Adopting a gene regulatory network approach to investigate toxicity through the adaptive stress response in teleost fish species.

Submitted by **Amy Louise Foreman** to the

University of Exeter

as a thesis for the degree of

Doctor of Philosophy by Research in Biological Sciences.

31st March 2018.

(Signature)……………………………………………………………………

ABSTRACT

Given current risks of pollutant exposures in aquatic environments, there is a growing need to generate reliable computational risk assessment methods to establish how adverse outcomes can be produced across exposure organisms. The adaptive stress response is widely targeted by pollutants of concern and includes transcription factors including nuclear factor (erythroid-derived 2)-like 2 (Nrf2), hypoxia inducible factor (HIF-1α), heat shock factor (HSF1), nuclear factor kappa-light-chain-enhancer of activated -B cells (NFkB), metal transcription factor 1 (MTF1), the aryl hydrocarbon receptor (AhR) and tumor protein P53 (P53). While these TFs are known to be activated by distinct inducers, less is understood about the regulatory links between factors, particularly at the transcription factor (TF) DNA-binding level.

In this thesis, a gene regulatory network (GRN) of adaptive-stress response factors that are key targets of chemical toxicity was constructed based on experimental evidence from mammalian cell-lines. The GRN was modeled using boolean logic and this identified a number of response outcomes that could be attributed to the activation of pathways including antioxidant defence processes and glucose metabolism. The GRN model illustrated that the activation of Nrf2, HIF-1α, AhR, MTF1 and HSF1 led to the same adverse outcomes, suggesting canalisation in stress response pathways.

The ability to use GRNs across different species is widely supported by the identification of TF binding sites (TFBS) within target genes. To assess the efficiency of using the mammalian GRN across teleost fish species, a comprehensive analysis of validated binding sites for the AhR, MTF1, HIF-1α and Nrf2 was conducted across fish-species in comparison to the mammalian consensus binding sequence. This showed variations in binding site composition across validated TFBS for HIF-1α and Nrf2 in fish compared to the mammalian consensus, preventing the identification of the functional sequences for these factors using traditional methods. To establish if such changes affected the efficiency to predict positive downstream target genes for Nrf2 and HIF-1α in mammals and across teleost fish species, random

forest classification models were used to compare the efficiency of multiple positional weight matrices (PWM) motifs of TFBS for Nrf2 and HIF-1α. Whilst the result from this analysis identified discrepancies in the ability to predict target genes based on the mammalian motif file used, mammalian motifs were able to predict target genes across fish species. Validated binding sites in fish species were then aligned to generate PWM motifs and sites were predicted across shared target genes hsp70 and hmox1 using both fish based and mammalian based models. This showed that whilst there was some overlap in identified sites across species, fish-specific motifs identified unique sites from mammalian models.

To validate the GRN, gene-expression responses across exposures traditionally associated with activating distinct adaptive stress response factors were collated across the literature. This showed support for some of the key responses identified in the model. Chemical exposure studies were then undertaken *in vivo* in embryo-larval zebrafish (2 and 4 dpf) to identify potential connectivity between the TFs NFkB, MTF1 and HIF-1α with Nrf2, a key factor in the adaptive stress-response and a regulator of antioxidant response processes. The inducer of Nrf2, tert-butylhydroquinone (tBHQ), was used to determine if there was a change in transcriptional output of *mtf1*, *hif1a* and *nfkb1* over time and with exposure concentration. This showed a significant difference in expression for *nfkb1* and alterations in expression of *mtf1* over prolonged exposure scenarios. In addition, the developmental expression of *nrf2a*, *mtf1*, *hif1α* and *nfkb1* from 2 hpf to 96 hpf showed differences between transcript levels with *hif1α* and *nfkb1* having the highest levels of expression compared to *nrf2a* and *mtf1*.

Overall, the research presented in this thesis provides a novel approach to assess the initiation of adaptive stress-response factors from molecular interactions. The research goes some way in establishing the feedback loops and connections between NFkB, MTF1, Nrf2, AhR, HIF-1α, HSF1 and P53. In doing so, the model generated in this thesis provides a novel approach of establishing outcomes under toxicant exposures.

Acknowledgements

Firstly, I would like to thank my supervisors for their support and patience. To Charles Tyler and Tetsu Kudoh for your guidance, discussions and for always having an open door. Thank you Tetsu for taking me on as an undergraduate and showing me the ropes. Thanks also go to Aya - your continuous passion and enthusiasm for your work is always inspiring. And to Anke - for your encouragement, top-tips and friendship; you've been instrumental at getting me through to the end and I am so grateful for all the support you have given me. I'd also like to thank Greg and the ARC team for keeping the zebrafish army in tip-top shape and to John Dowdle for bringing a smile to my face every morning.

This PhD adventure would have not been the same without the 201 family, past and present. It has been wonderful to work with you all over the last few years and I'm particularly grateful to Kat, Katie, Krista, Bas and Ross for your constant positivity and irreplaceable company. To Ruth, Shelly, Uncle Tony, Sara and Moreman, I'd like to thank you for making the Cave the most magical respite imaginable and for always providing such quality entertainment. To Jenny and Aoife, for your continuous and unwavering support and advice; you will always be setting my scientific gold standard. Thank you all for listening to my endless stories and constant demands for tea.

Rich, my truly charming desk-mate, housemate and GP sidekick, how you managed to put up with my mess, singing and stressful Spanish guitar songs for the last few years is a truly remarkable achievement - you deserve a knighthood! And to Gemma and Laura for never tiring of the "fish chat". I wouldn't have got to the end without the three of you.

Finally, I am incredibly lucky to have such a supportive family, thank you all for everything you have done for me. To Grandma, Grandad, Mum, Dad and Liz, I'd like to dedicate this work to you.

Table of Contents

List of Figures:

List of Tables:

List of Supplementary Figures and Tables:

List of General Abbreviations:

AO	acradine orange
AOP	Adverse outcome pathway
ARNT	Aryl-hydrocarbon nuclear translocator
ATAC-seq	Assay for transposable chromatin
AUC	Area under the curve
bp	Base-pairs
ChIP-seq	Chromatin-immonpercipitation
Ct	cycle threshold
DMSO	Dimethyl sulfoxide
DNA	deoxyribonucleic acid
EpRE	Electrophile response element
GFP	Green fluorescent protein
GLM	Generalised linear model
GRN	Gene-regulatory network
GSH	Glutathione (reduced)
GSSG	Glutathione (oxidised)
GSTp	Glutathione-s-transferase Pi
HIF1-a	Hypoxia inducible factor 1a
HMM	Hidden Markov Model
HMOX1	Heme-oxygenase 1
hpf	hours post-fertilization
HRE	Hypoxia response element
HSE	Heat-shock response element
HSF1	Heat-shock factor1
IKK	IkB Kinase
kb	Kilobase
KE	Key-event
Keap1	Kelch-like ECH associated protein 1
LUC	Luciferