00020	Citrate cycle (TCA cycle)	8
	hsa00040	Pentose and glucuronate interconversions	5
	hsa00051	Fructose and mannose metabolism	6
	hsa00500	Starch and sucrose metabolism	6
	hsa00520	Amino sugar and nucleotide sugar metabolism	8
	hsa00620	Pyruvate metabolism	10
	hsa00640	Propanoate metabolism	10
	hsa00650	Butanoate metabolism	11
	hsa00562	Inositol phosphate metabolism	7
	hsa00190	Oxidative phosphorylation	14
Lipid and nucleotide metabolism	hsa00071	Fatty acid metabolism	12
	hsa00140	Steroid hormone biosynthesis	10
	hsa00561	Glycerolipid metabolism	10
	hsa00564	Glycerophospholipid metabolism	11
	hsa00565	Ether lipid metabolism	5
	hsa00591	Linoleic acid metabolism	6
	hsa00230	Purine metabolism	23
	hsa00240	Pyrimidine metabolism	19
Amino acids and glycan metabolism	hsa00260	Glycine, serine and threonine metabolism	6
	hsa00280	Valine, leucine and isoleucine degradation	14
	hsa00290	Valine, leucine and isoleucine biosynthesis	5
	hsa00310	Lysine degradation	11
	hsa00330	Arginine and proline metabolism	8
	hsa00340	Histidine metabolism	6
	hsa00380	Tryptophan metabolism	11
	hsa00410	beta-Alanine metabolism	5
	hsa00450	Selenoamino acid metabolism	7
	hsa00480	Glutathione metabolism	11
	hsa00510	N-Glycan biosynthesis	10
	hsa00532	Chondroitin sulfate biosynthesis	6
	hsa00601	Glycosphingolipid biosynthesis	5
	hsa00511	Other glycan degradation	5
Metabolism of cofactor and vitamin and xenobiotics	hsa00760	Nicotinate and nicotinamide metabolism	6
	hsa00830	Retinol metabolism	10
	hsa00860	Porphyrin and chlorophyll metabolism	6
	hsa00980	Metabolism of xenobiotics by cytochrome P450	11
	hsa00982	Drug metabolism	11

(Continue next page)

(Continued)

Domain	KEGG ID	Pathway Name	Genes
Transcription, translation, folding sorting and degradation, replication and repair	hsa03020	RNA polymerase	5
	hsa03040	Spliceosome	21
	hsa03010	Ribosome	14
	hsa00970	Aminoacyl-tRNA biosynthesis	12
	hsa04120	Ubiquitin mediated proteolysis	17
	hsa03050	Proteasome	12
	hsa03018	RNA degradation	10
	hsa03440	Homologous recombination	5
Membrane transport and signal transduction and interaction	hsa02010	ABC transporters	8
	hsa04010	MAPK signaling pathway	18
	hsa04350	TGF-beta signaling pathway	6
	hsa04020	Calcium signaling pathway	7
	hsa04070	Phosphatidylinositol signaling system	10
	hsa04080	Neuroactive ligand-receptor interaction	9
	hsa04512	ECM-receptor interaction	8
Transport and catabolism and cell motility	hsa04144	Endocytosis	19
	hsa04142	Lysosome	22
	hsa04810	Regulation of actin cytoskeleton	17
Cell growth and death and cellular community	hsa04114	Oocyte meiosis	8
	hsa04115	p53 signaling pathway	7
	hsa04510	Focal adhesion	22
	hsa04530	Tight junction	12
Immune system	hsa04621	NOD-like receptor signaling pathway	5
	hsa04660	T cell receptor signaling pathway	6
	hsa04664	Fc epsilon RI signaling pathway	6
	hsa04666	Fc gamma R-mediated phagocytosis	8
	hsa04670	Leukocyte transendothelial migration	7
	hsa04062	Chemokine signaling pathway	12
Endocrine system	hsa04910	Insulin signaling pathway	13
	hsa04920	Adipocytokine signaling pathway	6
	hsa03320	PPAR signaling pathway	8
	hsa04912	GnRH signaling pathway	7
	hsa00150	Androgen and estrogen metabolism	8
	hsa04914	Progesterone-mediated oocyte maturation	7
Circulatory, excretory and nervous system	hsa04260	Cardiac muscle contraction	11
	hsa04270	Vascular smooth muscle contraction	10
	hsa04960	Aldosterone-regulated sodium reabsorption	6
	hsa04722	Neurotrophin signaling pathway	6
	hsa04360	Axon guidance	9

Table 5.4: List of significant pathways (1% FDR) represented by at least 5 genes associated with the component of the transcriptional response which is not affected by lipophilicity. Pathways are arranged according to the KEGG levels of organization. FDR values of each of the Kegg pathways are provided in the supplementary material.

Having performed a KEGG pathway analysis I then set to extend this comparison to a wider range of functional terms, using the gene ontology system. I identified 302 biological GO terms associated with compound lipophilicity and 153 associated with the component of the

transcriptional response which we identified to be free from any K_{ow} effect (1% FDR). 120 terms (78% of the lipophilicity independent signature) were in common while 182 (60% of the lipophilicity dependent signature) were specifically associated to the K_{ow} -affected molecular response and 33 (22% of the lipophilicity independent signature) specifically linked to the K_{ow} -free molecular response. The redundancy in the gene ontology terms was reduced by employing semantic similarity criteria as implemented in the REVIGO web-based application (table 5.5-5.6).

term_ID	description
GO:0000087	mitotic M phase
GO:0006512	ubiquitin cycle
GO:0007610	behavior
GO:0008283	cell proliferation
GO:0034622	cellular macromolecular complex assembly
GO:0048609	multicellular organismal reproductive process
GO:0051789	response to protein
GO:0048193	Golgi vesicle transport
GO:0016337	single organismal cell-cell adhesion
GO:0008284	positive regulation of cell proliferation
GO:0030029	actin filament-based process
GO:0051098	regulation of binding
GO:0040008	regulation of growth
GO:0007626	locomotory behavior
GO:0042592	homeostatic process
GO:0010035	response to inorganic substance
GO:0007059	chromosome segregation
GO:0032101	regulation of response to external stimulus
GO:0006952	defense response
GO:0009628	response to abiotic stimulus
GO:0022904	respiratory electron transport chain
GO:0051301	cell division
GO:0010564	regulation of cell cycle process
GO:0097190	apoptotic signaling pathway
GO:0006793	phosphorus metabolic process
GO:0051173	positive regulation of nitrogen compound metabolic process

term_ID	description
GO:0006366	transcription from RNA polymerase II promoter
GO:0030518	intracellular steroid hormone receptor signaling pathway
GO:0007586	digestion
GO:0051050	positive regulation of transport
GO:0030522	intracellular receptor signaling pathway
GO:0016477	cell migration
GO:0006936	muscle contraction
GO:0000165	MAPK cascade
GO:0048584	positive regulation of response to stimulus
GO:0051674	localization of cell
GO:0001501	skeletal system development
GO:0051240	positive regulation of multicellular organismal process
GO:0006417	regulation of translation
GO:0007186	G-protein coupled receptor signaling pathway
GO:0051592	response to calcium ion
GO:0034470	ncRNA processing
GO:0051094	positive regulation of developmental process
GO:0031279	regulation of cyclase activity
GO:0051339	regulation of lyase activity
GO:0030036	actin cytoskeleton organization
GO:0019933	cAMP-mediated signaling
GO:0016071	mRNA metabolic process
GO:0051101	regulation of DNA binding
GO:0000902	cell morphogenesis
GO:0006140	regulation of nucleotide metabolic process

Table 5.5: Biological processes (GO terms) found to be specifically enriched by lipophilicity dependent signatures. Redundant terms have been removed using REVIGO.

Interestingly, I identified cAMP-mediated signalling and G-protein coupled receptor signalling pathways to be specifically affected by lipophilicity dependent signatures. These pathways are related to each other and are important in activating internal signal transduction as a result of external stimuli. Particularly, G-protein coupled receptors, which are seven transmembrane receptors since they cross the membrane seven times, can activate signal transduction via cAMP signalling pathway. Being associated with membranes they are susceptible to lipophilicity effects. Consistent with this, MAPK cascade, as well as terms linked to transcription and translation activity, were affected. I also found terms linked to calcium ion response and homeostatic process to be affected by lipophilicity potentially due to membrane dysfunction. Other functions identified were associated with regulation of cell cycle processes, apoptosis signalling pathway and cytoskeleton organization.

term_ID	description
GO:0030001	metal ion transport
GO:0042493	response to drug
GO:0007017	microtubule-based process
GO:0016032	viral process
GO:0009314	response to radiation
GO:0006650	glycerophospholipid metabolic process
GO:0010608	posttranscriptional regulation of gene expression
GO:0000226	microtubule cytoskeleton organization
GO:0009968	negative regulation of signal transduction
GO:0001775	cell activation
GO:0042355	L-fucose catabolic process
GO:0006516	glycoprotein catabolic process
GO:0033674	positive regulation of kinase activity
GO:0006898	receptor-mediated endocytosis
GO:0051259	protein oligomerization
GO:0019637	organophosphate metabolic process
GO:0032990	cell part morphogenesis
GO:0007584	response to nutrient

Table 5.6: Biological processes (GO terms) found to be specifically enriched by lipophilicity independent signatures. Redundant terms have been removed using REVIGO.

Biological functions associated with lipophilicity are mainly linked to metabolic processes and, to some extent, to cellular organization.

5.4.3 Defining the functional profile of Juvenile and moulting hormones

In order to test whether the effects of EDs chemicals in *D. magna* are compatible with the disruption of the activity of endocrine hormones, I first defined transcriptional signatures linked to exposure to JH and MH. I identified a total of 861 (378 up and 483 down-regulated) and 509 (177 up and 332 down-regulated) differentially expressed genes (FDR<5%) for the MH (20-hydroxyecdysone) and the JH (methyl farnesoate) respectively (fig. 5.6). These genes actually represent a significant portion of the K_{ow} independent signature.

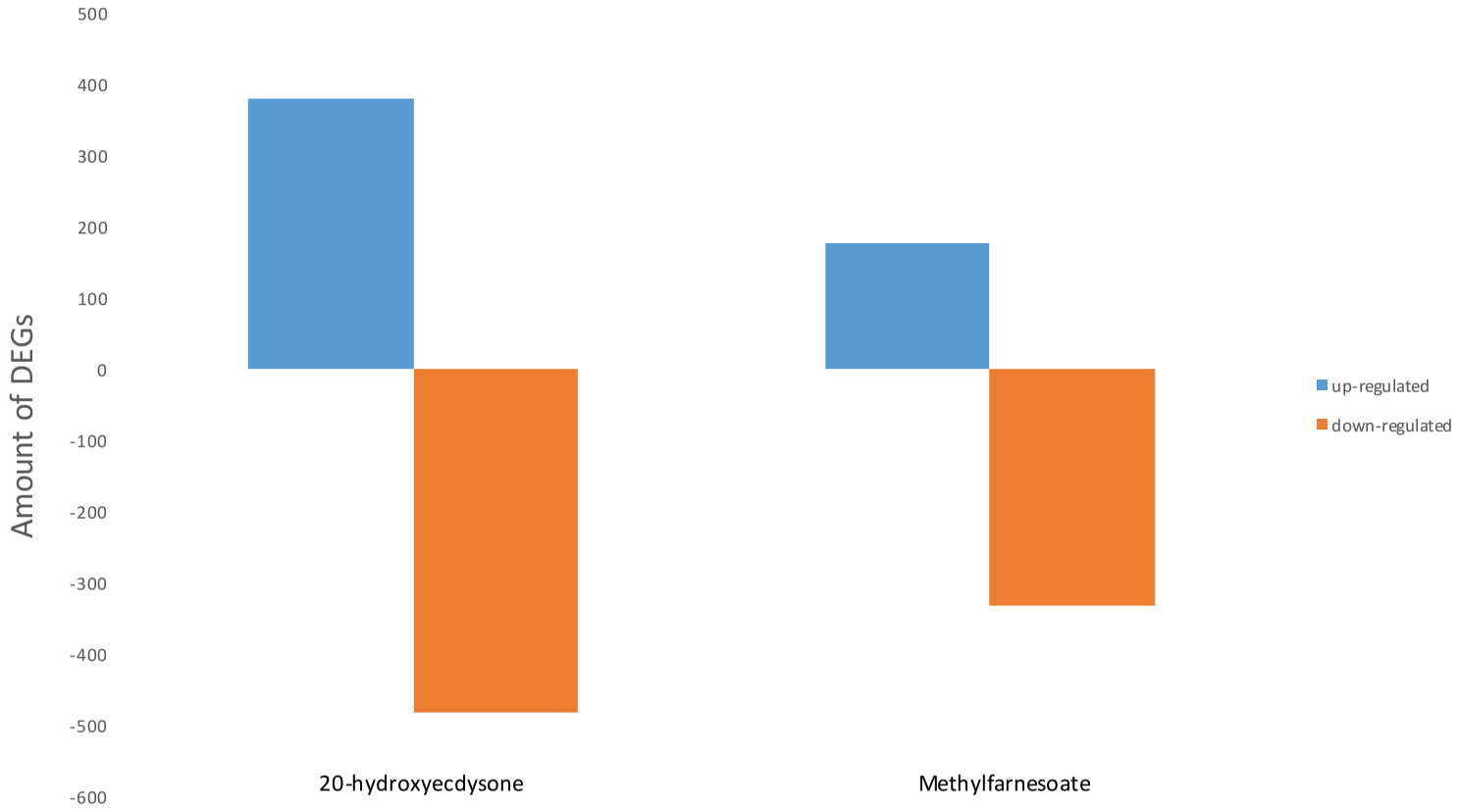


Figure 5.6: Differentially regulated genes associated with moulting and juvenile hormone exposures at 5% FDR. Blue and orange refers to up-regulated and down-regulated genes, respectively.

To characterize their functional profile, I identified KEGG pathways that were significantly enriched. The results of this analysis for the exposure to moulting hormone are represented in table 5.7 (up-regulated genes) and 5.8 (down-regulated genes).

Domain	KEGG ID	Pathway Name	Genes
Energy and amino acids metabolism	hsa00190	Oxidative phosphorylation	5
	hsa00280	Valine, leucine and isoleucine degradation	5
	hsa00330	Arginine and proline metabolism	5
Signal transduction	hsa04010	MAPK signaling pathway	7
Transport and catabolism	hsa04142	Lysosome	5
Cell motility and cellular community	hsa04810	Regulation of actin cytoskeleton	6
	hsa04510	Focal adhesion	7
	hsa04530	Tight junction	6
Immune system	hsa04666	Fc gamma R-mediated phagocytosis	5
	hsa04062	Chemokine signaling pathway	5

Table 5.7: List of KEGG pathways enriched for genes found to be up-regulated as a result of exposure to 20-hydroxyecdysone. FDR values of each of the Kegg pathways are provided in the supplementary material.

Domain	KEGG ID	Pathway Name	Genes
Carbohydrate and lipid metabolism	hsa00650	Butanoate metabolism	6
	hsa00071	Fatty acid metabolism	6
	hsa00140	Steroid hormone biosynthesis	6
	hsa00591	Linoleic acid metabolism	5
Nucleotide metabolism	hsa00230	Purine metabolism	6
Metabolism of amino acids and vitamins	hsa00380	Tryptophan metabolism	5
	hsa00830	Retinol metabolism	8
Xenobiotics degradation and metabolism	hsa00980	Metabolism of xenobiotics by cytochrome P450	8
	hsa00982	Drug metabolism - cytochrome P450	8
	hsa00983	Drug metabolism - other enzymes	6
Translation, folding, sorting and degradation	hsa03010	Ribosome	5
	hsa04120	Ubiquitin mediated proteolysis	7
	hsa03018	RNA degradation	7
Signal transduction	hsa04070	Phosphatidylinositol signaling system	5
ransport and catabolism	hsa04142	Lysosome	6
Cell growth and death	hsa04110	Cell cycle	6
Cellular community	hsa04510	Focal adhesion	8
Endocrine system	hsa04910	Insulin signaling pathway	5

Table 5.8: List of KEGG pathways enriched for genes found to be down-regulated as a result of exposure to 20-hydroxyecdysone. FDR values of each of the Kegg pathways are provided in the supplementary material.

Exposure to moulting hormone induces specific over-expression of genes in energy metabolism, represented by oxidative phosphorylation, as well as in cell motility and cellular motility and communication (actin cytoskeleton, tight junction and focal adhesion) and in functions related to immunity as chemokine signaling pathway and Fc Gamma R-mediated phagocytosis. Genes belonging to the oxidative phosphorylation pathway are ATP6V0D1 and ATP6V1D, subunits of vacuolar ATPase, and NDUFS8, NDUFA10 and UQCR10, which are subunits of the complex I and III involved in the electron transport chain. Cell cycle and insulin signaling pathway along with functions linked to translation (ribosome) were found to be down-regulated. Also, pathways underlying drug metabolism (metabolism of xenobiotics by cytochrome P450, drug metabolism – cytochrome P450 and drug metabolism – other enzymes) were down-regulated. Genes part of the cytochrome P450 have been shown to drive the biosynthesis of 20-hydroxyecdysone[342]. Interestingly, steroid hormone biosynthesis pathway was down-regulated. Specifically, genes belonging to the steroid hormone biosynthesis pathway includes CYP3A4, CYP3A5 and CYP3A7 which are members of the cytochrome P450 family involved in the steroid biosynthesis as well as xenobiotics metabolism, HSD17B12 which is the estradiol 17-beta-dehydrogenase 12 and is important for the conversion of estrone into estradiol, SULT1E1 that is a sulfotransferase enzyme that catalyse the sulphate conjugation of many hormones and UGT2B7 which is an UDP-glucuronosyltransferase playing an important role in regulating the level and the activity of estrogens.

The functional profile of the gene expression signature associated to JH also contained the terms steroid hormone biosynthesis and metabolism of xenobiotics by cytochrome P450 (table 5.9). Moreover, 83% of the steroid hormone biosynthesis genes hit by MH were also differentially regulated by JH (CYP3A5, CYP3A7, SULT1E1 and UGT2B7). The other gene specifically hit by JH is

HSD11B2, a hydroxysteroid 11-beta dehydrogenase 2. Similarly, purine metabolism was down-regulated in both the hormones (table 5.10).

Domain	KEGG ID	Pathway Name	Genes
Lipid and vitamins metabolism	hsa00140	Steroid hormone biosynthesis	5
	hsa00830	Retinol metabolism	5
Xenobiotics metabolism	hsa00983	Drug metabolism - other enzymes	5
Transport and catabolism	hsa04142	Lysosome	5

Table 5.9: List of KEGG pathways enriched for genes found to be up-regulated as a result of exposure to Methyl farnesoate. FDR values of each of the Kegg pathways are provided in the supplementary material.

Domain	KEGG ID	Pathway Name	Genes
Nucleotide metabolism	hsa00230	Purine metabolism	7
	hsa00240	Pyrimidine metabolism	7
Transcription, folding, sorting and degradation	hsa03040	Spliceosome	7
	hsa04120	Ubiquitin mediated proteolysis	5
	hsa03050	Proteasome	8
Transport and catabolism	hsa04142	Lysosome	7

Table 5.10: List of KEGG pathways enriched for genes found to be down-regulated as a result of exposure to Methyl farnesoate. FDR values of each of the Kegg pathways are provided in the supplementary material.

A comparison between the response to JH and MH (table 5.11) revealed that six pathways were modulated by both hormones. These are steroid hormone biosynthesis, retinol metabolism, purine metabolism, drug metabolism – other enzymes, ubiquitin-mediated proteolysis and lysosome. Interestingly, steroid hormone biosynthesis was up-regulated in juvenile hormone but down-regulated in moulting hormone suggesting an antagonizing activity. Retinol metabolism was found to be up-regulated in MH but down-regulated in JH. Purine metabolism was down-regulated in both but just 2 genes, CANT1 and HPRT1, were in common (we identified 6 and 7 genes to be differentially expressed in moulting and juvenile hormone respectively). “Drug

metabolism – other enzymes" pathway was found to up-regulated in the juvenile hormone but down-regulated in the moulting hormone.

Domain	Pathway	JH	MH
Carbohydrate and energy metabolism	Butanoate metabolism		DOWN
	Oxidative phosphorylation		UP
Lipid metabolism	Fatty acid metabolism		DOWN
	Steroid hormone biosynthesis	UP	DOWN
	Linoleic acid metabolism		DOWN
Amino acids and vitamin metabolism	Valine, leucine and isoleucine metabolism		UP
	Arginine and proline metabolism		UP
	Tryptophan		DOWN
	Retinol metabolism	UP	DOWN
Nucleotide metabolism	Purine metabolism	DOWN	DOWN
	Pirimidine metabolism	DOWN	
Xenobiotics biodegradation and biosynthesis	Metabolism of xenobiotics by cytochrome P450		DOWN
	Drug metabolism - cytochrome P450		DOWN
	Drug metabolism - other enzymes	UP	DOWN
Transcription and Translation	Spliceosome	DOWN	
	Ribosome		DOWN
Folding, sorting and degradation	Ubiquitin mediated proteolysis	DOWN	DOWN
	Proteasome	DOWN	
	RNA degradation		DOWN
Signal transduction	MAPK signaling pathway		UP
	Phosphatidylinositol signaling system		DOWN
Transport and catabolism	Lysosome	UP/DOWN	UP/DOWN
Cell motility and cellular community	Regulation of actin cytoskeleton		UP
	Focal adhesion		UP/DOWN
	Tight junction		UP
Immune system	Fc gamma R-mediated phagocytosis		UP
	Chemokine signaling pathway		UP
Endocrine system	Insulin signaling pathway		DOWN

Table 5.11: Pathway-level comparison between moulting and juvenile hormone. Pathways in red have been showed to be modulated by both the hormones.

I also investigate whether genes well-known to play a key role in the biosynthetic pathways of either juvenile or moulting hormone were differentially regulated. I retrieved 16 of these genes from KEGG (map00981). Surprisingly, none of them was found in our dataset. This is mainly due to the fact that KEGG lacks information about juvenile and moulting hormone pathways, especially for invertebrate species as crustaceans.

5.4.4 Identifying ED's with the ability to affect biological pathways underlying juvenile or moulting hormones activity

Having defined the transcriptional signatures linked to the two hormones, I set to test whether the EDs, as well as the additional non-ED chemicals in the panel, were able to mimic (or interfere with) the activity of either moulting or juvenile hormone.

I first performed an exploratory PCA analysis comparing all chemicals by their lipophilicity independent gene expression signature. (fig. 5.7).

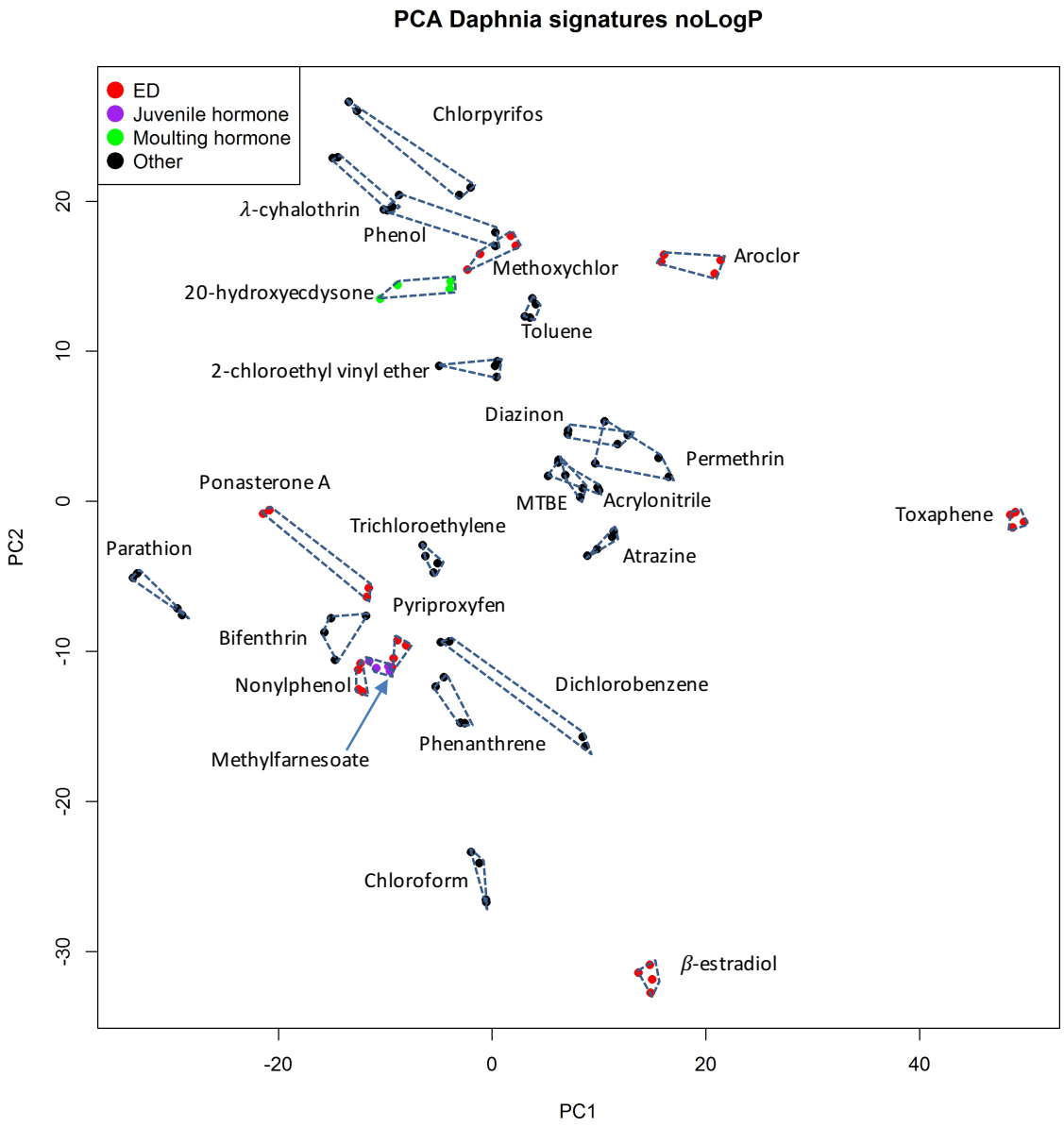


Figure 5.7: PCA analysis run on the lipohilicity-free response shows samples to cluster differently. PC1 and PC2 explain 15.7% and 13.4% of the variance respectively. The "noLogP" stated in the plot is to indicate the PCA have been run on the component of the transcritpional response which is free from lipophilicity effects.

I hypothesized that at least some of the ED's in our study may be able to mimic either juvenile or moulting hormone. The PCA suggested that this may indeed be the case and that some of the

chemicals may be mimicking either JH (nonylphenol, ponasteroneA and pyriproxyfen) or MH (methoxychlor and aroclor).

Since the first two components represent only the 29% of the variance, the PCA in figure 5.7 can only be considered indicative. In order to retain most of the variance and develop a more quantitative system of prediction, I computed the geometric distance between chemicals using the first 8 principal components which explain 80% of the variance. I finally calculated Euclidean distance between hormones of interest and the other compounds using the loadings of the first 8 principal components (table 5.12).

20-hydroxyecdysone	
Chemical	Euclidean dist.
β-estradiol	0.59
Phenol	0.60
Bifenthrin	0.63
λ-Cyhalothrin	0.65
Pyripoxyfen	0.70
Toluene	0.72
2-chloroethylvinyl ether	0.74
Acrylonitrile	0.74
Dichlorobenzene	0.75
MTBE	0.76
Aroclor1242	0.79
Methoxychlor	0.83
Methylfarnesoate	0.84
Atrazine	0.85
Diazinon	0.87
Nonylphenol	0.88
Chlorpyrifos	0.90
Parathion	0.91
Permethrin	0.93
Phenanthrene	0.93
Ponasterone A	0.96
Chloroform	0.96
Toxaphene	0.96
Trichloroethylene	1.02

Methylfarnesoate	
Chemical	Euclidean dist.
Aroclor1242	0.34
Diazinon	0.57
Chlorpyrifos	0.63
Atrazine	0.64
MTBE	0.66
Nonylphenol	0.67
Phenol	0.67
Acrylonitrile	0.71
Parathion	0.71
Ponasterone A	0.71
Bifenthrin	0.71
Dichlorobenzene	0.72
Pyripoxyfen	0.74
Trichloroethylene	0.76
β-estradiol	0.76
λ-Cyhalothrin	0.79
Methoxychlor	0.79
Chloroform	0.81
Toxaphene	0.81
20-hydroxyecdysone	0.84
2-chloroethylvinyl ether	0.85
Permethrin	0.94
Toluene	0.97
Phenanthrene	1.13

Table 5.12: Euclidean distances between hormones of interest and the other compounds using the principal components (PCs). Highlighted in red the ED's including the two hormones of interest.

I identified β-estradiol and aroclor1242 to potentially mimic moulting and juvenile hormone respectively. Moreover, I identified novel compounds to potentially affect moulting (bifenthrin and λ-cyhalothrin) and juvenile (Diazinon and Atrazine) hormone pathways. However, some of the known ED chemicals (e.g. Ponasterone A and toxaphene) show a very different profile to the reference hormones potentially indicating that they operate on a different mechanism.

5.4.5 Identification of pathways linked to moulting and juvenile hormone exposures that are affected by known and putative ED's

I set to run a comparative analysis using compounds functional profile to identify either Methyl farnesoate or ecdysone pathways that are perturbed by ED's. I first defined the transcriptional responses following exposure to each of the chemicals by selecting genes differentially expressed (5% FDR) (fig. 5.8).

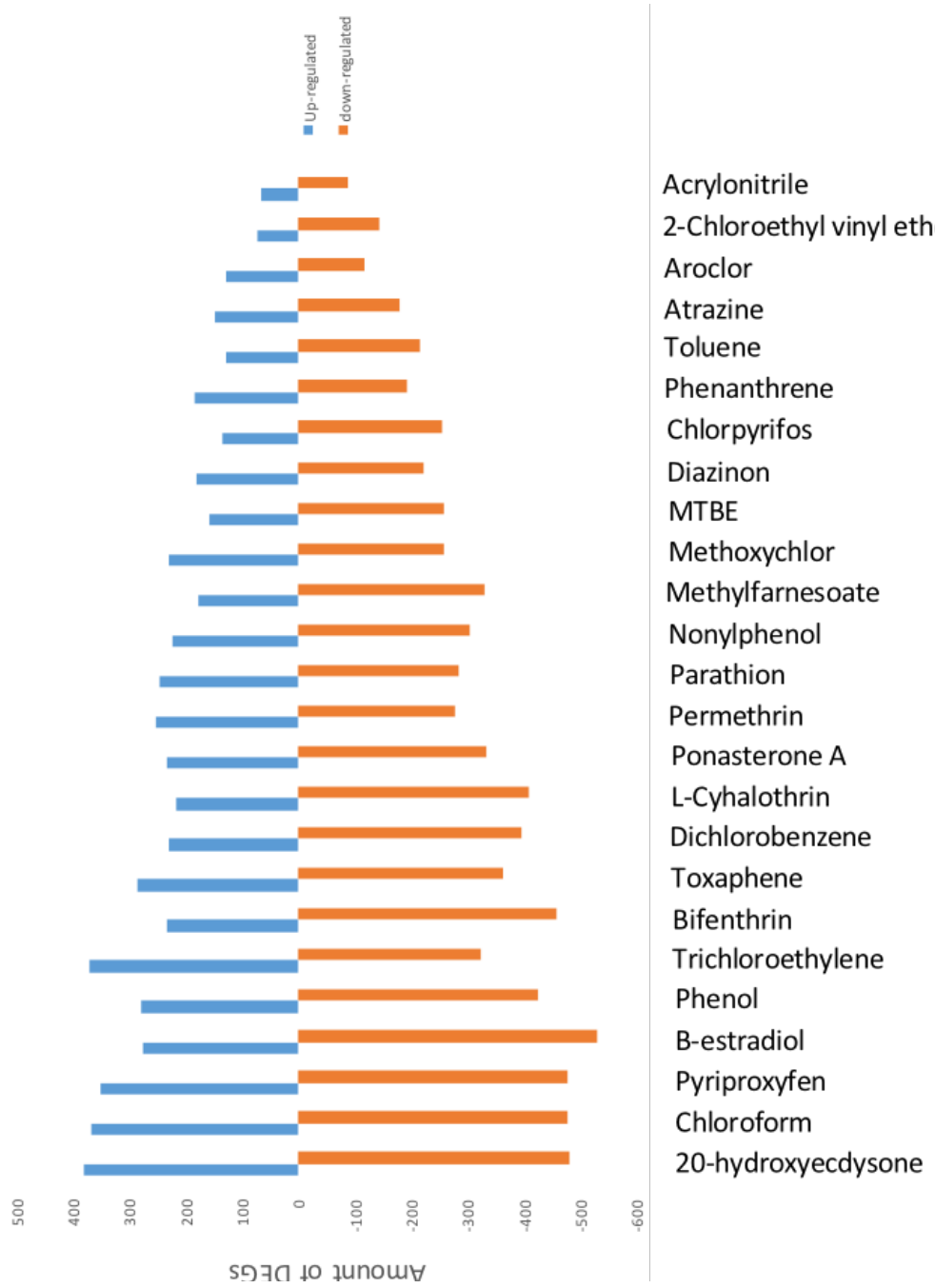


Figure 5.8: Genes found to be differentially expressed as result of exposure to each of the compound present in the study (5% FDR).

I then performed a functional enrichment analysis on each of the chemical exposures and identified either the KEGG pathways and the GO terms that overlapped with the 20-

hydroxyecdysone or Methyl farnesoate exposures. In order to reduce the redundancy in the gene ontology terms, I used semantic similarity criteria as implemented in the REVIGO web-based application.

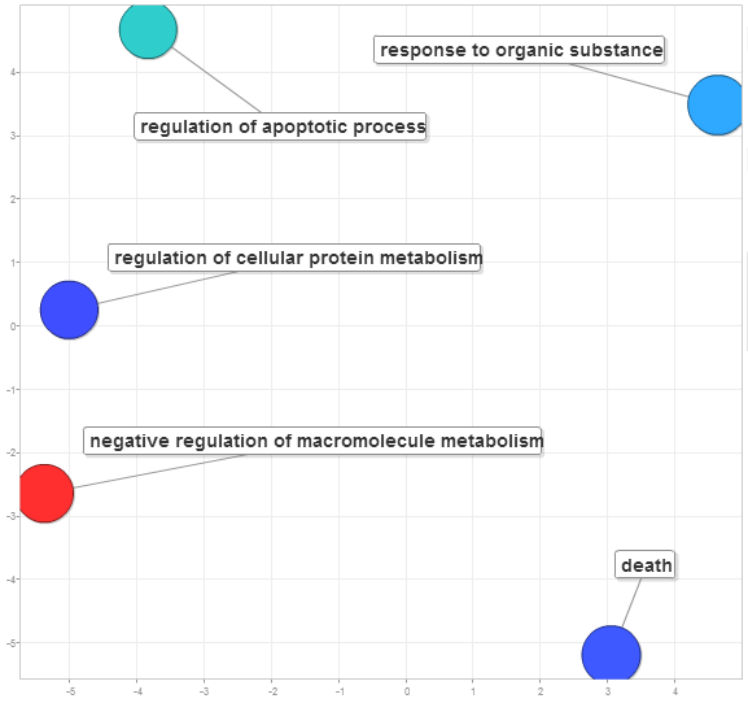
Aroclor was found to be the ED showing a more similar expression profile to the juvenile hormone. Comparative analysis at a pathway level shows nucleotide metabolism along with transcription and translation related functions to be in common (table 5.13).

Domain	KEGG ID	Pathway Name
Nucleotide metabolism	hsa00230	Purine metabolism
	hsa00240	Pyrimidine metabolism
Transcription and Translation	hsa03040	Spliceosome
	hsa00970	Aminoacyl-tRNA biosynthesis

Table 5.13: KEGG pathways enriched for genes found to be differentially expressed in both juvenile hormone and aroclor. Number og genes and FDR values of each of the Kegg pathways are provided in the supplementary material.

At a gene level, a comparative analysis identified five gene ontology terms (fig. 5.9).

Juvenile hormone – Aroclor similarities



term_ID	description
GO:0010033	response to organic substance
GO:0016265	death
GO:0032268	regulation of cellular protein metabolic process
GO:0042981	regulation of apoptotic process
GO:0010605	negative regulation of macromolecule metabolic process

Figure 5.9: Biological processes (GO terms) in common between methyl farnesoate and Aroclor. Functions are arranged within a semantic space which groups terms based on their similarity. The table reports the full list of overlapping GO terms.

Most of the shared transcriptional response is associated with cellular metabolic processes mainly related to proteins. Moreover, genes involved in the regulation of apoptotic processes and more in general with cell death appear to be affected. Genes shared by the two compounds that belonged to the "response to organic substance" included CAV1, which is caveolin-1 a tumor suppressor gene involved in cell cycle progression, DUOX1, a dual oxidase involved in antimicrobial defence at mucosal surfaces, HPRT1, a transferase playing a central role in the generation of purine nucleotide, SRRT, an RNA effector molecule involved in RNA-mediated gene silencing (RNAi) and promoting neural stem cell self-renewal, CASP3, which encode caspase 3 involved in apoptotic processes, HSPE1, a heat shock protein which functions as a chaperonin, GGH, a gamma-glutamyl hydrolase involved in pyrimidine metabolism, SLC34A2, a sodium-dependent phosphate transporter, and ID1, a DNA binding and transcriptional inhibitor involved in cell growth, senescence and differentiation.

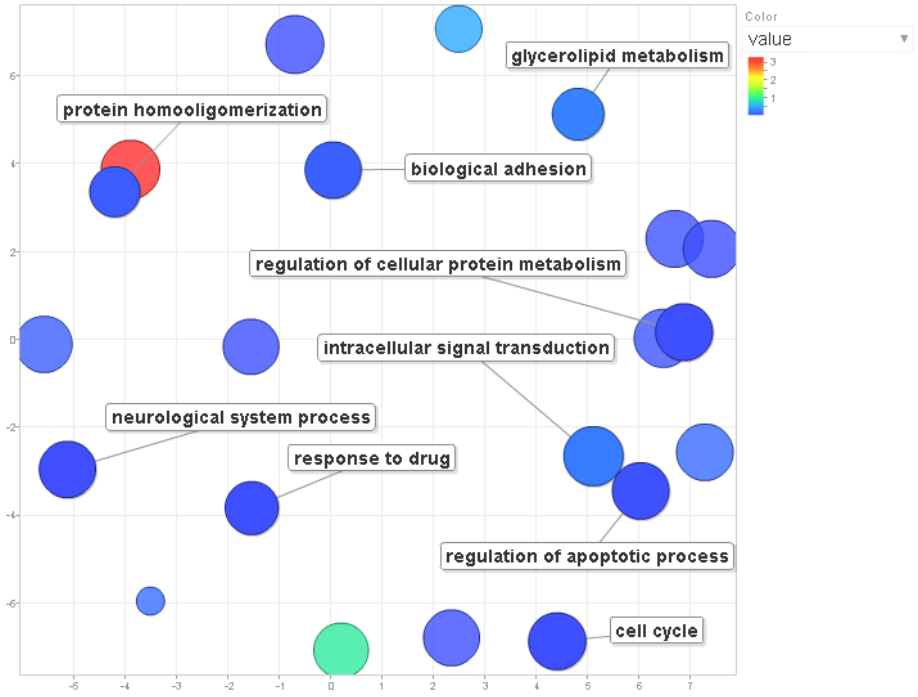
The response to β–estradiol is more similar to the moulting hormone and the functional comparative analysis confirm a high degree of overlap. The pathway level comparative analysis reveals these two chemicals to affect different levels of biological organization (table 5.14). Transcription and translation activity were both affected. As a consequence, carbohydrate, lipid, nucleotide and amino acids metabolism are differentially regulated. Interestingly, functions involved in the metabolism of xenobiotics, particularly cytochrome P450 related, are affected.

Domain	KEGG ID	Pathway Name
Carbohydrate metabolism	hsa00640	Propanoate metabolism
	hsa00650	Butanoate metabolism
Lipid metabolism	hsa00071	Fatty acid metabolism
	hsa00561	Glycerolipid metabolism
Nucleotide metabolism	hsa00230	Purine metabolism
	hsa00240	Pyrimidine metabolism
Amino acids metbolism	hsa00280	Valine, leucine and isoleucine degradation
	hsa00310	Lysine degradation
	hsa00330	Arginine and proline metabolism
	hsa00380	Tryptophan metabolism
	hsa00480	Glutathione metabolism
Glycan and vitamins metabolism	hsa00510	N-Glycan biosynthesis
	hsa00830	Retinol metabolism
Xenobiotics metabolism	hsa00980	Metabolism of xenobiotics by cytochrome P450
	hsa00982	Drug metabolism - cytochrome P450
	hsa00983	Drug metabolism - other enzymes
Transcription and translaton	hsa03040	Spliceosome
	hsa03010	Ribosome
	hsa00970	Aminoacyl-tRNA biosynthesis
Folding, sorting and degradation	hsa04120	Ubiquitin mediated proteolysis
	hsa03050	Proteasome
Signal transduction	hsa04010	MAPK signaling pathway
	hsa04070	Phosphatidylinositol signaling system
	hsa04512	ECM-receptor interaction
Transport and catabolism	hsa04144	Endocytosis
	hsa04142	Lysosome
Cell motility and cellular community	hsa04810	Regulation of actin cytoskeleton
	hsa04510	Focal adhesion
Immune system	hsa04666	Fc gamma R-mediated phagocytosis
	hsa04062	Chemokine signaling pathway
Endocrine system	hsa04910	Insulin signaling pathway
Circulatory system	hsa04260	Cardiac muscle contraction
	hsa04270	Vascular smooth muscle contraction
Development	hsa04360	Axon guidance

Table 5.14: KEGG pathways enriched for genes found to be differentially expressed in both moulting hormone and β-estradiol. Number og genes and FDR values of each of the Kegg pathways are provided in the supplementary material.

A gene-level comparative analysis was performed in order to identify more specific functions (fig. 5.10).

Moulting hormone – β-estradiol similarities



term_ID	description
GO:0000279	M phase
GO:0007155	cell adhesion
GO:0022610	biological adhesion
GO:0042493	response to drug
GO:0050877	neurological system process
GO:0030001	metal ion transport
GO:0016265	death
GO:0007049	cell cycle
GO:0051260	protein homooligomerization
GO:0032268	regulation of cellular protein metabolic process
GO:0008037	cell recognition
GO:0007017	microtubule-based process
GO:0046486	glycerolipid metabolic process
GO:0022402	cell cycle process
GO:0042981	regulation of apoptotic process
GO:0042127	regulation of cell proliferation
GO:0035556	intracellular signal transduction
GO:0043085	positive regulation of catalytic activity
GO:0043933	macromolecular complex subunit organization
GO:0019220	regulation of phosphate metabolic process
GO:0010605	negative regulation of macromolecule metabolic process

Figure 5.10: Biological processes (GO terms) in common between 20-hydroxyecdysone and β -estradiol. Functions are arranged within a semantic space which groups terms based on their similarity. The table reports the full list of overlapping GO terms.

Functions related to cell growth (cell cycle, M phase, regulation of cell proliferation, etc.) and death (regulation of apoptotic process) are affected by both the compounds. Interestingly, drug metabolism is enriched suggesting the potential involvement of cytochrome P450 which plays a central role in both xenobiotics metabolism and ecdysteroids biosynthesis. Genes involved in drug metabolism and shared by the two compounds include among the other, STAR, the steroidogenic acute regulatory protein involved in steroid hormone synthesis, SLC8A1 a sodium-calcium exchanger important for calcium homeostasis, ABCB1 and ABCA3, ABC transporters involved in drug resistance and PARK7, a positive regulator of androgen receptor-dependent transcription functioning as a sensor for oxidative stress. Other functions to be represented are linked to metabolic processes, neurological system process and cell adhesion.

Among the compounds not classified as endocrine disruptors, we identified Diazinon and Atrazine to show a functional profile which suggests they could potentially affect biological pathways underlying juvenile hormone activity. A pathway level comparative analysis shows Atrazine and Diazinon to affect steroid hormone pathways along with xenobiotics metabolism pathways similarly to juvenile hormone (table 5.15-5.16).

Domain	KEGG ID	Pathway Name
Lipid metabolism	hsa00140	Steroid hormone biosynthesis
Xenobiotics metabolism	hsa00980	Metabolism of xenobiotics by cytochrome P450
	hsa00982	Drug metabolism - cytochrome P450
	hsa00983	Drug metabolism - other enzymes
Folding, sorting and degradation	hsa04120	Ubiquitin mediated proteolysis
Transport and catabolism and cellular community	hsa04142	Lysosome
	hsa04510	Focal adhesion

Table 5.15: KEGG pathways enriched for genes found to be differentially expressed in both Juvenile hormone and Atrazine. Number og genes and FDR values of each of the Kegg pathways are provided in the supplementary material.

Domain	KEGG ID	Pathway Name
Lipid metabolism	hsa00140	Steroid hormone biosynthesis
nucleotide metabolism	hsa00230	Purine metabolism
	hsa00240	Pyrimidine metabolism
Xenobiotics metabolism	hsa00980	Metabolism of xenobiotics by cytochrome P450
	hsa00982	Drug metabolism - cytochrome P450
	hsa00983	Drug metabolism - other enzymes
Transcription and folding, sorting and degradation	hsa03040	Spliceosome
	hsa04120	Ubiquitin mediated proteolysis
	hsa03050	Proteasome
Transport and catabolism and cellular community	hsa04142	Lysosome
	hsa04510	Focal adhesion

Table 5.16: KEGG pathways enriched for genes found to be differentially expressed in both Juvenile hormone and Diazinon. Number og genes and FDR values of each of the Kegg pathways are provided in the supplementary material.

Gene-level comparative analysis confirm these findings (table 5.17). The functional overlap between Diazinon and Atrazine with Juvenile hormone is related to steroid metabolic process, the regulation of hormone levels and response to drugs. Other biological functions found to be affected are mainly related to cell growth and death processes as apoptosis and cell cycle and with the regulation of metabolic processes.

Atrazine	
term_ID	description
GO:0042493	response to drug
GO:0043085	positive regulation of catalytic activity
GO:0031145	anaphase-promoting complex-dependent proteasomal ubiquitin-dependent protein catabolic process
GO:0008202	steroid metabolic process
GO:0042592	homeostatic process
GO:0009719	response to endogenous stimulus
GO:0042981	regulation of apoptotic process
GO:0007049	cell cycle
GO:0006351	transcription, DNA-templated
GO:0010604	positive regulation of macromolecule metabolic process

Diazinon	
term_ID	description
GO:0042493	response to drug
GO:0043085	positive regulation of catalytic activity
GO:0043161	proteasome-mediated ubiquitin-dependent protein catabolic process
GO:0009719	response to endogenous stimulus
GO:0042981	regulation of apoptotic process
GO:0022402	cell cycle process
GO:0007049	cell cycle
GO:0006397	mRNA processing
GO:0008202	steroid metabolic process
GO:0031396	regulation of protein ubiquitination
GO:0010604	positive regulation of macromolecule metabolic process

Table 5.17: Functional overlap between Diazinon and Atrazine with Juvenile hormone.

Chemical showing a functional profile similar to the moulting hormone included bifenthrin and λ-cyhalothrin. Pathway level comparative analysis identified most of the transcriptional response shared by λ-cyhalothrin and MH to be associated with lipid, nucleotide and amino acids metabolism, translation-related functions, signal transduction and immune system (table 5.18).

Domain	KEGG ID	Pathway Name
Lipid metabolism	hsa00071	Fatty acid metabolism
Nucleotide metabolism	hsa00230	Purine metabolism
	hsa00240	Pyrimidine metabolism
Amino acids and glycan metabolism	hsa00280	Valine, leucine and isoleucine degradation
	hsa00480	Glutathione metabolism
	hsa00510	N-Glycan biosynthesis
Translation and folding, sorting and degradation	hsa03010	Ribosome
	hsa04120	Ubiquitin mediated proteolysis
	hsa03050	Proteasome
	hsa03018	RNA degradation
Signal transduction	hsa04010	MAPK signaling pathway
	hsa04070	Phosphatidylinositol signaling system
Transport and catabolism	hsa04144	Endocytosis
	hsa04142	Lysosome
Cellular community	hsa04510	Focal adhesion
Immune system	hsa04062	Chemokine signaling pathway

Table 5.18: KEGG pathways enriched for genes found to be differentially expressed in both moulting hormone and λ-cyhalothrin. Number og genes and FDR values of each of the Kegg pathways are provided in the supplementary material.

On the other hand, pathway-level functional overlap between MH and Bifenthrin shows a higher degree of similarities (table 5.19). Carbohydrate, lipid and amino acids metabolism are highly affected. Transcription and translation along with signal transduction are differentially regulated by both the compounds. Moreover, functions related to xenobiotics metabolism, particularly linked to the activity of cytochrome P450, are affected.

Domain	KEGG ID	Pathway Name
Carbohydrate metabolism	hsa00010	Glycolysis / Gluconeogenesis
	hsa00640	Propanoate metabolism
	hsa00650	Butanoate metabolism
Lipid metabolism	hsa00071	Fatty acid metabolism
	hsa00140	Steroid hormone biosynthesis
	hsa00561	Glycerolipid metabolism
	hsa00564	Glycerophospholipid metabolism
Nucleotide metabolism	hsa00230	Purine metabolism
	hsa00240	Pyrimidine metabolism
Amino acids metabolism	hsa00280	Valine, leucine and isoleucine degradation
	hsa00310	Lysine degradation
	hsa00330	Arginine and proline metabolism
	hsa00340	Histidine metabolism
	hsa00380	Tryptophan metabolism
	hsa00480	Glutathione metabolism
Vitamin metabolism	hsa00830	Retinol metabolism
Xenobiotics metabolism	hsa00980	Metabolism of xenobiotics by cytochrome P450
	hsa00982	Drug metabolism - cytochrome P450
Transcription and Translation	hsa03040	Spliceosome
	hsa03010	Ribosome
	hsa00970	Aminoacyl-tRNA biosynthesis
Folding, sorting and degradation	hsa04120	Ubiquitin mediated proteolysis
	hsa03050	Proteasome
Signal transduction	hsa04010	MAPK signaling pathway
	hsa04070	Phosphatidylinositol signaling system
	hsa04512	ECM-receptor interaction
Transport and catabolism	hsa04144	Endocytosis
	hsa04142	Lysosome
Cell motility and cellular community	hsa04810	Regulation of actin cytoskeleton
	hsa04510	Focal adhesion
Immune system	hsa04666	Fc gamma R-mediated phagocytosis
	hsa04062	Chemokine signaling pathway
Circulatory system and development	hsa04260	Cardiac muscle contraction
	hsa04360	Axon guidance

Table 5.19: KEGG pathways enriched for genes found to be differentially expressed in both moulting hormone and Bifenthrin. Number og genes and FDR values of each of the Kegg pathways are provided in the supplementary material.

Gene-level functional overlap between bifenthrin and λ-cyhalothrin with the moulting hormone reveal interesting insights and they are mainly related to cell growth and death processes as cell cycle and apoptosis, chromatin organization, neurological system process, ion homeostasis, cell

adhesion and regulation of metabolic processes (table 5.20). Interestingly, response to hormone and response to drug terms were both found for bifenthrin and λ-cyhalothrin.

λ-cyhalothrin	
term_ID	description
GO:0000279	M phase
GO:0007155	cell adhesion
GO:0016192	vesicle-mediated transport
GO:0022610	biological adhesion
GO:0042493	response to drug
GO:0050877	neurological system process
GO:0016265	death
GO:0007049	cell cycle
GO:0016032	viral process
GO:0032268	regulation of cellular protein metabolic process
GO:0009719	response to endogenous stimulus
GO:0000226	microtubule cytoskeleton organization
GO:0006650	glycerophospholipid metabolic process
GO:0007267	cell-cell signaling
GO:0022402	cell cycle process
GO:0006955	immune response
GO:0042981	regulation of apoptotic process
GO:0007268	synaptic transmission
GO:0006816	calcium ion transport
GO:0051259	protein oligomerization
GO:0043085	positive regulation of catalytic activity
GO:0010605	negative regulation of macromolecule metabolic process
GO:0009725	response to hormone
GO:0006355	regulation of transcription, DNA-templated
GO:0051276	chromosome organization
GO:0006325	chromatin organization
GO:0031396	regulation of protein ubiquitination
GO:0019226	transmission of nerve impulse
GO:0010033	response to organic substance
GO:0050890	cognition

Bifenthrin	
term_ID	description
GO:0000279	M phase
GO:0006898	receptor-mediated endocytosis
GO:0007155	cell adhesion
GO:0008219	cell death
GO:0022610	biological adhesion
GO:0042493	response to drug
GO:0050877	neurological system process
GO:0016265	death
GO:0006325	chromatin organization
GO:0032268	regulation of cellular protein metabolic process
GO:0044093	positive regulation of molecular function
GO:0007267	cell-cell signaling
GO:0022402	cell cycle process
GO:0007049	cell cycle
GO:0042127	regulation of cell proliferation
GO:0030001	metal ion transport
GO:0006351	transcription, DNA-templated
GO:0007268	synaptic transmission
GO:0007167	enzyme linked receptor protein signaling pathway
GO:0010605	negative regulation of macromolecule metabolic process
GO:0010604	positive regulation of macromolecule metabolic process
GO:0009725	response to hormone
GO:0006816	calcium ion transport
GO:0051276	chromosome organization
GO:0007601	visual perception
GO:0031396	regulation of protein ubiquitination
GO:0019226	transmission of nerve impulse
GO:0010033	response to organic substance
GO:0050953	sensory perception of light stimulus
GO:0050890	cognition

Table 5.20: Functional overlap between bifenthrin and λ-cyhalothrin with moulting hormone.

5.4.6 Prediction of endocrine disruption

I finally asked whether we could develop a biomarker signature to discriminate between endocrine disruptors and the other chemical compounds. Employing a statistical modeling technique that searches for smaller subsets of genes with the highest predictive power, I identified a model of endocrine disruption that included 25 genes (table 5.21).

Gene	Role	Frequency
KRTAP5-10	keratin associated protein 5-10	42
SLC51A	solute carrier family 51 alpha subunit	42
CHI3L2	chitinase 3 like 2	37
CYB5A	cytochrome b5 type A	35
PRKCE	protein kinase C epsilon	32
FAM162A	family with sequence similarity 162 member A	30
BTBD3	BTB domain containing 3	27
NKAP	NFKB activating protein	27
TMEM256	transmembrane protein 256	26
ADH5	alcohol dehydrogenase 5 (class III), chi polypeptide	24
OTOP2	otopetrin 2	24
RRP15	ribosomal RNA processing 15 homolog	24
SRL	sarcalumenin	24
ABHD16A	abhydrolase domain containing 16A	23
SLC4A7	solute carrier family 4 member 7	23
TUBB4B	tubulin beta 4B class IVb	23
RP2	retinitis pigmentosa 2	22
SDR42E1	short chain dehydrogenase/reductase family 42E, member 1	22
AGA	aspartylglucosaminidase	21
ATP1A1	ATPase Na+/K+ transporting subunit alpha 1	21
KMO	kynurenine 3-monooxygenase (kynurenine 3-hydroxylase)	21
BTBD1	BTB domain containing 1	20
FUT10	fucosyltransferase 10	20
NCAM2	neural cell adhesion molecule 2	20
PCCB	propionyl-CoA carboxylase beta subunit	20

Table 5.21: Genes representing the model we developed along with their frequencies which represent the times a gene was included into a model.

Those genes were used as input for a functional analysis in order to retrieve biological pathways they were involved in (table 5.22). Functions enriched are mainly linked to energy, lipid and amino acid metabolism and cytochrome P450 activity.

ADH5
hsa00982:Drug metabolism
hsa00010:Glycolysis / Gluconeogenesis
hsa00071:Fatty acid metabolism
hsa00350:Tyrosine metabolism
hsa00680:Methane metabolism
hsa00830:Retinol metabolism
hsa00980:Metabolism of xenobiotics by cytochrome P450
AGA
hsa04142:Lysosome
hsa00511:Other glycan degradation
ATP1A1
hsa04960:Aldosterone-regulated sodium reabsorption
hsa04260:Cardiac muscle contraction
KMO
hsa00380:Tryptophan metabolism
NCAM2
hsa05020:Prion diseases
hsa04514:Cell adhesion molecules (CAMs)
PCCB
hsa00640:Propanoate metabolism
hsa00280:Valine, leucine and isoleucine degradation
PRKCE
hsa04930:Type II diabetes mellitus
hsa04664:Fc epsilon RI signaling pathway
hsa04270:Vascular smooth muscle contraction
hsa04530:Tight junction
hsa04666:Fc gamma R-mediated phagocytosis

Table 5.22: Pathways enriched for genes identified by our model to discriminate ED's.

5.5 Discussion

In this chapter I have tested the hypothesis that chemicals that act as endocrine disruptors in vertebrates are affecting Daphnia via a mechanism that involve the juvenile and moulting hormones. Results are consistent with the original hypothesis. Moreover, the computational

pipeline I developed identified compounds which potentially could affect normal endocrine functions.

5.5.1 Basal toxicity is key toxicity mechanism in ecotoxicology

A key step in the analysis I have performed is to identify the component of the transcriptional response that is dependent on a basal toxicity mechanism. This has allowed me to isolate the transcriptional signatures that may be associated to a specific mechanism of action.

Basal toxicity or narcosis has classically been attributed to compounds acting by non-specifically disrupting the functioning of cell membranes. We have previously shown that a significant portion of the gene expression response to sub-lethal chemical exposure is linked to basal toxicity and that, at least in *Daphnia magna*, involves a transcriptional switch possibly driven by the release of calcium from the intracellular ER compartment[341]. This suggests that in the chronic exposure at sub-lethal doses that characterize the natural environment, basal toxicity may be a dominant factor affecting the biology of an organism.

The analysis described in this chapter focuses on K_{ow} to define a basal toxicity signature and indeed prove to be effective. However, my analysis has limitations. For example, I do not consider the fact that polar and non-polar Narcotic chemicals induce different responses in a fish acute toxicity test, which result in additional and more severe phenotypes following exposure to polar chemicals. Polar compounds disrupt cell membranes by binding the polar head groups of membrane lipids while non-polar compounds disrupt membrane integrity by a direct action with the hydrophobic membrane interior. At present, it is unclear what is the effect of this mechanism on gene expression and what role it may play in environmental toxicity. The relatively limited number of chemicals in this study prevented me to address whether polar and non-polar chemicals may induce a differential transcriptional response in *Daphnia magna*. I envisage that

by applying our approach to a sufficiently large number of chemicals may help to address this question.

5.5.2 Novel compounds candidates affecting endocrine system in *Daphnia magna*

Despite studies investigating endocrine disrupting chemical effects on invertebrates date back to almost two decades ago[343], mechanistic information of these effects has been achieved just recently. This has been possible by leveraging the power of omics technologies in the context of the AOP framework. In 2017, Song and collaborators developed the first invertebrate ED AOP where ecdysone receptor agonism leads to lethal molting disruption in Daphnia[28]. Moreover, they identified the release of chitobiase as a biomarker of molting disruption[344]. This is of paramount importance as it provides a valuable endpoint for the development of biological assays with the ability to assess endocrine disruption in aquatic invertebrate species.

The endocrine disruptors present in my chemical panel, with the exception of β–estradiol[345], have already been shown to affect either ecdysone or juvenile related pathways. Pyriproxyfen is a well-known analog of juvenile hormone[346]. Campos et al. demonstrated that Nonylphenol is able to deregulate ecdysone regulatory pathways in *D.magna*[347]. Methoxychlor and Aroclor have been demonstrated to be able to delay moulting in *D. magna*[348,349]. Toxaphene has previously been showed to play a role in sex determination in Daphnia promoting male offspring[350]. Ponasterone A has already been demonstrated to act as ecdysone analogue[351].

Here I demonstrated the effectiveness of a transcriptomic-based approach coupled with functional genomics for the identification of biological pathways underlying juvenile and moulting hormone pathways to be affected by environmentally relevant compounds.

Interestingly, my approach led to the identification of novel compounds with the potential to affect *D. magna* endocrine system by affecting either ecdysone or juvenile related molecular

pathways. Diazinon is an organophosphate insecticide whose toxicity is associated with the presence of strong oxidizing agents which can convert Diazinon into toxic phosphorus oxides as its active metabolite diaxozon[352]. Although the mode of action of Diazinon has been widely investigated as it inhibits the target site of the acetylcholinesterase, a few studies have tried to address the question whether this compound was able to affect endocrine system or reproduction in daphnids. Sanchez et al., exposed *Daphnia magna* to increasing concentration of Diazinon and they discovered that reproduction, in terms of total young for female and time to first reproduction, was significantly affected[353]. This finding is consistent with my hypothesis.

Atrazine is a worldwide used herbicide that was banned in the European Union in 2004 but it is still one of the most-widely used in many countries such as Australia and United States. In 1998 the US Environmental Protection Agency (EPA) started the Endocrine Disruptor Screening Program (EDSP) to determine whether a chemical has the potential to interact with the endocrine system and require more thorough testing. Results have shown Atrazine to have the potential to interact with the estrogen and androgen pathways in mammals and other wildlife however, EPA did not recommended further testing because Atrazine was not expected to impact current EPA-established regulatory endpoints for human health or ecological risk assessment[354]. The EPA's results have been criticized and the safety of Atrazine is still controversial as many studies have shown its ability to act as endocrine disruptor[355–357]. Interestingly, Palma et al. showed Atrazine to induce abnormalities during embryo development but they also discovered that Atrazine elicit its toxic effect without interfering with the ecdysteroid activity of Daphnia[358]. Even if a crosstalk between ecdysteroid and juvenile hormone pathways is present, the finding of Palma and collaborators does not exclude the possibility that Atrazine affect the juvenile hormone pathway and further laboratory validation is needed to prove it.

On the other hand, I identified two environmentally relevant compounds showing a transcriptional profile similar to the ecdysone hormone. Bifenthrin, is a pyrethroid insecticide with a well recognize toxicity towards aquatic organisms. Its neurotoxicity is associated with the ability of this chemical to alter the normal functioning of sodium channels affecting the membrane potential. The effect of bifenthrin on reproduction in *Daphnia magna* was already investigated[359]. *Daphnia magna* (F0 generation) was exposed to different bifenthrin concentrations. Reproduction (number of young for female) as well as length in offspring from the first brood (F1 1st) was significantly reduced. Offspring from the third brood (F1 3rd) were able to restore reproduction but length was still significantly affected. There are currently no studies investigating its effect on ecdysteroids at molecular level and our hypothesis, upon laboratory validation, may confirm that providing new insight about the molecular mechanism of Bifenthrin toxicity.

λ-cyhalothrin is a man-made mixture of iosomers of cyhalotrhin, an organic compound, that is used as a pesticide for its ability to remain effective for longer period of times. As all the other pyrethroids, it acts by disrupting the functioning of the nervous system. There are no studies that have focused on elucidating λ-cyhalothrin adverse outcomes on the endocrine system in *D. magna*.

My results suggest that Diazinon and Atrazine on one side and λ-cyhalothrin and bifenthrin on the other one, could potentially lead to reproduction and moulting impairment by affecting juvenile and moulting hormone pathways, respectively, and further laboratory validation is necessary to confirm my findings.

Chapter 6: Discussion

The overarching aim of my thesis was to prove the effectiveness of a computational biology approach to infer the effects of chemicals in organisms of environmental relevance. The results of my work have indeed proven that system biology and data driven science have great potential in investigating mechanisms of environmental toxicity, particularly in relation to the application of pathway inference for the identification of mechanistic biomarkers and *in silico* identification of compounds with potential toxicity. The purpose of this section is to discuss the general implications of the work in relation to the development of alternative risk assessment procedures.

6.1 Biomarkers of chemical toxicity

Risk assessment of chemical exposure to organisms and environment is particularly challenging due to the diversity in chemical MoA, the different organism's sensitivity and the fact that effects produced are usually measurable after prolonged exposures. Ideally, we would like to regulate the use of potentially dangerous chemicals before any effects are observed on individual species and ecosystems. Biomarkers, intended as early molecular changes linked to long term chronic toxicity can be indeed very useful to detect adverse effects long before the ecosystem is endangered. Ideally, biomarkers should be informative of specific toxicity mechanisms reflecting the progression between exposure and adverse effect [360,361]. A monitoring program built on such powerful diagnostic indicators could indeed be useful to develop appropriate policies for prevention[362,363]. Biomarkers need to be thoroughly validated in order to be useful but the "gold standards" for the process of validation still need to be defined[364].

Developing and validating biomarkers with the characteristics mentioned above is a challenging task. However, the advent of functional genomics and data driven science has created

considerable expectations[365–367]. The work described in chapter 2 represent a proof of concept study that address this important need. I successfully identified *in vitro* and *in vivo* molecular signatures triggered by chemical exposure which are diagnostic of chemical Mode of Action (MoA). Moreover, the approach developed is a promising tool for chemicals grouping and read-across and it can provide essential knowledge for the development of specific AOPs. By using the biomarkers we have developed on in vitro cell line assays, we may be able to predict toxic effects of related chemicals without the need of in vivo testing, reducing the number of animals, the cost and the time needed. Moreover, detailed understanding of toxic mechanisms and their downstream effects on the ecosystem would help scientist to make safe decision about the handling and treatment of municipal and industrial waste material. The computational approach I developed offer essential knowledge in this direction which can be used to improve current risk assessment procedures.

My proof of concept study is an important step but, further work is needed to translate putative biomarkers into a useful diagnostic tool for environmental risk assessment. Such a task is an integrative process that requires discovery, integration of genomic data with phenotypical endpoints, development of predictive statistical models and the ability to deliver such predictive information to authorities in a format easy to be implemented.

The greatest challenges to biomarker discovery are the platform and data analysis validation. The quick development of omics technologies posed a rush to identify new genomic biomarkers and analytical validation steps may have been superficially assessed. With regard to data analysis, on the other hand, discriminating between the signal associated with a specific phenotype and the background may be challenging for different reasons including the strength of the signal, the presence of confounding factors and sample heterogeneity. Moreover, the

large number of variables collected requires the development of statistical methods able to achieve the desired power. The MicroArray Quality Control Project (MAQC), a community-wide effort lead by the Food and Drug Administration (FDA), is currently tackling many of these issues for both microarray[368,369] and sequencing[370] platforms. With regards to microarray experiments, which technology has been extensively used throughout the whole thesis, is worth mentioning the development of a set of standards by the Microarray Gene Expression Data Society as requirement for the publication of microarray experiments[371]. Ecotoxicological studies have focused on the effects of acute exposures to single toxicants at high doses hence biomarkers contribute little to the prediction of direct consequences at a population, community and ecosystem level. With the development of ecotoxicogenomics[372], the scientific community is focusing on exposures to mixtures of chemical agents at lower doses, which better represent realistic environmental concentrations, able to trigger a threat to populations and communities. In order to provide a continuum of toxic responses from molecules to ecosystems effort has to be put towards linking molecular signatures with alterations in the genetic pool of the affected population.

6.2 Cell lines to inform about *in vivo* toxicity

The development of alternatives for toxicity testing is a major step that needs to be undertaken and a big effort has already been put to address this need. In the introduction, we extensively discussed the 3Rs framework for establishing these new alternatives for toxicity testing. In chapter 3, I presented a work based on these alternatives and we tried to address the question whether the transcriptional state of a cell line exposed to a given chemical could be used as a biosensor to predict toxicity in a more complex organism such as the zebrafish embryo. My work actually shows I can successfully map the transcriptional response of a cell line to an embryo.

This is of paramount importance and precisely fit the vision of the NC3R framework to reduce the need of animal testing and to develop *in silico* methods for toxicity predictions.

While my study explores an interesting objective, it remains a proof of concept and should be used as a pilot study to drive further effort in this direction. My design suffers of a limitation associated with the absence of time series following exposure. The response can change quite rapidly in the timeframe of the experiment and different time points may be informative of different biological processes and can have a different predictive power in inferring the adverse phenotypic outcome. Moreover, the small number of chemicals in the panel does not provide enough robustness for the development of any quantitative models (i.e. basal toxicity model). I envisage that a better experimental design where these limitations are taken into account would have the potential to strengthen this approach and provide essential knowledge towards the NC3R vision.

6.3 Omics and endocrine disruption

Omics technologies offer the ability to investigate the complexities of gene-environment interactions at a deeper level than more conventional methodologies by investigating biological pathways behaviour following chemical perturbations in a higher throughput fashion[373,374]. Transcriptome profiling can be seen as a tool to predict toxic outcomes of exposure to particular compounds. In combination with System Biology methods, it has the potential to investigate chemical MoA which are paramount in refining current risk assessment procedures. Canonical mechanism of action of endocrine disruptors refers to the ability to mimic or block the transcriptional activity elicited by the naturally circulating hormones by binding to steroid hormone receptors. However, in the last decades new MoA of endocrine disruptors have been described[375]. With regards to the MoA that affect reproduction, the approach I developed in

chapter 4 shows I can successfully predict chemicals able to alter ovary development through mechanisms which are different from the classical endocrine disruption. Moreover, it allows a rapid screening of chemical compounds with endocrine-related effects supporting the NC3R view of reducing animal testing. Finally, dynamical models taking into account different time-windows of a given biological process represent a promising tool in identifying molecular KEs in the context of the AOP framework which can in turn provide essential knowledge for the development of risk assessment procedures.

While the knowledge of endocrine disruption activity in higher vertebrate species is well documented, information on their effects in invertebrate organisms is still limited[336,376–378]. This is due to a lack of knowledge of invertebrate endocrinology[335,379] and the low number of invertebrate genomes sequenced[380,381]. However, studies aiming to investigate endocrine disrupting effects in invertebrate species are increasing[28,340,382]. In chapter 5, I developed an approach to compare gene expression signatures representing Daphnia response to both juvenile and moulting hormones, which are the hormones that drive organism's moulting, reproduction and growth, with gene expression signatures representing response to a panel of chemicals including endocrine disruptors. By using this approach, I proved that well known vertebrate EDs have the ability to interfere with the normal Daphnia hormonal processes. Moreover, my approach led to the identification of novel compounds with the ability to affect juvenile and ecdysone-related pathways supporting the use of Daphnia as a rapid screening tool of endocrine disruption. The approach I developed can be potentially applied to different species where the hormones driving fundamental biological processes have been characterized providing not just the ability to screening different compounds but also to investigate chemical MoA. Acquiring new knowledge of MoA or KEs of a given chemical compound will enable the possibility to refine or develop new AOPs that can be used to drive risk assessment procedures.

6.4 Challenges and future directions

The fast development of technologies able to measure molecular-level endpoints has raised an interest in the scientific community to evaluate the ability of such measurements to be incorporated into modern risk assessment procedures[383]. However, these molecular-level measurements are not yet integrated into existing health risk assessment procedures. The AOP framework, which we extensively discussed in the introduction, is an initiative to promote the use of mechanistic biological information in risk assessment. Development of AOPs is a key step to undertake in order to rapidly characterize the risk that a large number of chemicals pose to the environment and human health. The main challenge in developing these pathway-based frameworks is to define the key events and their relationships that take place between the molecular initiating event and the adverse outcome. In this context, high-content datasets can facilitate the understanding of molecular toxicity pathways and the effects occurring at the basal biological levels of organization[26,384,385] . Although omics data are able to provide knowledge of how chemical compounds interfere with biological process[386] , there is still no understanding of how these data could be incorporated in the context of risk assessment. To fill this gap, many challenges still need to be addressed.

One of these challenges is associated with the lack of standardised statistical approaches for interpreting differentially expressed genes, protein or metabolites as a result of chemical exposure or to provide quantitative information to support the link between omics data and phenotypic outcome[4,387–389]. System Biology methods, along with computational approaches need to be developed to link results coming from different biological assays and end-points across different levels of biological organization, especially at the population level, which would greatly facilitate extrapolation of *in vivo* impacts from *in vitro* effects. More specifically, development of more complex modelling techniques able to incorporate biological processes

(i.e. compensatory homeostatic mechanisms) is needed in order for the field to progress. Moreover, there is a lack of gene annotations especially regarding non-model species. Most of the time, researchers have to investigate transcriptome changes following either exposures or diseases on small subset of genes having just a partial overview of the molecular changes occurring in a given tissue or cell hence making biological interpretation extremely difficult. The lack of reference genomes for these species make the use of whole genome sequencing not ideal. The scientific community is already successfully tackling this issue by applying *de-novo* transcriptome assembly approaches[390,391]. Wang and Gribskov outlined four main reason for which *de-novo* transcriptome assembly is beneficial even when a reference genome is available[392].

The other great challenge researchers have to face in order to understand the toxic processes at the molecular level is the integration of transcriptomic, proteomic and metabolomics technologies as well as with toxicity data (i.e. toxicological endpoints). As previously highlighted, transcriptional responses may not accurately reflect important toxicologically relevant biological responses since often important changes in proteins and metabolites are not detectable by simply studying the levels of mRNA. Moreover, even if transcriptomics, proteomics and metabolomics target different molecules they may anyway regulate the same biological pathways at different levels hence the need to integrate these different molecular data. The ability to link these different data with toxicity data would allow the detection of more sensitive toxicity signatures and biomarkers that could potentially lead to improvement in environmental monitoring and risk assessment.

Another challenge is associated with the lack of comparative cross-species sensitivity data. The method of extrapolating data from standard laboratory model species to environmentally

relevant species, which is the standard in current risk assessment approaches, is characterised by a high degree of uncertainty due to different range of sensitivity to pollutants, different reproductive strategies and differences in physiology and life history[393,394]. Changes in molecular targets over the course of evolution represent a significant source of inter-species difference in relation to chemical sensitivity. Despite many biological systems (i.e. reproductive and detoxification) being conserved across species, it has been demonstrated that small structural variation, such as differences in amino acid residue of specific receptor ligand-binding pockets, can significantly affect chemical sensitivity[395,396]. Understanding the impact of these differences in sensitivity between species is a crucial step to successfully apply data extrapolation.

Lastly, the current application of *in vitro* assays cannot be used to fully replace *in vivo* approaches however, studies demonstrating the potential of *in vitro* datasets in predicting apical toxic outcomes and mechanism of toxicity are increasing[397]. The biggest challenge is representing whole organism metabolism and physiology in a single *in vitro* system which present different ADME properties. Variation between the nominal and the freely available concentration that will reach the site of toxic action may affect the toxicity outcome masking true toxicity effects. These variations are mainly associated with the chemical binding to either components of the media or in vitro system constituent (i.e. plastic). To improve the accuracy of dose-response relationships the understanding of these mechanisms is crucial. However, the results so far are encouraging as many alternative methods have been established and a reasonable number of them have been accepted for regulatory purposes either nationally (i.e. OECD) and internationally (i.e. EPA)[398].

Chapter 7: Conclusions

In this thesis, I aimed to demonstrate the potential of system biology and data-driven science in identifying novel mechanisms of action of environmental toxicity and to develop a set of

biomarkers able to improve risk assessment and environmental monitoring. Furthermore, I aimed to support the use of cell lines as alternative for toxicity testing. In the beginning of this thesis I outlined a number of aims that have come to the following conclusions based on my studies:

1. System biology approaches can be used to identify biomarkers of chemical toxicity (chapter 2)
2. A cell culture system can be informative of whole organism toxicity supporting the use of cell lines as alternative for toxicity testing (chapter 3)
3. Advanced computational techniques have the power to identify novel chemical compounds with endocrine disruption activity (chapter 4 and 5)

The work presented in this dissertation offers essential knowledge that can be used for risk assessment and environmental monitoring.

Bibliography

(1) Regenmortel, M. H. V. Van. Reductionism and complexity in molecular biology. *EMBO Reports.* **2004**, *5* (11), 1016–1020.

(2) Fang, F. C.; Casadevall, A. Reductionistic and holistic science. *Infection and Immunity.* **2011**, *79* (4), 1401–1404.

(3) Kitano, H. Systems biology: a brief overview. *Science (80-.).* **2002**, *295* (5560), 1662–1664.

(4) Garcia-Reyero, N.; Perkins, E. J. Systems biology: leading the revolution in ecotoxicology. *Environmental Toxicology and Chemistry.* **2011**, *30* (2), 265–273.

(5) Hood, L.; Perlmutter, R. M. The impact of systems approaches on biological problems in drug discovery. *Nature Biotechnology.* **2004**, *22* (10), 1215–1217.

(6) Karahalil, B. Overview of systems biology and omics technologies. *Current Medicinal Chemistry.* **2016**, *23* (37), 4221–4230.

(7) Schneider, M. V.; Orchard, S. Omics technologies, data and bioinformatics principles. *Methods in molecular biology (Clifton, N.J.)*; 2011; Vol. 719, pp 3–30.

(8) Martyniuk, C. J.; Prucha, M. S.; Doperalski, N. J.; Antczak, P.; Kroll, K. J.; Falciani, F.; Barber, D. S.; Denslow, N. D. Gene expression networks underlying ovarian development in wild largemouth bass (*Micropterus salmoides*). *PLoS One* **2013**, *8* (3), e59093.

(9) Chen, J.; Zhou, A.; Xie, S.; Wang, C.; Lv, Z.; Zou, J. Comparative proteomic identification of mature and immature sperm in the catfish *cranoglanis bouderius*. *PLoS One* **2016**, *11* (3), e0151254.

(10) Ali, S. E.; Farag, M. A.; Holvoet, P.; Hanafi, R. S.; Gad, M. Z. A comparative metabolomics approach reveals early biomarkers for metabolic response to acute myocardial infarction. *Scientific Reports.* **2016**, *6* (1), 36359.

(11) Davidsen, P. K.; Turan, N.; Egginton, S.; Falciani, F. Multilevel functional genomics data integration as a tool for understanding physiology: a network biology perspective. *Journal of Applied Physiology.* **2016**, *120* (3), 297–309.

(12) Ge, Y.; Wang, D.-Z.; Chiu, J.-F.; Cristobal, S.; Sheehan, D.; Silvestre, F.; Peng, X.; Li, H.; Gong, Z.; Hong Lam, S.; et al. Environmental OMICS: current status and future directions. *Journal of Integrated OMICS* **2013**, *3* (2), 75–87.

(13) Allen, T. E. H.; Goodman, J. M.; Gutsell, S.; Russell, P. J. A history of the molecular initiating event. *Chemical Research in Toxicology.* **2016**, *29* (12), 2060–2070.

(14) Beyer, J.; Petersen, K.; Song, Y.; Ruus, A.; Grung, M.; Bakke, T.; Tollefsen, K. E. Environmental risk assessment of combined effects in aquatic ecotoxicology: A discussion paper. *Marine Environmental Research.* **2014**, *96*, 81–91.

(15) Brander, S. M.; Jeffries, K. M.; Cole, B. J.; DeCourten, B. M.; White, J. W.; Hasenbein, S.; Fangue, N. A.; Connon, R. E. Transcriptomic changes underlie altered egg protein production and reduced fecundity in an estuarine model fish exposed to bifenthrin. *Aquatic Toxicology.* **2016**, *174*, 247–260.

(16) Uren Webster, T. M.; Santos, E. M. Global transcriptomic profiling demonstrates induction of oxidative stress and of compensatory cellular stress responses in brown trout exposed to glyphosate and Roundup. *BMC Genomics* **2015**, *16* (1), 32.

(17) Kumar, G.; Denslow, N. D. Gene expression profiling in fish toxicology: A review. *Reviews*

of environmental contamination and toxicology; 2016; Vol. 241, pp 1–38.

(18) Galland, C.; Dupuy, C.; Loizeau, V.; Danion, M.; Auffret, M.; Quiniou, L.; Laroche, J.; Pichereau, V. Proteomic analysis of the European flounder *Platichthys flesus* response to experimental PAH–PCB contamination. *Marine Pollution Bulletin.* **2015**, *95* (2), 646–657.

(19) Peng, H.-F.; Bao, X.-D.; Zhang, Y.; Huang, L.; Huang, H.-Q. Identification of differentially expressed proteins of brain tissue in response to methamidophos in flounder (*Paralichthys olivaceus*). *Fish & Shellfish Immunology.* **2015**, *44* (2), 555–565.

(20) Yan, L.; Gong, C.; Zhang, X.; Zhang, Q.; Zhao, M.; Wang, C. Perturbation of metabonome of embryo/larvae zebrafish after exposure to fipronil. *Environmental Toxicology and Pharmacology.* **2016**, *48*, 39–45.

(21) Qiao, Q.; Le Manach, S.; Sotton, B.; Huet, H.; Duvernois-Berthet, E.; Paris, A.; Duval, C.; Ponger, L.; Marie, A.; Blond, A.; et al. Deep sexual dimorphism in adult medaka fish liver highlighted by multi-omic approach. *Scientific Reports.* **2016**, *6*, 32459.

(22) Norris, J. L.; Farrow, M. A.; Gutierrez, D. B.; Palmer, L. D.; Muszynski, N.; Sherrod, S. D.; Pino, J. C.; Allen, J. L.; Spraggins, J. M.; Lubbock, A. L. R.; et al. Integrated, high-throughput, multiomics platform enables data-driven construction of cellular responses and reveals global drug mechanisms of action. *Journal of Proteome Research.* **2017**, *16* (3), 1364–1375.

(23) Wang, L.; Espinoza, H. M.; MacDonald, J. W.; Bammler, T. K.; Williams, C. R.; Yeh, A.; Louie, K. W.; Marcinek, D. J.; Gallagher, E. P. Olfactory transcriptional analysis of salmon exposed to mixtures of chlorpyrifos and malathion reveal novel molecular pathways of neurobehavioral injury. *Toxicological Sciences.* **2016**, *149* (1), 145–157.

(24) Oliveira, H. H. P.; Liebel, S.; Rossi, S. C.; Azevedo, A. C. B.; Barrera, E. A. L.; Garcia, J. R. E.; Grötzner, S. R.; Neto, F. F.; Randi, M. A. F.; Ribeiro, C. A. O. Mixtures of benzo(a)pyrene, dichlorodiphenyltrichloroethane and tributyltin are more toxic to neotropical fish Rhamdia quelen than isolated exposures. *Ecotoxicology and Environmental Safety.* **2015**, *122*, 106–115.

(25) OECD. OECD.org - OECD http://www.oecd.org/ (accessed Jul 4, 2017).

(26) Ankley, G. T.; Bennett, R. S.; Erickson, R. J.; Hoff, D. J.; Hornung, M. W.; Johnson, R. D.; Mount, D. R.; Nichols, J. W.; Russom, C. L.; Schmieder, P. K.; et al. Adverse outcome pathways: A conceptual framework to support ecotoxicology research and risk assessment. *Environmental Toxicology and Chemistry.* **2010**, *29* (3), 730–741.

(27) Willett, C.; Caverly Rae, J.; Goyak, K. O.; Minsavage, G.; Westmoreland, C.; Andersen, M.; Avigan, M.; Duché, D.; Harris, G.; Hartung, T.; et al. Building shared experience to advance practical application of pathway-based toxicology: liver toxicity mode-of-action. *ALTEX* **2014**, *31* (4), 500–519.

(28) Song, Y.; Villeneuve, D. L.; Toyota, K.; Iguchi, T.; Tollefsen, K. E. Ecdysone receptor agonism leading to lethal molting disruption in arthropods: review and adverse outcome pathway development. *Environmental Science & Technology.* **2017**, *51* (8), 4142–4157.

(29) Perkins, E. J.; Antczak, P.; Burgoon, L.; Falciani, F.; Garcia-Reyero, N.; Gutsell, S.; Hodges, G.; Kienzler, A.; Knapen, D.; McBride, M.; et al. Adverse outcome pathways for regulatory applications: examination of four case studies with different degrees of completeness and scientific confidence. *Toxicological Sciences.* **2015**, *148* (1), 14–25.

(30) Lee, J. W.; Won, E.-J.; Raisuddin, S.; Lee, J.-S. Significance of adverse outcome pathways

in biomarker-based environmental risk assessment in aquatic organisms. *Journal of Environmental Sciences.* **2015**, *35*, 115–127.

(31) Howe, K.; Clark, M. D.; Torroja, C. F.; Torrance, J.; Berthelot, C.; Muffato, M.; Collins, J. E.; Humphray, S.; McLaren, K.; Matthews, L.; et al. The zebrafish reference genome sequence and its relationship to the human genome. *Nature* **2013**, *496* (7446), 498–503.

(32) Bournele, D.; Beis, D. Zebrafish models of cardiovascular disease. *Heart Failure Reviews.* **2016**, *21* (6), 803–813.

(33) Corallo, D.; Candiani, S.; Ori, M.; Aveic, S.; Tonini, G. P. The zebrafish as a model for studying neuroblastoma. *Cancer Cell International.* **2016**, *16* (1), 82.

(34) Aspatwar, A.; Tolvanen, M. E. E.; Ojanen, M. J. T.; Barker, H. R.; Saralahti, A. K.; Bäuerlein, C. A.; Ortutay, C.; Pan, P.; Kuuslahti, M.; Parikka, M.; et al. Inactivation of ca10a and ca10b genes leads to abnormal embryonic development and alters movement pattern in zebrafish. *PLoS One* **2015**, *10* (7), e0134263.

(35) Ablain, J.; Zon, L. I. Tissue-specific gene targeting using CRISPR/Cas9. *Methods in Cell Biology.* **2016**, *135*, 189–202.

(36) Mesens, N.; Crawford, A. D.; Menke, A.; Hung, P. D.; Van Goethem, F.; Nuyts, R.; Hansen, E.; Wolterbeek, A.; Van Gompel, J.; De Witte, P.; et al. Are zebrafish larvae suitable for assessing the hepatotoxicity potential of drug candidates? *Journal of Applied Toxicology.* **2015**, *35* (9), 1017–1029.

(37) Truong, L.; Simonich, M. T.; Tanguay, R. L. Better, faster, cheaper: Getting the most out of high-throughput screening with zebrafish. *Methods in molecular biology (Clifton, N.J.)*; 2016; Vol. 1473, pp 89–98.

(38) Stegeman, J. J.; Goldstone, J. V.; Hahn, M. E. 10 – Perspectives on zebrafish as a model in environmental toxicology. *Fish Physiology*; 2010; Vol. 29, pp 367–439.

(39) Dai, Y.-J.; Jia, Y.-F.; Chen, N.; Bian, W.-P.; Li, Q.-K.; Ma, Y.-B.; Chen, Y.-L.; Pei, D.-S. Zebrafish as a model system to study toxicology. *Environmental Toxicology and Chemistry.* **2014**, *33* (1), 11–17.

(40) OECD. OECD Guidelines for the testing of chemicals. Section 2: effects on biotic systems test no. 236: Fish embryo acute toxicity (FET) test. OECD Publishing July 26, 2013.

(41) Lammer, E.; Carr, G. J.; Wendler, K.; Rawlings, J. M.; Belanger, S. E.; Braunbeck, T. Is the fish embryo toxicity test (FET) with the zebrafish (*Danio rerio*) a potential alternative for the fish acute toxicity test? *Comparative Biochemistry and Physiology Part C: Toxicology and Pharmacology.* **2009**, *149* (2), 196–209.

(42) Knöbel, M.; Busser, F. J. M.; Rico-Rico, A.; Kramer, N. I.; Hermens, J. L. M.; Hafner, C.; Tanneberger, K.; Schirmer, K.; Scholz, S. Predicting adult fish acute lethality with the zebrafish embryo: relevance of test duration, endpoints, compound properties, and exposure concentration analysis. *Environmental Science & Technology.* **2012**, *46* (17), 9690–9700.

(43) Belanger, S. E.; Rawlings, J. M.; Carr, G. J. Use of fish embryo toxicity tests for the prediction of acute fish toxicity to chemicals. *Environmental Toxicology and Chemistry.* **2013**, *32* (8), 1768–1783.

(44) Braunbeck, T.; Kais, B.; Lammer, E.; Otte, J.; Schneider, K.; Stengel, D.; Strecker, R. The fish embryo test (FET): origin, applications, and future. *Environmental Science and Pollution Research International.* **2015**, *22* (21), 16247–16261.

(45) Ebert, D. A Genome for the environment. *Science (80-.).* **2011**, *331* (6017).

(46) Plaistow, S. J.; Collin, H. Phenotypic integration plasticity in *Daphnia magna* : an integral facet of G × E interactions. *Journal of Evolutionary Biology.* **2014**, *27* (9), 1913–1920.

(47) Regier, J. C.; Shultz, J. W.; Kambic, R. E. Pancrustacean phylogeny: hexapods are terrestrial crustaceans and maxillopods are not monophyletic. *Proceedings of the Royal Society B: Biological Sciences.* **2005**, *272* (1561), 395–401.

(48) Nardi, F. Hexapod origins: monophyletic or paraphyletic? *Science (80-.).* **2003**, *299* (5614), 1887–1889.

(49) McTaggart, S. J.; Wilson, P. J.; Little, T. J. *Daphnia magna* shows reduced infection upon secondary exposure to a pathogen. *Biology Letters.* **2012**, *8* (6), 972–975.

(50) Kim, H. J.; Koedrith, P.; Seo, Y. R. Ecotoxicogenomic approaches for understanding molecular mechanisms of environmental chemical toxicity using aquatic invertebrate, daphnia model organism. *International Journal of Molecular Science.* **2015**, *16*, 12261–12287.

(51) Le, Q.-A. V.; Sekhon, S. S.; Lee, L.; Ko, J. H.; Min, J. Daphnia in water quality biomonitoring - “omic” approaches. *Toxicology and Environmental Health Science.* **2016**, *8* (1), 1–6.

(52) Daphnia Genomics Consortium https://wiki.cgb.indiana.edu/display/DGC/Home (accessed Apr 14, 2017).

(53) Colbourne, J. K.; Pfrender, M. E.; Gilbert, D.; Thomas, W. K.; Tucker, A.; Oakley, T. H.; Tokishita, S.; Aerts, A.; Arnold, G. J.; Basu, M. K.; et al. The ecoresponsive genome of

Daphnia pulex. *Science (80-.).* **2011**, *331* (6017), 555–561.

(54) Colbourne, J. K.; Singan, V. R.; Gilbert, D. G. wFleaBase: the Daphnia genome database. *BMC Bioinformatics* **2005**, *6* (1), 45.

(55) Orsini, L.; Gilbert, D.; Podicheti, R.; Jansen, M.; Brown, J. B.; Solari, O. S.; Spanier, K. I.; Colbourne, J. K.; Rush, D.; Decaestecker, E.; et al. *Daphnia magna* transcriptome by RNA-Seq across 12 environmental stressors. *Scientific Data.* **2016**, *3*, 160030.

(56) Richter, C. A.; Martyniuk, C. J.; Annis, M. L.; Brumbaugh, W. G.; Chasar, L. C.; Denslow, N. D.; Tillitt, D. E. Methylmercury-induced changes in gene transcription associated with neuroendocrine disruption in largemouth bass (*Micropterus salmoides*). *General and Comparative Endocrinology.* **2014**, *203*, 215–224.

(57) Mehinto, A. C.; Prucha, M. S.; Colli-Dula, R. C.; Kroll, K. J.; Lavelle, C. M.; Barber, D. S.; Vulpe, C. D.; Denslow, N. D. Gene networks and toxicity pathways induced by acute cadmium exposure in adult largemouth bass (*Micropterus salmoides*). *Aquatic Toxicology.* **2014**, *152*, 186–194.

(58) Chen, Y.-J.; Liu, Y.-J.; Yang, H.-J.; Yuan, Y.; Liu, F.-J.; Tian, L.-X.; Liang, G.-Y.; Yuan, R.-M. Effect of dietary oxidized fish oil on growth performance, body composition, antioxidant defence mechanism and liver histology of juvenile largemouth bass *Micropterus salmoides*. *Aquaculture Nutrition.* **2012**, *18* (3), 321–331.

(59) Li, S.; Liu, H.; Bai, J.; Zhu, X. Transcriptome assembly and identification of genes and SNPs associated with growth traits in largemouth bass (*Micropterus salmoides*). *Genetica.* **2017**, *145* (2), 175–187.

(60) Byadgi, O.; Chen, C.-W.; Wang, P.-C.; Tsai, M.-A.; Chen, S.-C. De Novo transcriptome

analysis of differential functional gene expression in Largemouth bass (*Micropterus salmoides*) after challenge with *Nocardia seriolae*. *International Journal of Molecular Science.* **2016**, *17* (8), 1315.

(61) Dominguez, G. A.; Quattro, J. M.; Denslow, N. D.; Kroll, K. J.; Prucha, M. S.; Porak, W. F.; Grier, H. J.; Sabo-Attwood, T. L. Identification and transcriptional modulation of the Largemouth bass, *Micropterus salmoides*, vitellogenin receptor during oocyte development by insulin and sex steroids. *Biology of Reproduction.* **2012**, *87* (3), 67.

(62) Martyniuk, C. J.; Doperalski, N. J.; Prucha, M. S.; Zhang, J.-L.; Kroll, K. J.; Conrow, R.; Barber, D. S.; Denslow, N. D. High contaminant loads in Lake Apopka's riparian wetland disrupt gene networks involved in reproduction and immune function in largemouth bass. *Comparative Biochemistry and Physiology Part D: Genomics and Proteomics.* **2016**, *19*, 140–150.

(63) Collí-Dulá, R. C.; Martyniuk, C. J.; Streets, S.; Denslow, N. D.; Lehr, R. Molecular impacts of perfluorinated chemicals (PFASs) in the liver and testis of male largemouth bass (*Micropterus salmoides*) in Minnesota Lakes. *Comparative Biochemistry and Physiology Part D: Genomics and Proteomics.* **2016**, *19*, 129–139.

(64) Williams, T. D.; Davies, I. M.; Wu, H.; Diab, A. M.; Webster, L.; Viant, M. R.; Chipman, J. K.; Leaver, M. J.; George, S. G.; Moffat, C. F.; et al. Molecular responses of European flounder (*Platichthys flesus*) chronically exposed to contaminated estuarine sediments. *Chemosphere.* **2014**, *108*, 152–158.

(65) Petersen, A. M.; Dillon, D.; Bernhardt, R. R.; Torunsky, R.; Postlethwait, J. H.; von Hippel, F. A.; Loren Buck, C.; Cresko, W. A. Perchlorate disrupts embryonic androgen synthesis

and reproductive development in threespine stickleback without changing whole-body levels of thyroid hormone. *General and Comparative Endocrinolology.* **2015**, *210*, 130–144.

(66) Cavallin, J. E.; Jensen, K. M.; Kahl, M. D.; Villeneuve, D. L.; Lee, K. E.; Schroeder, A. L.; Mayasich, J.; Eid, E. P.; Nelson, K. R.; Milsk, R. Y.; et al. Pathway-based approaches for assessment of real-time exposure to an estrogenic wastewater treatment plant effluent on fathead minnow reproduction. *Environmental Toxicology and Chemistry.* **2016**, *35* (3), 702–716.

(67) Council, N. R. *Toxicity Testing in the 21st Century*; National Academies Press: Washington, D.C., 2007.

(68) ECHA https://echa.europa.eu/home (accessed Apr 15, 2017).

(69) Singh, J. The national centre for the replacement, refinement, and reduction of animals in research. *Journal of Pharmacology and Pharmacotherapeutics.* **2012**, *3* (1), 87–89.

(70) Russell, W. M. S.; Burch, R. L. *The principles of human experimental technique*; Universities Federation for Animal Welfare: Wheathampstead, 1992.

(71) Sewell, F.; Edwards, J.; Prior, H.; Robinson, S. Opportunities to apply the 3Rs in safety assessment programs. *ILAR Journal.* **2016**, *57* (2), 234–245.

(72) Goh, J.-Y.; Weaver, R. J.; Dixon, L.; Platt, N. J.; Roberts, R. A. Development and use of in vitro alternatives to animal testing by the pharmaceutical industry 1980–2013. *Toxicology Research.* **2015**, *4* (5), 1297–1307.

(73) Lee, L. E. J.; Dayeh, V. R.; Schirmer, K.; Bols, N. C. Applications and potential uses of fish

gill cell lines: examples with RTgill-W1. *In Vitro Cellular and Developmental Biology - Animal.* **2009**, *45* (3–4), 127–134.

(74) Tanneberger, K.; Knöbel, M.; Busser, F. J. M.; Sinnige, T. L.; Hermens, J. L. M.; Schirmer, K. Predicting fish acute toxicity using a fish gill cell line-based toxicity assay. *Environmental Science and Technology.* **2013**, *47* (2), 1110–1119.

(75) Yazdani, M.; Andresen, A. M. S.; Gjøen, T. Short-term effect of bisphenol-a on oxidative stress responses in Atlantic salmon kidney cell line: a transcriptional study. *Toxicology Mechanisms and Methods.* **2016**, *26* (4), 295–300.

(76) Scholz, S.; Sela, E.; Blaha, L.; Braunbeck, T.; Galay-Burgos, M.; García-Franco, M.; Guinea, J.; Klüver, N.; Schirmer, K.; Tanneberger, K.; et al. A European perspective on alternatives to animal testing for environmental hazard identification and risk assessment. *Regulatory Toxicology and Pharmacology.* **2013**, *67* (3), 506–530.

(77) Bols, N. C.; Barlian, A.; Chirino-Trejo, M.; Caldwell, S. J.; Goegan, P.; Lee, L. E. J. Development of a cell line from primary cultures of rainbow trout, *Oncorhynchus mykiss* (Walbaum), gills. *Journal of Fish Diseases.* **1994**, *17* (6), 601–611.

(78) Sangion, A.; Gramatica, P. Hazard of pharmaceuticals for aquatic environment: Prioritization by structural approaches and prediction of ecotoxicity. *Environmental International.* **2016**, *95*, 131–143.

(79) Raies, A. B.; Bajic, V. B. In silico toxicology: computational methods for the prediction of chemical toxicity. *WIREs Computational Molecular Science.* **2016**, *6*, 147–172.

(80) Burden, N.; Benstead, R.; Clook, M.; Doyle, I.; Edwards, P.; Maynard, S. K.; Ryder, K.; Sheahan, D.; Whale, G.; van Egmond, R.; et al. Advancing the 3Rs in regulatory

ecotoxicology: A pragmatic cross-sector approach. *Integrated Environmental Assessment and Management.* **2016**, *12* (3), 417–421.

(81) Weaver, J. M. J.; Ross-Innes, C. S.; Fitzgerald, R. C. The "–omics" revolution and oesophageal adenocarcinoma. *Nature Reviews Gastroenterology & Hepatology.* **2013**, *11* (1), 19–27.

(82) Lockhart, D. J.; Dong, H.; Byrne, M. C.; Follettie, M. T.; Gallo, M. V; Chee, M. S.; Mittmann, M.; Wang, C.; Kobayashi, M.; Horton, H.; et al. Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nature Biotechnology.* **1996**, *14* (13), 1675–1680.

(83) Wang, Z.; Gerstein, M.; Snyder, M. RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics.* **2009**, *10* (1), 57–63.

(84) Amin, S.; Prentis, P. J.; Gilding, E. K.; Pavasovic, A.; Hou, R.; Bao, Z.; Wang, S.; Su, H.; Li, Y.; Du, H.; et al. Assembly and annotation of a non-model gastropod (*Nerita melanotragus*) transcriptome: a comparison of De novo assemblers. *BMC Research Notes.* **2014**, *7* (1), 488.

(85) Melwani, A. R.; Thompson, E. L.; Raftos, D. A. Differential proteomic response of Sydney rock oysters (*Saccostrea glomerata*) to prolonged environmental stress. *Aquatic Toxicology.* **2016**, *173*, 53–62.

(86) Trapp, J.; Armengaud, J.; Pible, O.; Gaillard, J.-C.; Abbaci, K.; Habtoul, Y.; Chaumot, A.; Geffard, O. Proteomic investigation of male *Gammarus fossarum* , a freshwater crustacean, in response to endocrine disruptors. *Journal of Proteome Research.* **2015**, *14* (1), 292–303.

(87) Bouatra, S.; Aziat, F.; Mandal, R.; Guo, A. C.; Wilson, M. R.; Knox, C.; Bjorndahl, T. C.; Krishnamurthy, R.; Saleem, F.; Liu, P.; et al. The human urine metabolome. *PLoS One* **2013**, *8* (9), e73076.

(88) Jackowski, K. Multinuclear NMR spectroscopy in the gas phase. *Journal of Molecular Structure.* **2006**, *786* (2–3), 215–219.

(89) Tripathi, P.; Somashekar, B. S.; Ponnusamy, M.; Gursky, A.; Dailey, S.; Kunju, P.; Lee, C. T.; Chinnaiyan, A. M.; Rajendiran, T. M.; Ramamoorthy, A. HR-MAS NMR tissue metabolomic signatures cross-validated by mass spectrometry distinguish bladder cancer from benign disease. *Journal of Proteome Research.* **2013**, *12* (7), 3519–3528.

(90) Atiqullah, M.; Anantawaraskul, S.; Emwas, A.-H. M.; Al-Harthi, M. A.; Hussain, I.; Ul-Hamid, A.; Hossaen, A. Effects of supported (n BuCp) 2 ZrCl 2 catalyst active-center distribution on ethylene–1-hexene copolymer backbone heterogeneity and thermal behaviors. *Industrial & Engineering Chemistry Research.* **2013**, *52* (27), 9359–9373.

(91) Abriata, L. A.; Zaballa, M.-E.; Berry, R. E.; Yang, F.; Zhang, H.; Walker, F. A.; Vila, A. J. Electron spin density on the axial His ligand of high-spin and low-spin nitrophorin 2 probed by heteronuclear NMR spectroscopy. *Inorganic Chemistry.* **2013**, *52* (3), 1285–1295.

(92) Emwas, A.-H. M. The strengths and weaknesses of NMR spectroscopy and mass spectrometry with particular focus on metabolomics research. *Methods in molecular biology (Clifton, N.J.)*; 2015; Vol. 1277, pp 161–193.

(93) Wu, H.; Xu, L.; Yu, D.; Ji, C. Differential metabolic responses in three life stages of mussels Mytilus galloprovincialis exposed to cadmium. *Ecotoxicology* **2017**, *26* (1), 74–

80.

(94) Taylor, N. S.; Kirwan, J. A.; Johnson, C.; Yan, N. D.; Viant, M. R.; Gunn, J. M.; McGeer, J. C. Predicting chronic copper and nickel reproductive toxicity to *Daphnia pulex-pulicaria* from whole-animal metabolic profiles. *Environmental Pollution.* **2016**, *212*, 325–329.

(95) Chomczynski, P.; Sacchi, N. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Analytical Biochemistry.* **1987**, *162* (1), 156–159.

(96) Chomczynski, P.; Sacchi, N. The single-step method of RNA isolation by acid guanidinium thiocyanate–phenol–chloroform extraction: twenty-something years on. *Nature Protocols.* **2006**, *1* (2), 581–585.

(97) Sykacek, P.; Kreil, D. P.; Meadows, L. A.; Auburn, R. P.; Fischer, B.; Russell, S.; Micklem, G. The impact of quantitative optimization of hybridization conditions on gene expression analysis. *BMC Bioinformatics* **2011**, *12*, 73.

(98) Long, J. The Science Creative Quarterly https://www.scq.ubc.ca/image-bank/ (accessed Apr 16, 2017).

(99) Wu, Z. A review of statistical methods for preprocessing oligonucleotide microarrays. *Statistcal Methods in Medical Research.* **2009**, *18* (6), 533–541.

(100) Cleveland, W. S. Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistic Association.* **1979**, *74* (368), 829–836.

(101) Bolstad, B. M.; Irizarry, R. A.; Astrand, M.; Speed, T. P. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias.

Bioinformatics **2003**, *19* (2), 185–193.

(102) Miller, J. A.; Cai, C.; Langfelder, P.; Geschwind, D. H.; Kurian, S. M.; Salomon, D. R.; Horvath, S. Strategies for aggregating gene expression data: the collapseRows R function. *BMC Bioinformatics* **2011**, *12*, 322.

(103) Shrestha, A.; Koju, R. P.; Beresford, S. A. A.; Gary Chan, K. C.; Karmacharya, B. M.; Fitzpatrick, A. L. Food patterns measured by principal component analysis and obesity in the Nepalese adult. *Heart Asia.* **2016**, *8* (1), 46–53.

(104) Smith, A. D. A. C.; Emmett, P. M.; Newby, P. K.; Northstone, K. Dietary patterns obtained through principal components analysis: the effect of input variable quantification. *British journal of Nutrition.* **2013**, *109* (10), 1881–1891.

(105) Pascoal, C.; Oliveira, M. R.; Pacheco, A.; Valadas, R. Detection of outliers using robust principal component analysis: A simulation study; Springer Berlin Heidelberg, 2010; pp 499–507.

(106) Brown, R. J. C.; Brown, A. S. Principal component analysis as an outlier detection tool for polycyclic aromatic hydrocarbon concentrations in ambient air. *Water, Air, Soil Pollution.* **2012**, *223* (7), 3807–3816.

(107) Ringnér, M. What is principal component analysis? *Nature Biotechnology.* **2008**, *26* (3), 303–304.

(108) Lin, I.-H.; Chen, D.-T.; Chang, Y.-F.; Lee, Y.-L.; Su, C.-H.; Cheng, C.; Tsai, Y.-C.; Ng, S.-C.; Chen, H.-T.; Lee, M.-C.; et al. Hierarchical clustering of breast cancer methylomes revealed differentially methylated and expressed breast cancer genes. *PLoS One* **2015**, *10* (2), e0118453.

(109) Alon, U.; Barkai, N.; Notterman, D. A.; Gish, K.; Ybarra, S.; Mack, D.; Levine, A. J. Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. *Proceedings of the National Academy of Sciences of the United States of America.* **1999**, *96* (12), 6745–6750.

(110) Herrero, J.; Valencia, A.; Dopazo, J. A hierarchical unsupervised growing neural network for clustering gene expression patterns. *Bioinformatics* **2001**, *17* (2), 126–136.

(111) Kohonen, T. *Self-Organization and Associative Memory*; Springer Berlin Heidelberg, 1989.

(112) Fritzke, B. Growing cell structures a self-organizing network for unsupervised and supervised learning. *Neural Networks.* **1994**, *7* (9), 1441–1460.

(113) Macqueen, J. Some methods for classification and analysis of multivariate observations. *Proceedings of the Fifth Berkeley Symposium of Mathematical Statistics and Probability.* **1967**, *1*, Statistics, 281--297.

(114) Belacel, N.; Wang, Q. (Christa); Cuperlovic-Culf, M. Clustering methods for microarray gene expression data. *Omics: A Journal of Integrative Biology.* **2006**, *10* (4), 507–531.

(115) Almudevar, A. Multiple hypothesis testing: a methodological overview. *Methods in Molecular Biology.* **2013**, *972*, 37–55.

(116) Benjamini, Y.; Hochberg, Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B.* **1995**, *57* (1), 289–300.

(117) Tusher, V. G.; Tibshirani, R.; Chu, G. Significance analysis of microarrays applied to the

ionizing radiation response. *Proceedings of the Natlional Academy of Sciences of the United States of America.* **2001**, *98* (9), 5116–5121.

(118) Hosack, D. A.; Dennis, G.; Sherman, B. T.; Lane, H.; Lempicki, R. A.; Lane, H. C.; Lempicki, R. A.; Steenbeke, T.; Khazanie, P.; Gupta, N. DAVID: Database for annotation, visualization, and integrated discovery. *Genome Biology.* **2003**, *4* (6), P4.

(119) Al-Shahrour, F.; Diaz-Uriarte, R.; Dopazo, J. FatiGO: a web tool for finding significant associations of Gene Ontology terms with groups of genes. *Bioinformatics* **2004**, *20* (4), 578–580.

(120) Reimand, J.; Kull, M.; Peterson, H.; Hansen, J.; Vilo, J. g:Profiler--a web-based toolset for functional profiling of gene lists from large-scale experiments. *Nucleic Acids Research.* **2007**, *35* (Web Server issue), W193-200.

(121) Subramanian, A.; Tamayo, P.; Mootha, V. K.; Mukherjee, S.; Ebert, B. L.; Gillette, M. A.; Paulovich, A.; Pomeroy, S. L.; Golub, T. R.; Lander, E. S.; et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America.* **2005**, *102* (43), 15545–15550.

(122) Supek, F.; Bošnjak, M.; Škunca, N.; Šmuc, T. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One* **2011**, *6* (7), e21800.

(123) Wang, L.; Song, X.; Gu, L.; Li, X.; Cao, S.; Chu, C.; Cui, X.; Chen, X.; Cao, X. NOT2 proteins promote polymerase II-dependent transcription and interact with multiple MicroRNA biogenesis factors in Arabidopsis. *The Plant Cell* **2013**, *25* (2), 715–727.

(124) Riising, E. M.; Helin, K. A new role for the polycomb group protein Ezh1 in promoting

transcription. *Molecular Cell.* **2012**, *45* (2), 145–146.

(125) Villaverde, A. F.; Banga, J. R. Reverse engineering and identification in systems biology: strategies, perspectives and challenges. *Journal of the Royal Society Interface.* **2014**, *11* (91), 20130505.

(126) Stuart, J. M.; Segal, E.; Koller, D.; Kim, S. K. A gene-coexpression network for global discovery of conserved genetic modules. *Science (80-.).* **2003**, *302* (5643), 249–255.

(127) Shannon, C. E. A Mathematical Theory of Communication. *The bell System Technical Journal.* **1948**, *27*, 379–423.

(128) Margolin, A. a; Nemenman, I.; Basso, K.; Wiggins, C.; Stolovitzky, G.; Dalla Favera, R.; Califano, A. ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinformatics* **2006**, *7 Suppl 1*, S7.

(129) Cover, T. M.; Thomas, J. A. *Elements of information theory*; Wiley, 1991.

(130) Zoppoli, P.; Morganella, S.; Ceccarelli, M. TimeDelay-ARACNE: Reverse engineering of gene networks from time-course data by an information theoretic approach. *BMC Bioinformatics* **2010**, *11*, 154.

(131) Cantone, I.; Marucci, L.; Iorio, F.; Ricci, M. A.; Belcastro, V.; Bansal, M.; Santini, S.; di Bernardo, M.; di Bernardo, D.; Cosma, M. P.; et al. A yeast synthetic network for in vivo assessment of reverse-engineering and modeling approaches. *Cell.* **2009**, *137* (1), 172–181.

(132) Villaverde, A. F.; Ross, J.; Mor?n, F.; Banga, J. R.; Belcastro, V. MIDER: Network inference with mutual information distance and entropy reduction. *PLoS One* **2014**, *9* (5), e96732.

(133) Kontkanen, P.; Kontkanen, P.; Myllymäki, P.; Silander, T.; Tirri, H.; Grunwald, P. Comparing predictive inference methods for discrete domains. *Proceedings of the sixth International Workshop on Artificial Intelligence and Statistics.* **1997**, 311--318.

(134) Myllymaki, P.; Silander, T.; Tirri, H.; Uronen, P. B-COURSE: A web-based tool for bayesian and causal data analysis. *International Journal of Artificial Intelligence Tools.* **2002**, *11* (3), 369–387.

(135) Marcot, B. G.; Holthausen, R. S.; Raphael, M. G.; Rowland, M. M.; Wisdom, M. J. Using Bayesian belief networks to evaluate fish and wildlife population viability under land management alternatives from an environmental impact statement. *Forest Ecology and Management.* **2001**, *153* (1–3), 29–42.

(136) Yu, J.; Smith, V. A.; Wang, P. P.; Hartemink, A. J.; Jarvis, E. D. Advances to Bayesian network inference for generating causal networks from observational biological data. *Bioinformatics.* **2004**, *20* (18), 3594–3603.

(137) Dondelinger, F.; Husmeier, D.; Lèbre, S. Dynamic Bayesian networks in molecular plant science: inferring gene regulatory networks from multiple gene expression time series. *Euphytica.* **2012**, *183* (3), 361–377.

(138) Bonneau, R. Learning biological networks: from modules to dynamics. *Nature Chemical Biology.* **2008**, *4* (11), 658–664.

(139) Jesmin; Rashid, M. S.; Jamil, H.; Hontecillas, R.; Bassaganya-Riera, J.; King, H.; Aubert, R.; Herman, W.; Flegal, K.; Graubard, B.; et al. Gene regulatory network reveals oxidative stress as the underlying molecular mechanism of type 2 diabetes and hypertension. *BMC Medical Genomics.* **2010**, *3* (1), 45.

(140) Lin, E.; Lane, H.-Y. Machine learning and systems genomics approaches for multi-omics data. *Biomarker Research.* **2017**, *5*, 2.

(141) Altman, N. S. An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician.* **1992**, *46* (3), 175–185.

(142) Alpaydın, E. Introduction to machine learning third editionn. *MIT Press* **2014**.

(143) Ju, H.; Brasier, A. R. Variable selection methods for developing a biomarker panel for prediction of dengue hemorrhagic fever. *BMC Research Notes.* **2013**, *6*, 365.

(144) Fernandez, M.; Sarai, A.; Ahmad, S. Recognition of drug-target interaction patterns using genetic algorithm-optimized Bayesian-regularized neural networks and support vector machines. *2009 IEEE International Conference on Systems, Man and Cybernetics*; IEEE, 2009; pp 498–503.

(145) Tibshirani, R.; Hastie, T.; Narasimhan, B.; Chu, G. Diagnosis of multiple cancer types by shrunken centroids of gene expression. *Proceedings of the National Academy of Sciences of the United States of America.* **2002**, *99* (10), 6567–6572.

(146) Trägårdh, M.; Chappell, M. J.; Ahnmark, A.; Lindén, D.; Evans, N. D.; Gennemark, P. Input estimation for drug discovery using optimal control and Markov chain Monte Carlo approaches. *Journal of Pharmacokinetics and Pharmacodynamics.* **2016**, *43* (2), 207–221.

(147) Beheshti, I.; Demirel, H.; Matsuda, H.; Alzheimer's Disease Neuroimaging Initiative. Classification of Alzheimer's disease and prediction of mild cognitive impairment-to-Alzheimer's conversion from structural magnetic resource imaging using feature ranking and a genetic algorithm. *Computers in Biology and Medicine.* **2017**, *83*, 109–119.

(148) Sainani, K. L. Multivariate regression: The pitfalls of automated variable selection. *PM&R* **2013**, *5*, 791–794.

(149) Trevino, V.; Falciani, F. GALGO: an R package for multivariate variable selection using genetic algorithms. *Bioinformatics* **2006**, *22* (9), 1154–1156.

(150) Falciani, F.; Diab, A. M.; Sabine, V.; Williams, T. D.; Ortega, F.; George, S. G.; Chipman, J. K. Hepatic transcriptomic profiles of European flounder (*Platichthys flesus*) from field sites and computational approaches to predict site from stress gene responses following exposure to model toxicants. *Aquatic Toxicology.* **2008**, *90* (2), 92–101.

(151) European Commission. European Commission's Joint Research Centre (JRC) https://ec.europa.eu/jrc/en (accessed Jul 4, 2017).

(152) US Army. U.S. Army Engineer Research and Development Center http://www.erdc.usace.army.mil/ (accessed Jul 4, 2017).

(153) OECD; EPA. Aopwiki https://aopwiki.org/ (accessed Jul 4, 2017).

(154) OECD; EPA. Effectopedia | The online encyclopedia of adverse outcome pathwasy https://www.effectopedia.org/ (accessed Jul 4, 2017).

(155) Wittwehr, C.; Aladjov, H.; Ankley, G.; Byrne, H. J.; de Knecht, J.; Heinzle, E.; Klambauer, G.; Landesmann, B.; Luijten, M.; MacKay, C.; et al. How adverse outcome pathways can aid the development and use of computational prediction models for regulatory toxicology. *Toxicological Sciences.* **2017**, *155* (2), 326–336.

(156) Altenburger, R.; Ait-Aissa, S.; Antczak, P.; Backhaus, T.; Barcel?, D.; Seiler, T.-B.; Brion, F.; Busch, W.; Chipman, K.; de Alda, M. L.; et al. Future water quality monitoring - Adapting

tools to deal with mixtures of pollutants in water resource management. *Science of the Total Environment.* **2015**, *512–513*, 540–551.

(157) Knight, K. Implementing the 3Rs: improving experimental approaches in animal biology. *Journal of Experimental Biology.* **2016**, *219* (16), 2414–2415.

(158) Stokes, W. Animals and the 3Rs in toxicology research and testing: The way forward. *Human & Experimental Toxicology.* **2015**, *34* (12), 1297–1303.

(159) Basketter, D. A.; Clewell, H.; Kimber, I.; Rossi, A.; Blaauboer, B.; Burrier, R.; Daneshian, M.; Eskes, C.; Goldberg, A.; Hasiwa, N.; et al. A roadmap for the development of alternative (non-animal) methods for systemic toxicity testing - t4 report*. *ALTEX* **2012**, *29* (1), 3–91.

(160) Olson, H.; Betton, G.; Robinson, D.; Thomas, K.; Monro, A.; Kolaja, G.; Lilly, P.; Sanders, J.; Sipes, G.; Bracken, W.; et al. Concordance of the toxicity of pharmaceuticals in humans and in animals. *Regulatory Toxicology and. Pharmacology.* **2000**, *32* (1), 56–67.

(161) Hoehme, S.; Brulport, M.; Bauer, A.; Bedawy, E.; Schormann, W.; Hermes, M.; Puppe, V.; Gebhardt, R.; Zellmer, S.; Schwarz, M.; et al. Prediction and validation of cell alignment along microvessels as order principle to restore tissue architecture in liver regeneration. *Proceedings of the National Academy of Sciences of the United States of America.* **2010**, *107* (23), 10371–10376.

(162) Schliess, F.; Hoehme, S.; Henkel, S. G.; Ghallab, A.; Driesch, D.; Böttger, J.; Guthke, R.; Pfaff, M.; Hengstler, J. G.; Gebhardt, R.; et al. Integrated metabolic spatial-temporal model for the prediction of ammonia detoxification during liver damage and regeneration. *Hepatology.* **2014**, *60* (6), 2040–2051.

(163) Ghallab, A. Letter to the editor: In vitro test systems and their limitations. *EXCLI Journal Journal.* **2013**, *1212*, 1024–1026.

(164) Kilemade, M.; Lyons-Alcantara, M.; Rose, T.; Fitzgerald, R.; Mothersill, C. Rainbow trout primary epidermal cell proliferation as an indicator of aquatic toxicity: an in vitro/in vivo exposure comparison. *Aquatic Toxicology.* **2002**, *60* (1), 43–59.

(165) Tanneberger, K.; Rico-Rico, A.; Kramer, N. I.; Busser, F. J. M.; Hermens, J. L. M.; Schirmer, K. Effects of solvents and dosing procedure on chemical toxicity in cell-based in vitro assays. *Environmental Science & Technology.* **2010**, *44* (12), 4775–4781.

(166) Schirmer, K. Proposal to improve vertebrate cell cultures to establish them as substitutes for the regulatory testing of chemicals and effluents using fish. *Toxicology.* **2006**, *224* (3), 163–183.

(167) Bury, N. R.; Schnell, S.; Hogstrand, C. Gill cell culture systems as models for aquatic environmental monitoring. *Journal of Experimental Biology.* **2014**, *217* (Pt 5), 639–650.

(168) Klapper, W.; Heidorn, K.; Kühne, K.; Parwaresch, R.; Krupp, G. Telomerase activity in "immortal" fish. *FEBS Letters.* **1998**, *434* (3), 409–412.

(169) Evans, D. H.; Piermarini, P. M.; Choe, K. P. The multifunctional fish gill: Dominant site of gas exchange, osmoregulation, acid-base regulation, and excretion of nitrogenous waste. *Physiology Reviews.* **2005**, *85* (1), 97–177.

(170) Organisation for Economic Co-operation and Development. *Test No. 203 : fish, acute toxicity test*; OECD Publishing, 1992.

(171) Nagel, R. DarT: The embryo test with the Zebrafish *Danio rerio*--a general model in

ecotoxicology and toxicology. *ALTEX.* **2002**, *19 Suppl 1*, 38–48.

(172) Dahm, R.; Geisler, R. Learning from small fry: The zebrafish as a genetic model organism for aquaculture fish species. *Marine Biotechnology.* **2006**, *8* (4), 329–345.

(173) Aleström, P.; Holter, J. L.; Nourizadeh-Lillabadi, R. Zebrafish in functional genomics and aquatic biomedicine. *Trends in Biotechnology.* **2006**, *24* (1), 15–21.

(174) Parng, C. In vivo zebrafish assays for toxicity testing. *Current Opinions in Drug Discovery & Development.* **2005**, *8* (1), 100–106.

(175) Braunbeck, T.; Boettcher, M.; Hollert, H.; Kosmehl, T.; Lammer, E.; Leist, E.; Rudolf, M.; Seitz, N. Towards an alternative for the acute fish LC(50) test in chemical assessment: the fish embryo toxicity test goes multi-species -- an update. *ALTEX.* **2005**, *22* (2), 87–102.

(176) Henn, K.; Braunbeck, T. Dechorionation as a tool to improve the fish embryo toxicity test (FET) with the zebrafish (*Danio rerio*). *Comparative Biochemistry and Physiology Part C: Toxicology and Pharmacology.* **2011**, *153* (1), 91–98.

(177) Kais, B.; Schneider, K. E.; Keiter, S.; Henn, K.; Ackermann, C.; Braunbeck, T. DMSO modifies the permeability of the zebrafish (*Danio rerio*) chorion-Implications for the fish embryo test (FET). *Aquatic Toxicology.* **2013**, *140–141*, 229–238.

(178) Klüver, N.; König, M.; Ortmann, J.; Massei, R.; Paschke, A.; Kühne, R.; Scholz, S. Fish embryo toxicity test: Identification of compounds with weak toxicity and analysis of behavioral effects to improve prediction of acute toxicity for neurotoxic compounds. *Environmental Science & Technology.* **2015**, *49* (11), 7002–7011.

(179) Dekant, W. The role of biotransformation and bioactivation in toxicity; Birkhäuser Basel, 2009; pp 57–86.

(180) Klüver, N.; Ortmann, J.; Paschke, H.; Renner, P.; Ritter, A. P.; Scholz, S. Transient overexpression of adh8a increases allyl alcohol toxicity in zebrafish embryos. *PLoS One* **2014**, *9* (3), e90619.

(181) Piña, B.; Barata, C. A genomic and ecotoxicological perspective of DNA array studies in aquatic environmental risk assessment. *Aquatic Toxicology.* **2011**, *105* (3–4 Suppl), 40–49.

(182) Martyniuk, C. J.; Griffitt, R. J.; Denslow, N. D. Omics in aquatic toxicology: not just another microarray. *Environmental Toxicology and Chemistry.* **2011**, *30* (2), 263–264.

(183) Schirmer, K.; Tanneberger, K.; Kramer, N. I.; Völker, D.; Scholz, S.; Hafner, C.; Lee, L. E. J.; Bols, N. C.; Hermens, J. L. M. Developing a list of reference chemicals for testing alternatives to whole fish toxicity tests. *Aquatic Toxicology.* **2008**, *90* (2), 128–137.

(184) Kerr, M. K.; Churchill, G. A. Experimental design for gene expression microarrays. *Biostatistics.* **2001**, *2* (2), 183–201.

(185) Berthelot, C.; Brunet, F.; Chalopin, D.; Juanchich, A.; Bernard, M.; Noël, B.; Bento, P.; Da Silva, C.; Labadie, K.; Alberti, A.; et al. The rainbow trout genome provides novel insights into evolution after whole-genome duplication in vertebrates. *Nature Communication.* **2014**, *5*, 3657.

(186) Moreno-Hagelsieb, G.; Latimer, K. Choosing BLAST options for better detection of orthologs as reciprocal best hits. *Bioinformatics.* **2008**, *24* (3), 319–324.

(187) Ritchie, M. E.; Phipson, B.; Wu, D.; Hu, Y.; Law, C. W.; Shi, W.; Smyth, G. K. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research.* **2015**, *43* (7), e47.

(188) Kim, S.-M.; Yoo, J.-A.; Baek, J.-M.; Cho, K.-H. Diethyl phthalate exposure is associated with embryonic toxicity, fatty liver changes, and hypolipidemia via impairment of lipoprotein functions. *Toxicology In Vitro.* **2015**, *30* (1 Pt B), 383–393.

(189) Xu, T.; Zhao, J.; Hu, P.; Dong, Z.; Li, J.; Zhang, H.; Yin, D.; Zhao, Q. Pentachlorophenol exposure causes Warburg-like effects in zebrafish embryos at gastrulation stage. *Toxicology and Applied Pharmacology.* **2014**, *277* (2), 183–191.

(190) DeMicco, A.; Cooper, K. R.; Richardson, J. R.; White, L. A. Developmental neurotoxicity of pyrethroid insecticides in zebrafish embryos. *Toxicological Sciences.* **2010**, *113* (1), 177–186.

(191) Wolf, J. C.; Wheeler, J. R. A critical review of histopathological findings associated with endocrine and non-endocrine hepatic toxicity in fish models. *Aquatic Toxicology.* **2018**, *197*, 60–78.

(192) Pascussi, J.-M.; Gerbal-Chaloin, S.; Duret, C.; Daujat-Chavanieu, M.; Vilarem, M.-J.; Maurel, P. The tangle of nuclear receptors that controls xenobiotic metabolism and transport: crosstalk and consequences. *Annual Review of Pharmacology and Toxicology.* **2008**, *48* (1), 1–32.

(193) Ma, Y.; Han, J.; Guo, Y.; Lam, P. K. S.; Wu, R. S. S.; Giesy, J. P.; Zhang, X.; Zhou, B. Disruption of endocrine function in in vitro H295R cell-based and in in vivo assay in zebrafish by 2,4-dichlorophenol. *Aquatic Toxicology.* **2011**, *106–107* (107), 173–181.

(194) Jeffries, K.; Komoroske, L.; Truong, J.; Werner, I.; Hasenbein, M.; Hasenbein, S.; Fangue, N.; Connon, R. The transcriptome-wide effects of exposure to a pyrethroid pesticide on the Critically Endangered delta smelt Hypomesus transpacificus. *Endangered Species Research.* **2015**, *28* (1), 43–60.

(195) Genomics, B.; Connon, R. E.; Geist, J.; Pfeiff, J.; Loguinov, A. V; D 'abronzo, L. S.; Wintz, H.; Vulpe, C. D.; Werner, I. Linking mechanistic and behavioral responses to sublethal esfenvalerate exposure in the endangered delta smelt; *Hypomesus transpacificus* (Fam. Osmeridae). *BMC Genomics.* **2009**, *10*.

(196) Wang, H.; Chen, T.; Weng, C.; Yang, C.; Tang, M.; Wang, H.; Chen, T.; Weng, C.; Yang, C.; Tang, M. Acrolein preferentially damages nucleolus eliciting ribosomal stress and apoptosis in human cancer cells. *Oncotarget.* **2016**, *7* (49), 80450–80464.

(197) Golla, U.; Bandi, G.; Tomar, R. S. Molecular cytotoxicity mechanisms of allyl alcohol (acrolein) in budding yeast. *Chemical Research in Toxicology.* **2015**, *28* (6), 1246–1264.

(198) Garcia-Reyero, N.; Escalon, B. L.; Loh, P.; Laird, J. G.; Kennedy, A. J.; Berger, B.; Perkins, E. J. Assessment of Chemical Mixtures and Groundwater Effects on *Daphnia magna* Transcriptomics. *Environmental Science & Technology.* **2012**, *46* (1), 42–50.

(199) United Nations Environment Programme (UNEP). United Nations Environment Programme (UNEP) http://www.unep.org/ (accessed Sep 15, 2016).

(200) Judson, R.; Kavlock, R.; Martin, M.; Reif, D.; Houck, K.; Knudsen, T.; Richard, A.; Tice, R. R.; Whelan, M.; Xia, M.; et al. Perspectives on validation of high-throughput assays supporting 21st century toxicity testing. *ALTEX.* **2013**, *30* (1), 51–56.

(201) Shukla, S. J.; Huang, R.; Austin, C. P.; Xia, M. The future of toxicity testing: a focus on in

vitro methods using a quantitative high-throughput screening platform. *Drug Discovery Today.* **2010**, *15* (23–24), 997–1007.

(202) Dix, D. J.; Houck, K. A.; Martin, M. T.; Richard, A. M.; Setzer, R. W.; Kavlock, R. J. The ToxCast program for prioritizing toxicity testing of environmental chemicals. *Toxicological Sciences.* **2007**, *95* (1), 5–12.

(203) Kavlock, R. J.; Austin, C. P.; Tice, R. R. Toxicity testing in the 21st century: Implications for human health risk assessment. *Risk Analysis.* **2009**, *29* (4), 485–487.

(204) Szymański, P.; Markowicz, M.; Mikiciuk-Olasik, E. Adaptation of high-throughput screening in drug discovery—toxicological screening tests. *International Journal of Molecular Sciences.* **2012**, *13* (12), 427–452.

(205) Padilla, S.; Corum, D.; Padnos, B.; Hunter, D. L.; Beam, A.; Houck, K. A.; Sipes, N.; Kleinstreuer, N.; Knudsen, T.; Dix, D. J.; et al. Zebrafish developmental screening of the ToxCast™ Phase I chemical library. *Reproductive Toxicology.* **2012**, *33* (2), 174–187.

(206) George, S.; Lin, S.; Ji, Z.; Thomas, C. R.; Li, L.; Mecklenburg, M.; Meng, H.; Wang, X.; Zhang, H.; Xia, T.; et al. Surface defects on plate-shaped silver nanoparticles contribute to its hazard potential in a fish gill cell line and zebrafish embryos. *ACS Nano.* **2012**, *6* (5), 3745–3759.

(207) Rivero-Wendt, C. L. G.; Oliveira, R.; Monteiro, M. S.; Domingues, I.; Soares, A. M. V. M.; Grisolia, C. K. Steroid androgen 17α-methyltestosterone induces malformations and biochemical alterations in zebrafish embryos. *Environmental Toxicology and Pharmacology.* **2016**, *44*, 107–113.

(208) Tu, W.; Xu, C.; Lu, B.; Lin, C.; Wu, Y.; Liu, W. Acute exposure to synthetic pyrethroids

causes bioconcentration and disruption of the hypothalamus-pituitary-thyroid axis in zebrafish embryos. *Science of the Total Environment.* **2016**, *542* (Pt A), 876–885.

(209) Segner, H. Cytotoxicity assays with fish cells as an alternative to the acute lethality test with fish. *Alternatives to Laboratory Animals.* **2004**, *32* (4), 375–382.

(210) Mothersill, C.; Austin, B. *In vitro methods in aquatic toxicology*; Springer, 2003.

(211) Castaño, A.; Bols, N.; Braunbeck, T.; Dierickx, P.; Halder, M.; Isomaa, B.; Kawahara, K.; Lee, L. E. J.; Mothersill, C.; Pärt, P.; et al. The use of fish cells in ecotoxicology. The report and recommendations of ECVAM Workshop 47. *Alternatives to Laboratory Animals.* **2003**, *31* (3), 317–351.

(212) Gulden, M.; Seibert, H. In vitro-in vivo extrapolation of toxic potencies for hazard and risk assessment – problems and new developments. *ALTEX.* **2005**, *22* (special issue 2).

(213) Heringa, M. B.; Schreurs, R. H. M. M.; Busser, F.; van der Saag, P. T.; van der Burg, B.; Hermens, J. L. M. Toward more useful in vitro toxicity data with measured free concentrations. *Environmental Science & Technology*. **2004**, *38* (23), 6263–6270.

(214) Quignot, N.; Hamon, J.; Bois, F. Y. Extrapolating in vitro results to predict human toxicity. *In vitro toxicology system*. 2014; pp 531–550.

(215) Yoon, M.; Campbell, J. L.; Andersen, M. E.; Clewell, H. J. Quantitative *in vitro* to *in vivo* extrapolation of cell-based toxicity assay results. *Critical Reviews in Toxicology.* **2012**, *42* (8), 633–652.

(216) Mielke, H.; Di Consiglio, E.; Kreutz, R.; Partosch, F.; Testai, E.; Gundert-Remy, U. The importance of protein binding for the in vitro–in vivo extrapolation (IVIVE)—example of

ibuprofen, a highly protein-bound substance. *Archives of Toxicology.* **2017**, *91* (4), 1663–1670.

(217) Verhaar, H. J. M.; van Leeuwen, C. J.; Hermens, J. L. M. Classifying environmental pollutants. *Chemosphere.* **1992**, *25* (4), 471–491.

(218) Verhaar, H. J. M.; Urrestarazu Ramos, E.; Hermens, J. L. M. Classifying environmental pollutants. 2: Separation of class 1 (baseline toxicity) and class 2 ("polar narcosis") type compounds based on chemical descriptors. *Journal of Chemometrics.* **1996**, *10* (2), 149–162.

(219) Verhaar, H. J. .; Solbé, J.; Speksnijder, J.; van Leeuwen, C. J.; Hermens, J. L. . Classifying environmental pollutants: Part 3. External validation of the classification system. *Chemosphere.* **2000**, *40* (8), 875–883.

(220) Barron, M. G.; Lilavois, C. R.; Martin, T. M. MOAtox: A comprehensive mode of action and acute aquatic toxicity database for predictive model development. *Aquatic Toxicology.* **2015**, *161*, 102–107.

(221) Bradbury, S.; Carlson, R.; Henry, T. Polar narcosis in aquatic organisms. In *Aquatic Toxicology and Hazard Assessment: 12th Volume*; ASTM International: 100 Barr Harbor Drive, PO Box C700, West Conshohocken, PA 19428-2959, 1989; pp 59-59–15.

(222) Roberts, D.; Costello, J. Mechanisms of action for general and polar narcosis: A difference in dimension. *QSAR & Combinatorial Science.* **2003**, *22* (2), 226–233.

(223) Fafandel, M.; Bihari, N.; Perić, L.; Cenov, A. Effect of marine pollutants on the acid DNase activity in the hemocytes and digestive gland of the mussel *Mytilus galloprovincialis*. *Aquatic Toxicology.* **2008**, *86* (4), 508–513.

(224) Whitehead, A. Interactions between oil-spill pollutants and natural stressors can compound ecotoxicological effects. *Integrative and Comparative Biology.* **2013**, *53* (4), 635–647.

(225) Johnston, J. E.; Hoffman, K.; Wing, S.; Lowman, A. Fish consumption patterns and mercury advisory knowledge among fishers in the haw river basin. *North Carolina Medical Journal.* **2016**, *77* (1), 9–14.

(226) Turyk, M. E.; Bhavsar, S. P.; Bowerman, W.; Boysen, E.; Clark, M.; Diamond, M.; Mergler, D.; Pantazopoulos, P.; Schantz, S.; Carpenter, D. O. Risks and benefits of consumption of Great Lakes fish. *Environmental Health and Perspectives.* **2012**, *120* (1), 11–18.

(227) Gurney, W. S. C. Modeling the demographic effects of endocrine disruptors. *Environmental Health and Perspectives.* **2006**, *114 Suppl 1* (Suppl 1), 122–126.

(228) Kidd, K. A.; Blanchfield, P. J.; Mills, K. H.; Palace, V. P.; Evans, R. E.; Lazorchak, J. M.; Flick, R. W.; Swackhamer, D. Collapse of a fish population after exposure to a synthetic estrogen. *Proceedings of the National Academy of Sciences of the United States of America.* **2007**, *22* (21), 8897–8901.

(229) Gore, A.; Crews, D.; Doan, L.; La Merrill, M.; Patisaul, H.; Zota, A. *Introduction to endocrine disrupting chemicals (EDCs): A guide for public interest organizations and policy-makers*; 2014.

(230) World Health Organization. *Global assessment of the state-of-the-science of endocrine disruptors*; World Health Organization, 2013.

(231) Kavlock, R. J.; Daston, G. P.; DeRosa, C.; Fenner-Crisp, P.; Gray, L. E.; Kaattari, S.; Lucier, G.; Luster, M.; Mac, M. J.; Maczka, C.; et al. Research needs for the risk assessment of

health and environmental effects of endocrine disruptors: a report of the U.S. EPA-sponsored workshop. *Environmental Health and Perspectives.* **1996**, *104 Suppl 4*, 715–740.

(232) Rudel, R. A.; Perovich, L. J. Endocrine disrupting chemicals in indoor and outdoor air. *Atmospheric Environment. (1994).* **2009**, *43* (1), 170–181.

(233) Hotchkiss, A. K.; Rider, C. V.; Blystone, C. R.; Wilson, V. S.; Hartig, P. C.; Ankley, G. T.; Foster, P. M.; Gray, C. L.; Gray, L. E. Fifteen years after "Wingspread" - environmental endocrine disrupters and human and wildlife health: where we are today and where we need to go. *Toxicological Sciences.* **2008**, *105* (2), 235–259.

(234) Abdel-moneim, A.; Coulter, D. P.; Mahapatra, C. T.; Sepúlveda, M. S. Intersex in fishes and amphibians: population implications, prevalence, mechanisms and molecular biomarkers. *Journal of Applied Toxicology.* **2015**, *35* (11), 1228–1240.

(235) Lackey, B. R.; Gray, S. L.; Henricks, D. M.; Chang, C.; Katzenellenbogen, J. A. 20.; Sugiyama, Y. Crosstalk and considerations in endocrine disruptor research. *Medical Hypotheses.* **2001**, *56* (6), 644–647.

(236) Vandenberg, L. N.; Colborn, T.; Hayes, T. B.; Heindel, J. J.; Jacobs, D. R.; Lee, D.-H.; Shioda, T.; Soto, A. M.; vom Saal, F. S.; Welshons, W. V.; et al. Hormones and endocrine-disrupting chemicals: Low-dose effects and nonmonotonic dose responses. *Endocrine Reviews.* **2012**, *33* (3), 378–455.

(237) Patino, R. Manipulations of the reproductive system of fishes by means of exogenous chemicals. *The progressive Fish-Culturist.* **1997**, *59* (2), 118–128.

(238) Patiño, R.; Sullivan, C. V. Ovarian follicle growth, maturation, and ovulation in teleost

fish. *Fish Physiology and Biochemistry*. **2002**, *26* (1), 57–70.

(239) Grier, J.; Uribe, M.; Patino, R. The ovary, folliculogenesis and oogenesis in teleosts. *Reproductive Biology and phylogeny of fishes (Agnathans bony fishes)* **2009**, 25–84.

(240) Martyniuk, C. J.; Kroll, K. J.; Porak, W. F.; Steward, C.; Grier, H. J.; Denslow, N. D. Seasonal relationship between gonadotropin, growth hormone, and estrogen receptor mRNA expression in the pituitary gland of Largemouth bass. *General and Comparative Endocrinology*. **2009**, *163* (3), 306–317.

(241) Menn, F. Le; Cerdà, J.; Babin, P. J. Ultrastructural aspects of the ontogeny and differentiation of ray-finned fish ovarian follicles. *The Fish Oocyte*; Springer Netherlands: Dordrecht, 2007; pp 1–37.

(242) França, G. F.; Grier, H. J.; Quagio-Grassiotto, I. A new vision of the origin and the oocyte development in the ostariophysi applied to *Gymnotus sylvius* (Teleostei, Gymnotiformes). *Neotropical Ichthyology.* **2010**, *8* (4), 787–804.

(243) Bazzoli, N.; Pereira Godinho, H. Cortical alveoli in oocytes of freshwater neotropical teleost fish. *Bollettino di Zoologia.* **1994**, *61*, 301–308.

(244) Mattingly, C. J.; Colby, G. T.; Forrest, J. N.; Boyer, J. L. The Comparative Toxicogenomics Database (CTD). *Environmental Health and Perspectives.* **2003**, *111* (6), 793–795.

(245) Sepúlveda, M. S.; Johnson, W. E.; Higman, J. C.; Denslow, N. D.; Schoeb, T. R.; Gross, T. S. An evaluation of biomarkers of reproductive function and potential contaminant effects in Florida largemouth bass (*Micropterus salmoidesfloridanus*) sampled from the St. Johns River. *Science of the Total Environment.* **2002**, *289* (1–3), 133–144.

(246) Garcia-Reyero, N.; Griffitt, R. J.; Liu, L.; Kroll, K. J.; Farmerie, W. G.; Barber, D. S.; Denslow, N. D. Construction of a robust microarray from a non-model species (largemouth bass) using pyrosequencing technology. *Journal of Fish Biology.* **2008**, *72* (9), 2354–2376.

(247) Sonnenschein, C.; Soto, A. M. An updated review of environmental estrogen and androgen mimics and antagonists. *The Journal of Steroid Biochemistry and Molecular Biology.* **1998**, *65* (1–6), 143–150.

(248) Bayley, M.; Junge, M.; Baatrup, E. Exposure of juvenile guppies to three antiandrogens causes demasculinization and a reduced sperm count in adult males. *Aquatic Toxicology* **2002**, *56* (4), 227–239.

(249) Ankley, G. T.; Gray, L. E. Cross-species conservation of endocrine pathways: a critical analysis of tier 1 fish and rat screening assays with 12 model chemicals. *Environmental Toxicology and Chemistry.* **2013**, *32* (5), 1084–1087.

(250) Smith, R. N.; Aleksic, J.; Butano, D.; Carr, A.; Contrino, S.; Hu, F.; Lyne, M.; Lyne, R.; Kalderimis, A.; Rutherford, K.; et al. InterMine: a flexible data warehouse system for the integration and analysis of heterogeneous biological data. *Bioinformatics.* **2012**, *28* (23), 3163–3165.

(251) Blum, J.; Nyagode, B. A.; James, M. O.; Denslow, N. D. Effects of the pesticide methoxychlor on gene expression in the liver and testes of the male largemouth bass (*Micropterus salmoides*). *Aquatic Toxicology.* **2008**, *86* (4), 459–469.

(252) Garcia-Reyero, N.; Barber, D.; Gross, T.; Johnson, K.; Sepulveda, M.; Szabo, N.; Denslow, N. Dietary exposure of largemouth bass to OCPs changes expression of genes important

for reproduction. *Aquatic Toxicology.* **2006**, *78* (4), 358–369.

(253) Sabo-Attwood, T.; Kroll, K. J.; Denslow, N. D. Differential expression of largemouth bass (*Micropterus salmoides*) estrogen receptor isotypes alpha, beta, and gamma by estradiol. *Molecular and Cellular Endocrinology.* **2004**, *218* (1–2), 107–118.

(254) Biales, A. D.; Bencic, D. C.; Lazorchak, J. L.; Lattier, D. L. A quantitative real-time polymerase chain reaction method for the analysis of vitellogenin transcripts in model and nonmodel fish species. *Environmental Toxicology and Chemistry.* **2007**, *26* (12), 2679.

(255) Giguère, V. Canonical signaling and nuclear activity of mTOR-a teamwork effort to regulate metabolism and cell growth. *FEBS Journal.* **2018**.

(256) Stempin, S.; Andres, S.; Bumke Scheer, M.; Rode, A.; Nau, H.; Seidel, A.; Lampen, A. Valproic acid and its derivatives enhanced estrogenic activity but not androgenic activity in a structure dependent manner. *Reproductive Toxicology.* **2013**, *42*, 49–57.

(257) Gustavsen, M. W.; von Krogh, K.; Taubøll, E.; Zimmer, K. E.; Dahl, E.; Olsaker, I.; Ropstad, E.; Verhaegen, S. Differential effects of antiepileptic drugs on steroidogenesis in a human in vitro cell model. *Acta Neurologica Scandinavica Supplementum.* **2009**, No. 189, 14–21.

(258) Esquifino, A. I.; Moreno, M. L.; Agrasal, C.; Villanúa, M. A. Effects of cyclosporine on ovarian function in sham-operated and pituitary-grafted young female rats. *Proceedings of the Society for Experimental Biology and Medicine.* **1995**, *208* (4), 397–403.

(259) Shivanandappa, T.; Krishnakumari, M. K.; Majumder, S. K. Testicular atrophy in *Gallus domesticus* fed acute doses of copper fungicides. *Poultry Science.* **1983**, *62* (2), 405–408.

(260) Faßbender, C.; Braunbeck, T. Reproductive and genotoxic effects in zebrafish after chronic exposure to methyl methanesulfonate in a multigeneration study. *Ecotoxicology.* **2013**, *22* (5), 825–837.

(261) Georgescu, B.; Georgescu, C.; Dărăban, S.; Bouaru, A.; Paşcalău, S. Heavy metals acting as endocrine disrupters. *Scientific Papers: Animal Science and Biotechnology.* **2011**, *44* (2).

(262) Dowdy, J.; Brower, S.; Miller, M. R. Acetaminophen exhibits weak antiestrogenic activity in human endometrial adenocarcinoma (Ishikawa) cells. *Toxicological Sciences.* **2003**, *72* (1), 57–65.

(263) Kristensen, D. M.; Lesné, L.; Le Fol, V.; Desdoits-Lethimonier, C.; Dejucq-Rainsford, N.; Leffers, H.; Jégou, B. Paracetamol (acetaminophen), aspirin (acetylsalicylic acid) and indomethacin are anti-androgenic in the rat foetal testis. *International Journal of Andrology.* **2012**, *35* (3), 377–384.

(264) Omran, N. E.; Salama, W. M. The endocrine disruptor effect of the herbicides atrazine and glyphosate on *Biomphalaria alexandrina* snails. *Toxicology and Industrial Health* **2016**, *32* (4), 656–665.

(265) Shu, X.; Hu, X.; Zhou, S.; Xu, C.; Qiu, Q.; Nie, S.; Xie, M. [Effect of Quercetin exposure during the prepubertal period on ovarian development and reproductive endocrinology of mice]. *Yao Xue Xue Bao.* **2011**, *46* (9), 1051–1057.

(266) Duong, A.; Steinmaus, C.; McHale, C. M.; Vaughan, C. P.; Zhang, L. Reproductive and developmental toxicity of formaldehyde: a systematic review. *Mutation Research.* **2011**, *728* (3), 118–138.

(267) Kim, J.; Hayton, W. L.; Schultz, I. R. Modeling the brain–pituitary–gonad axis in salmon. *Marine Environmental Research.* **2006**, *62*, S426–S432.

(268) Li, Z.; Kroll, K. J.; Jensen, K. M.; Villeneuve, D. L.; Ankley, G. T.; Brian, J. V; Sepúlveda, M. S.; Orlando, E. F.; Lazorchak, J. M.; Kostich, M.; et al. A computational model of the hypothalamic: pituitary: gonadal axis in female fathead minnows (*Pimephales promelas*) exposed to 17α-ethynylestradiol and 17β-trenbolone. *BMC System Biology.* **2011**, *5*, 63.

(269) Li, Z.; Villeneuve, D. L.; Jensen, K. M.; Ankley, G. T.; Watanabe, K. H.; Kidd, K. A computational model for asynchronous oocyte growth dynamics in a batch-spawning fish. *Canadian Journal of Fisheries and Aquatic Sciences.* **2011**, *68* (9), 1528–1538.

(270) Gillies, K.; Krone, S. M.; Nagler, J. J.; Schultz, I. R. A computational model of the rainbow trout hypothalamus-pituitary-ovary-liver axis. *PLoS Computational Biology.* **2016**, *12* (4), e1004874.

(271) Guzmán, J. M.; Luckenbach, J. A.; Yamamoto, Y.; Swanson, P. Expression profiles of Fsh-regulated ovarian genes during oogenesis in coho salmon. *PLoS One.* **2014**, *9* (12), e114176.

(272) Dorrington, J. H.; Armstrong, D. T. Follicle-stimulating hormone stimulates estradiol-17beta synthesis in cultured Sertoli cells. *Proceedings of the National Academy of Sciences of the United States of America. U. S. A.* **1975**, *72* (7), 2677–2681.

(273) Breton, T. S.; Anderson, J. L.; Goetz, F. W.; Berlinsky, D. L. Identification of ovarian gene expression patterns during vitellogenesis in Atlantic cod (*Gadus morhua*). *General and Comparative Endocrinology.* **2012**, *179* (2), 296–304.

(274) Ornostay, A.; Cowie, A. M.; Hindle, M.; Baker, C. J. O.; Martyniuk, C. J. Classifying

chemical mode of action using gene networks and machine learning: A case study with the herbicide linuron. *Comparative Biochemistry and Physiology Part D: Genomics and Proteomics* **2013**, *8* (4), 263–274.

(275) Yang, Y.; Maxwell, A.; Zhang, X.; Wang, N.; Perkins, E. J.; Zhang, C.; Gong, P. Differential reconstructed gene interaction networks for deriving toxicity threshold in chemical risk assessment. *BMC Bioinformatics.* **2013**, *14 Suppl 14* (Suppl 14), S3.

(276) Noh, H.; Gunawan, R. Inferring gene targets of drugs and chemical compounds from gene expression profiles. *Bioinformatics.* **2016**, *32* (14), 2120–2127.

(277) Hooper, M. J.; Ankley, G. T.; Cristol, D. A.; Maryoung, L. A.; Noyes, P. D.; Pinkerton, K. E. Interactions between chemical and climate stressors: A role for mechanistic toxicology in assessing climate change risks. *Environmental Toxicology and Chemistry.* **2013**, *32* (1), 32–48.

(278) Groh, K. J.; Carvalho, R. N.; Chipman, J. K.; Denslow, N. D.; Halder, M.; Murphy, C. A.; Roelofs, D.; Rolaki, A.; Schirmer, K.; Watanabe, K. H. Development and application of the adverse outcome pathway framework for understanding and predicting chronic toxicity: I. Challenges and research needs in ecotoxicology. *Chemosphere.* **2015**, *120*, 764–777.

(279) Cardoso-Vera, J. D.; Islas-Flores, H.; SanJuan-Reyes, N.; Montero-Castro, E. I.; Galar-Martínez, M.; García-Medina, S.; Elizalde-Velázquez, A.; Dublán-García, O.; Gómez-Oliván, L. M. Comparative study of diclofenac-induced embryotoxicity and teratogenesis in Xenopus laevis and Lithobates catesbeianus, using the frog embryo teratogenesis assay: Xenopus (FETAX). *Science of the Total Environment.* **2016**, *574*, 467–475.

(280) Stoddard, K. I.; Huggett, D. B. Early life stage (ELS) toxicity of sucralose to fathead

minnows, *Pimephales promelas*. *Bulletin of Environmental Contamination and Toxicology.* **2014**, *93* (4), 383–387.

(281) Griesbach, R. J. Biochemistry and genetics of flower color. *Plant Breedings Reviews.* **2005**, *25*, 89–114.

(282) Takahashi, A.; Ohnishi, T. The significance of the study about the biological effects of solar ultraviolet radiation using the exposed facility on the international space station. *Biological Sciences in Space.* **2004**, *18* (4), 255–260.

(283) Ross, J. A.; Kasum, C. M. Dietary flavonoids : bioavailability, metabolic effects, and safety. *Annual Review of Nutrition.* **2002**, *22* (1), 19–34.

(284) Prince, P. S. M.; Sathya, B. Pretreatment with Quercetin ameliorates lipids, lipoproteins and marker enzymes of lipid metabolism in isoproterenol treated cardiotoxic male Wistar rats. *European Journal of Pharmacology.* **2010**, *635* (1–3), 142–148.

(285) Hillis, W. E. *Wood extractives and their significance to the pulp and paper industries.*; Academic Press, 1962.

(286) Sjöström, E. *Wood chemistry : fundamentals and applications*; Academic Press, 1993.

(287) Pês, T. S.; Saccol, E. M. H.; Ourique, G. M.; Londero, É. P.; Gressler, L. T.; Golombieski, J. I.; Glanzner, W. G.; Llesuy, S. F.; Gonçalves, P. B. D.; Neto, J. R.; et al. Quercetin in the diet of silver catfish: Effects on antioxidant status, blood parameters and pituitary hormone expression. *Aquaculture.* **2016**, *458*, 100–106.

(288) Padma, V. V.; Lalitha, G.; Shirony, N. P.; Baskaran, R. Effect of Quercetin against lindane induced alterations in the serum and hepatic tissue lipids in wistar rats. *Asian Pacific*

Journal of Tropical Biomedicine. **2012**, *2* (11), 910–915.

(289) Hu, Q.-H.; Zhang, X.; Pan, Y.; Li, Y.-C.; Kong, L.-D. Allopurinol, Quercetin and Rutin ameliorate renal NLRP3 inflammasome activation and lipid accumulation in fructose-fed rats. *Biochemical Pharmacology.* **2012**, *84* (1), 113–125.

(290) Zhai, S.-W.; Liu, S.-L. Effects of dietary Quercetin on growth performance, serum lipids level and body composition of Tilapia (*Oreochromis Niloticus*). *Italian Journal of Animal Sciences.* **2013**, *12* (4), e85.

(291) Kikuchi, K.; Furuta, T.; Iwata, N.; Onuki, K.; Noguchi, T. Effect of dietary lipid levels on the growth, feed utilization, body composition and blood characteristics of tiger puffer Takifugu rubripes. *Aquaculture.* **2009**, *298* (1–2), 111–117.

(292) Naseer, Z.; Ahmad, E.; Epikmen, E. T.; Uçan, U.; Boyacioğlu, M.; İpek, E.; Akosy, M. Quercetin supplemented diet improves follicular development, oocyte quality, and reduces ovarian apoptosis in rabbits during summer heat stress. *Theriogenology.* **2017**, *96*, 136–141.

(293) Kang, J.-T.; Kwon, D.-K.; Park, S.-J.; Kim, S.-J.; Moon, J.-H.; Koo, O.-J.; Jang, G.; Lee, B.-C. Quercetin improves the in vitro development of porcine oocytes by decreasing reactive oxygen species levels. *Journal of Veterinary Sciences.* **2013**, *14* (1), 15–20.

(294) Weber, L. P.; Kiparissis, Y.; Hwang, G. S.; Niimi, A. J.; Janz, D. M.; Metcalfe, C. D. Increased cellular apoptosis after chronic aqueous exposure to nonylphenol and Quercetin in adult medaka (*Oryzias latipes*). *Comparative Biochemistry and Physiology C: Toxicoogy and Pharmacology.* **2002**, *131* (1), 51–59.

(295) Niederreither, K.; Dollé, P. Retinoic acid in development: towards an integrated view.

Naure Review Genetics. **2008**, *9* (7), 541–553.

(296) Rhinn, M.; Dolle, P. Retinoic acid signalling during development. *Development.* **2012**, *139* (5), 843–858.

(297) Larange, A.; Cheroutre, H. Retinoic acid and retinoic acid receptors as pleiotropic modulators of the immune system. *Annual Review of Immunology.* **2016**, *34* (1), 369–394.

(298) Agrimson, K. S.; Hogarth, C. A. Germ cell commitment to oogenic versus spermatogenic pathway: The role of retinoic acid. *Results and problems in cell differentiation*. 2016; Vol. 58, pp 135–166.

(299) Teletin, M.; Vernet, N.; Ghyselinck, N. B.; Mark, M. Roles of retinoic acid in germ cell differentiation. *Current topics in developmental biology*. 2017; Vol. 125, pp 191–225.

(300) Kumar, S.; Dollé, P.; Ghyselinck, N. B.; Duester, G. Endogenous retinoic acid signaling is required for maintenance and regeneration of cornea. *Experimental Eye Research.* **2017**, *154*, 190–195.

(301) Baldwin, H. E.; Nighland, M.; Kendall, C.; Mays, D. A.; Grossman, R.; Newburger, J. 40 years of topical tretinoin use in review. *Journal of Drugs in Dermatology.* **2013**, *12* (6), 638–642.

(302) Noy, N. Non-classical transcriptional activity of retinoic acid. *Sub-cellular biochemistry*; 2016; Vol. 81, pp 179–199.

(303) Christian, M. S.; Mitala, J. J.; Powers, W. J.; McKenzie, B. E.; Latriano, L. A developmental toxicity study of tretinoin emollient cream (Renova) applied topically to New Zealand

white rabbits. *Journal of the American Academy of Dermatology.* **1997**, *36* (3 Pt 2), S67-76.

(304) Seegmiller, R. E.; Ford, W. H.; Carter, M. W.; Mitala, J. J.; Powers, W. J. A developmental toxicity study of tretinoin administered topically and orally to pregnant Wistar rats. *Journal of the American Academy of Dermatology.* **1997**, *36* (3), S60–S66.

(305) Zhu, Y.; Zhu, Y.; Yin, H.; Zhou, H.; Wan, X.; Zhu, J.; Zhang, T. All-trans-retinoic acid induces short forelimb malformation during mouse embryo development by inhibiting chondrocyte maturation rather than by evoking excess cell death. *Toxicology Letters.* **2012**, *211* (2), 172–186.

(306) Pu, Y.; Wang, Z.; Bian, Y.; Zhang, F.; Yang, P.; Li, Y.; Zhang, Y.; Liu, Y.; Fang, F.; Cao, H.; et al. All- *trans* retinoic acid improves goat oocyte nuclear maturation and reduces apoptotic cumulus cells during *in vitro* maturation. *Animal Sciences Journal.* **2014**, *85* (9), 833–839.

(307) Lagarde, F.; Beausoleil, C.; Belcher, S. M.; Belzunces, L. P.; Emond, C.; Guerbet, M.; Rousselle, C. Non-monotonic dose-response relationships and endocrine disruptors: a qualitative method of assessment. *Environmental Health.* **2015**, *14*, 13.

(308) Shanle, E. K.; Xu, W. Endocrine disrupting chemicals targeting estrogen receptor signaling: identification and mechanisms of action. *Chemical Research in Toxicology.* **2011**, *24* (1), 6–19.

(309) Takayanagi, S.; Tokunaga, T.; Liu, X.; Okada, H.; Matsushima, A.; Shimohigashi, Y. Endocrine disruptor bisphenol A strongly binds to human estrogen-related receptor γ (ERRγ) with high constitutive activity. *Toxicology Letters.* **2006**, *167* (2), 95–105.

(310) Lewbart, G.; Lewbart, G. *Invertebrate medicine*; Wiley-Blackwell, 2012.

(311) Brück, K. Functions of the endocrine system. *Human Physiology*; Springer Berlin Heidelberg: Berlin, Heidelberg, 1983; pp 658–687.

(312) LaFont, R. The endocrinology of invertebrates. *Ecotoxicology.* **2000**, *9* (1/2), 41–57.

(313) Hartenstein, V. The neuroendocrine system of invertebrates: a developmental and evolutionary perspective. *Journal of Endocrinology.* **2006**, *190* (3), 555–570.

(314) Carlisle, D. B.; Passano, L. M. The X-organ of crustacea. *Nature.* **1953**, *171* (4363), 1070–1071.

(315) Nakatsuji, T.; Lee, C.-Y.; Watson, R. D. Crustacean molt-inhibiting hormone: Structure, function, and cellular mode of action. *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology.* **2009**, *152* (2), 139–148.

(316) Treerattrakool, S.; Panyim, S.; Chan, S.-M.; Withyachumnarnkul, B.; Udomkit, A. Molecular characterization of gonad-inhibiting hormone of *Penaeus monodon* and elucidation of its inhibitory role in vitellogenin expression by RNA interference. *FEBS Journal.* **2008**, *275* (5), 970–980.

(317) Wainwright, G.; Webster, S. G.; Wilkinson, M. C.; Chung, J. S.; Rees, H. H. Structure and significance of mandibular organ-inhibiting hormone in the crab, *Cancer pagurus*. Involvement in multihormonal regulation of growth and reproduction. *Journal of Biological Chemistry.* **1996**, *271* (22), 12749–12754.

(318) Aréchiga, H.; Cabrera-Peralta, C.; Huberman, A. Functional characterization of the neurodepressing hormone in the crayfish. *Journal of Neurobiology.* **1979**, *10* (4), 409–

422.

(319) Chung, J. S.; Zmora, N.; Katayama, H.; Tsutsui, N. Crustacean hyperglycemic hormone (CHH) neuropeptidesfamily: Functions, titer, and binding to target tissues. *General and Comparative Endocrinology.* **2010**, *166* (3), 447–454.

(320) Mykles, D. L. Ecdysteroid metabolism in crustaceans. *Journal of Steroid Biochemistry and Molecular Biology.* **2011**, *127* (3–5), 196–203.

(321) Spindler, K.-D.; Hönl, C.; Tremmel, C.; Braun, S.; Ruff, H.; Spindler-Barth, M. Ecdysteroid hormone action. *Cellular and Molecular Life Sciences.* **2009**, *66* (24), 3837–3850.

(322) Techa, S.; Chung, J. S. Ecdysteroids regulate the levels of molt- inhibiting hormone (MIH) expression in the blue crab, *Callinectes sapidus*. *PLoS One* **2015**, *10* (4), 1–19.

(323) Yao, T.-P.; Forman, B. M.; Jiang, Z.; Cherbas, L.; Chen, J.-D.; McKeown, M.; Cherbas, P.; Evans, R. M. Functional ecdysone receptor is the product of EcR and Ultraspiracle genes. *Nature.* **1993**, *366* (6454), 476–479.

(324) Sumiya, E.; Ogino, Y.; Toyota, K.; Miyakawa, H.; Miyagawa, S.; Iguchi, T. *Neverland* regulates embryonic moltings through the regulation of ecdysteroid synthesis in the water flea *Daphnia magna* , and may thus act as a target for chemical disruption of molting. *Journal of Applied Toxicology.* **2016**, *36* (11), 1476–1485.

(325) Hannas, B. R.; Wang, Y. H.; Thomson, S.; Kwon, G.; Li, H.; LeBlanc, G. A. Regulation and dysregulation of vitellogenin mRNA accumulation in daphnids (*Daphnia magna*). *Aquatic Toxicology.* **2011**, *101* (2), 351–357.

(326) Riddiford, L. M. Juvenile hormone action: A 2007 perspective. *Journal of Insect*

Physiology. **2008**, *54* (6), 895–901.

(327) Yin, C.-M.; Zou, B.-X.; Jiang, M.; Li, M.-F.; Qin, W.; Potter, T. L.; Stoffolano, J. G. Identification of juvenile hormone III bisepoxide (JHB3), juvenile hormone III and methyl farnesoate secreted by the corpus allatum of *Phormia regina* (Meigen), in vitro and function of JHB3 either applied alone or as a part of a juvenoid blend. *Journal of Insect Physiology.* **1995**, *41* (6), 473–479.

(328) Jindra, M.; Bellés, X.; Shinoda, T. Molecular basis of juvenile hormone signaling. *Current Opinions in Insect Sciences.* **2015**, *11*, 39–46.

(329) Reddy, P. R.; Nagaraju, G. P. C.; Reddy, P. S. Involvement of methyl farnesoate in the regulation of molting and reproduction in the freshwater crab *Oziotelphusa senex* senex. *Journal of Crustacean Biology.* **2004**, *24* (3), 511–515.

(330) Lobbia, S.; Futahashi, R.; Fujiwara, H. Modulation of the ecdysteroid-induced cell death by juvenile hormone during pupal wing development of Lepidoptera. *Archives of Insect Biochemistry and Physiology.* **2007**, *65* (3), 152–163.

(331) Jones, G.; Jones, D.; Fang, F.; Xu, Y.; New, D.; Wu, W.-H. Juvenile hormone action through a defined enhancer motif to modulate ecdysteroid-activation of natural core promoters. *Comparative Biochemistry and Physiology Part B: Biochemistry & Molecular Biology.* **2012**, *161* (3), 219–225.

(332) Hiruma, K.; Shinoda, T.; Malone, F.; Riddiford, L. M. Juvenile hormone modulates 20-hydroxyecdysone-inducible ecdysone receptor and ultraspiracle gene expression in the tobacco hornworm, *Manduca sexta*. *Development Genes and Evolution.* **1999**, *209* (1), 18–30.

(333) Dubrovsky, E. Hormonal cross talk in insect development. *Trends in Endocrinology and Metabolism.* **2005**, *16* (1), 6–11.

(334) Olmstead, A. W.; LeBlanc, G. A. Insecticidal juvenile hormone analogs stimulate the production of male offspring in the crustacean *Daphnia magna*. *Environmental Health and Perspectives.* **2003**, *111* (7), 919–924.

(335) LeBlanc, G. A. Crustacean endocrine toxicology: a review. *Ecotoxicology.* **2007**, *16* (1), 61–81.

(336) Soin, T.; Smagghe, G. Endocrine disruption in aquatic insects: a review. *Ecotoxicology.* **2007**, *16* (1), 83–93.

(337) Rodríguez, E. M.; Medesani, D. A.; Fingerman, M. Endocrine disruption in crustaceans due to pollutants: A review. *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology.* **2007**, *146* (4), 661–671.

(338) Shang, S. Differentially expressed genes in *Daphnia magna* exposed to methyl farnesoate, 2008. Thesis at Berkeley University.

(339) Ginjupalli, G. K.; Baldwin, W. S. The time- and age-dependent effects of the juvenile hormone analog pesticide, pyriproxyfen on *Daphnia magna* reproduction. *Chemosphere.* **2013**, *92* (9), 1260–1266.

(340) Giraudo, M.; Douville, M.; Cottin, G.; Houde, M. Transcriptomic, cellular and life-history responses of Daphnia magna chronically exposed to benzotriazoles: Endocrine-disrupting potential and molting effects. *PLoS One.* **2017**, *12* (2), e0171763.

(341) Antczak, P.; White, T. A.; Giri, A.; Michelangeli, F.; Viant, M. R.; Cronin, M. T. D.; Vulpe,

C.; Falciani, F. Systems biology approach reveals a calcium-dependent mechanism for basal toxicity in *Daphnia magna*. *Environmental Science & Technology.* **2015**, *49* (18), 11132–11140.

(342) Rewitz, K. F.; Gilbert, L. I. Daphnia Halloween genes that encode cytochrome P450s mediating the synthesis of the arthropod molting hormone: evolutionary implications. *BMC Evolutionary Biology.* **2008**, *8* (1), 60.

(343) DeFur, P. L.; International SETAC Workshop on Endocrine Disruption in Invertebrates: Endocrinology, T. *Endocrine disruption in invertebrates : endocrinology, testing and assessment*; SETAC Press, 1999.

(344) Song, Y.; Evenseth, L. M.; Iguchi, T.; Tollefsen, K. E. Release of chitobiase as an indicator of potential molting disruption in juvenile *Daphnia magna* exposed to the ecdysone receptor agonist 20-hydroxyecdysone. *Journal of Toxicology and Environmental Health Part A.* **2017**, *80* (16–18), 954–962.

(345) Kashian, D. R.; Dodson, S. I. Effects of vertebrate hormones on development and sex determination in *Daphnia magna*. *Environmental Toxicology and Chemistry.***2004**, *23* (5), 1282.

(346) Toyota, K.; Kato, Y.; Miyakawa, H.; Yatsu, R.; Mizutani, T.; Ogino, Y.; Miyagawa, S.; Watanabe, H.; Nishide, H.; Uchiyama, I.; et al. Molecular impact of juvenile hormone agonists on neonatal *Daphnia magna*. *Journal of Applied Toxicology.* **2014**, *34* (5), 537–544.

(347) Campos, B.; Garcia-Reyero, N.; Rivetti, C.; Escalon, L.; Habib, T.; Tauler, R.; Tsakovski, S.; Piña, B.; Barata, C. Identification of metabolic pathways in *Daphnia magna* explaining

hormetic effects of selective serotonin reuptake inhibitors and 4-nonylphenol using transcriptomic and phenotypic responses. *Environmental Science & Technology.* **2013**, *47* (16), 9434–9443.

(348) Baer, K. N.; Owens, K. D. Evaluation of selected endocrine disrupting compounds on sex determination in *Daphnia magna* using reduced photoperiod and different feeding rates. *Bulletin of Environmental Contamination and Toxicology.* **1999**, *62* (2), 214–221.

(349) Zou, E.; Fingerman, M. Effects of estrogenic xenobiotics on molting of the water flea, *Daphnia magna*. *Ecotoxicology and Environmental Safety.* **1997**, *38* (3), 281–285.

(350) Kashian, D. R.; Dodson, S. I. Effects of common-use pesticides on developmental and reproductive processes in Daphnia. *Toxicology and Industrial Health.* **2002**, *18* (5), 225–235.

(351) Baldwin, W. S.; Bailey, R.; Long, K. E.; Klaine, S. Incomplete ecdysis is an indicator of ecdysteroid exposure in *Daphnia magna*. *Environmental Toxicology and Chemistry.* **2001**, *20* (7), 1564–1569.

(352) Kretschmann, A.; Ashauer, R.; Preuss, T. G.; Spaak, P.; Escher, B. I.; Hollender, J. Toxicokinetic model describing bioconcentration and biotransformation of Diazinon in *Daphnia magna*. *Environmental Science & Technology.* **2011**, *45* (11), 4995–5002.

(353) Sánchez, M.; Ferrando, M. D.; Sancho, E.; Andreu, E. Physiological perturbations in several generations of *Daphnia magna* straus exposed to Diazinon. *Ecotoxicology and Environmental Safety.* **2000**, *46* (1), 87–94.

(354) US EPA, O. Endocrine Disruptor Screening Program Tier 1 Screening Determinations and Associated Data Evaluation Records.

(355) Hayes, T. B.; Anderson, L. L.; Beasley, V. R.; de Solla, S. R.; Iguchi, T.; Ingraham, H.; Kestemont, P.; Kniewald, J.; Kniewald, Z.; Langlois, V. S.; et al. Demasculinization and feminization of male gonads by atrazine: Consistent effects across vertebrate classes. *Journal of Steroid Biochemistry and Molecular Biology.* **2011**, *127* (1), 64–73.

(356) Foradori, C. D.; Hinds, L. R.; Hanneman, W. H.; Handa, R. J. Effects of atrazine and its withdrawal on gonadotropin-releasing hormone neuroendocrine function in the adult female Wistar rat. *Biology of Reproduction.* **2009**, *81* (6), 1099–1105.

(357) Kucka, M.; Pogrmic-Majkic, K.; Fa, S.; Stojilkovic, S. S.; Kovacevic, R. Atrazine acts as an endocrine disrupter by inhibiting cAMP-specific phosphodiesterase-4. *Toxicology and Applied Pharmacology.* **2012**, *265* (1), 19–26.

(358) Palma, P.; Palma, V. L.; Matos, C.; Fernandes, R. M.; Bohn, A.; Soares, A. M. V. M.; Barbosa, I. R. Effects of atrazine and endosulfan sulphate on the ecdysteroid system of *Daphnia magna*. *Chemosphere.* **2009**, *74* (5), 676–681.

(359) Ye, W.-H.; Wen, Y.-Z.; Liu, W.-P.; Wang, Z.-Q. Effects of bifenthrin on *Daphnia magna* during chronic toxicity test and the recovery test. *Journal of Environmental Sciences (China).* **2004**, *16* (5), 843–846.

(360) Owen, R.; Galloway, T. S.; Hagger, J. A.; Jones, M. B.; Depledge, M. H. Biomarkers and environmental risk assessment: Guiding principles from the human health field. *Marine Pollution Bulletin.* **2008**, *56* (4), 613–619.

(361) Vainio, H. Use of biomarkers in risk assessment. *International journal of hygiene and environmental health*. **2001**, *204* (2–3), 91–102.

(362) Bonassi, S.; Au, W. W. Biomarkers in molecular epidemiology studies for health risk

prediction. *Mutation Research.* **2002**, *511* (1), 73–86.

(363) Bonassi, S.; Neri, M.; Puntoni, R. Validation of biomarkers as early predictors of disease. *Mutation Research.* **2001**, *480–481*, 349–358.

(364) Firestein, G. S. A biomarker by any other name... *Nature Clinical Practice Rheumatology.* **2006**, *2* (12), 635–635.

(365) Bertagnolli, M. M. The forest and the trees: pathways and proteins as colorectal cancer biomarkers. *Journal of Clinical Oncology.* **2009**, *27* (35), 5866–5867.

(366) Zhang, D. Y.; Ye, F.; Gao, L.; Liu, X.; Zhao, X.; Che, Y.; Wang, H.; Wang, L.; Wu, J.; Song, D.; et al. Proteomics, pathway array and signaling network-based medicine in cancer. *Cell Division.* **2009**, *4* (1), 20.

(367) Ptitsyn, A. A.; Weil, M. M.; Thamm, D. H. Systems biology approach to identification of biomarkers for metastatic progression in cancer. *BMC Bioinformatics.* **2008**, *9* (Suppl 9), S8.

(368) MAQC Consortium, M.; Shi, L.; Reid, L. H.; Jones, W. D.; Shippy, R.; Warrington, J. A.; Baker, S. C.; Collins, P. J.; de Longueville, F.; Kawasaki, E. S.; et al. The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements. *Nature Biotechnology.* **2006**, *24* (9), 1151–1161.

(369) Shi, L.; Campbell, G.; Jones, W. D.; Campagne, F.; Wen, Z.; Walker, S. J.; Su, Z.; Chu, T.-M.; Goodsaid, F. M.; Pusztai, L.; et al. The MicroArray Quality Control (MAQC)-II study of common practices for the development and validation of microarray-based predictive models. *Nature Biotechnology.* **2010**, *28* (8), 827–838.

(370) Su, Z.; Łabaj, P. P.; Li, S.; Thierry-Mieg, J.; Thierry-Mieg, D.; Shi, W.; Wang, C.; Schroth, G. P.; Setterquist, R. A.; Thompson, J. F.; et al. A comprehensive assessment of RNA-seq accuracy, reproducibility and information content by the Sequencing Quality Control Consortium. *Nature Biotechnology.* **2014**, *32* (9), 903–914.

(371) Brazma, A.; Hingamp, P.; Quackenbush, J.; Sherlock, G.; Spellman, P.; Stoeckert, C.; Aach, J.; Ansorge, W.; Ball, C. A.; Causton, H. C.; et al. Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. *Nature Genetics.* **2001**, *29* (4), 365–371.

(372) Snape, J. R.; Maund, S. J.; Pickford, D. B.; Hutchinson, T. H. Ecotoxicogenomics: the challenge of integrating genomics into aquatic and terrestrial ecotoxicology. *Aquatic Toxicology.* **2004**, *67* (2), 143–154.

(373) Messerlian, C.; Martinez, R. M.; Hauser, R.; Baccarelli, A. A. “Omics” and endocrine-disrupting chemicals — new paths forward. *Nature Reviews Endocrinology.* **2017**, *13* (12), 740–748.

(374) Oliveira, E.; Barata, C.; Piña, B. Endocrine disruption in the omics era: New views, new hazards, new approaches. *Open Biotechnology Journal.* **2015**, *9* (1).

(375) Yang, O.; Kim, H. L.; Weon, J.-I.; Seo, Y. R. Endocrine-disrupting chemicals: Review of toxicological mechanisms using molecular pathway analysis. *Journal of Cancer Prevention.* **2015**, *20* (1), 12–24.

(376) Hutchinson, T. H. Small is useful in endocrine disrupter assessment—four key recommendations for aquatic invertebrate research. *Ecotoxicology.* **2007**, *16* (1), 231–238.

(377) Lafont, R.; Mathieu, M. Steroids in aquatic invertebrates. *Ecotoxicology.* **2007**, *16* (1), 109–130.

(378) Matthiessen, P.; Gibbs, P. E. Critical appraisal of the evidence for tributyltin-mediated endocrine disruption in mollusks. *Environmental Toxicology and Chemistry.* **1998**, *17* (1), 37–43.

(379) Fong, P. P.; Ford, A. T. The biological effects of antidepressants on the molluscs and crustaceans: A review. *Aquatic Toxicology.* **2014**, *151*, 4–13.

(380) Piña, B.; Casado, M.; Quiros, L. Analysis of gene expression as a new tool in ecotoxicology and environmental monitoring. *TrAC Trends in Analytical Chemistry.* **2007**, *26* (11), 1145–1154.

(381) Piña, B.; Barata, C. A genomic and ecotoxicological perspective of DNA array studies in aquatic environmental risk assessment. *Aquatic Toxicology.* **2011**, *105* (3–4), 40–49.

(382) Plahuta, M.; Tišler, T.; Toman, M. J.; Pintar, A. Toxic and endocrine disrupting effects of wastewater treatment plant influents and effluents on a freshwater isopod *Asellus aquaticus* (Isopoda, Crustacea). *Chemosphere.* **2017**, *174*, 342–353.

(383) SCENIHR (Scientific Committee on Emerging and Newly Identified Health Risks). Addressing the new challenges for risk assessment: Discussion paper approved for public consultation in view of receiving feedback from stakeholders for its further development. **2012**.

(384) Villeneuve, D. L.; Crump, D.; Garcia-Reyero, N.; Hecker, M.; Hutchinson, T. H.; LaLone, C. A.; Landesmann, B.; Lettieri, T.; Munn, S.; Nepelska, M.; et al. Adverse outcome pathway (AOP) development I: strategies and principles. *Toxicological Sciences.* **2014**, *142* (2),

312–320.

(385) Villeneuve, D. L.; Crump, D.; Garcia-Reyero, N.; Hecker, M.; Hutchinson, T. H.; LaLone, C. A.; Landesmann, B.; Lettieri, T.; Munn, S.; Nepelska, M.; et al. Adverse outcome pathway development II: best practices. *Toxicological Sciences.* **2014**, *142* (2), 321–330.

(386) Sturla, S. J.; Boobis, A. R.; FitzGerald, R. E.; Hoeng, J.; Kavlock, R. J.; Schirmer, K.; Whelan, M.; Wilks, M. F.; Peitsch, M. C. Systems toxicology: From basic research to risk assessment. *Chemical Research in Toxicology.* **2014**, *27* (3), 314–329.

(387) Ankley, G. T.; Villeneuve, D. L. The fathead minnow in aquatic toxicology: Past, present and future. *Aquatic Toxicology.* **2006**, *78* (1), 91–102.

(388) ECETOC (2008). Workshop report 11 - workshop on the application of 'omics in toxicology and ecotoxicology: case studies and risk assessment http://www.ecetoc.org/publication/workshop-report-11-workshop-on-the-application-of-omics-in-toxicology-and-ecotoxicology-case-studies-and-risk-assessment/ (accessed Nov 18, 2017).

(389) ECETOC (2010). Workshop report 19 - omics in (eco)toxicology: case studies and risk Assessment http://www.ecetoc.org/publication/workshop-report-19-omics-in-ecotoxicology-case-studies-and-risk-assessment/ (accessed Nov 18, 2017).

(390) Xu, E. G.; Mager, E. M.; Grosell, M.; Hazard, E. S.; Hardiman, G.; Schlenk, D. Novel transcriptome assembly and comparative toxicity pathway analysis in mahi-mahi (*Coryphaena hippurus*) embryos and larvae exposed to Deepwater Horizon oil. *Scientific Reports.* **2017**, *7*, 44546.

(391) Ip, J. C. H.; Leung, P. T. Y.; Ho, K. K. Y.; Qiu, J. W.; Leung, K. M. Y. De novo transcriptome

assembly of the marine gastropod *Reishia clavigera* for supporting toxic mechanism studies. *Aquatic Toxicology.* **2016**, *178*, 39–48.

(392) Wang, S.; Gribskov, M. Comprehensive evaluation of *de novo* transcriptome assembly programs and their effects on differential gene expression analysis. *Bioinformatics.* **2016**, *26* (3), btw625.

(393) Hecker, M. (2016). *Non-model species in ecological risk assessment. In: A systems biology approach to advancing adverse outcome pathways for risk assessment. Murphy C. and Reyero N.*; Eds., Springer, New York, U., Ed.

(394) Kramer, V. J.; Etterson, M. A.; Hecker, M.; Murphy, C. A.; Roesijadi, G.; Spade, D. J.; Spromberg, J. A.; Wang, M.; Ankley, G. T. Adverse outcome pathways and ecological risk assessment: Bridging to population-level effects. *Environmental Toxicology and Chemistry.* **2011**, *30* (1), 64–76.

(395) Doering, J. A.; Farmahin, R.; Wiseman, S.; Beitel, S. C.; Kennedy, S. W.; Giesy, J. P.; Hecker, M. Differences in activation of aryl hydrocarbon receptors of white sturgeon relative to lake sturgeon are predicted by identities of key amino acids in the ligand binding domain. *Environmental Science & Technology.* **2015**, *49* (7), 4681–4689.

(396) Karchner, S. I.; Franks, D. G.; Kennedy, S. W.; Hahn, M. E. The molecular basis for differential dioxin sensitivity in birds: Role of the aryl hydrocarbon receptor. *Proceedings of the National Academy of Sciences of the United States of America.* **2006**, *103* (16), 6252–6257.

(397) Drwal, M. N.; Siramshetty, V. B.; Banerjee, P.; Goede, A.; Preissner, R.; Dunkel, M. Molecular similarity-based predictions of the Tox21 screening outcome. *Frontiers in*

Environmental Sciences. **2015**, *3*, 54.

(398) Validated and Accepted Alternative Methods http://alttox.org/mapp/table-of-validated-and-accepted-alternative-methods/ (accessed Apr 26, 2017).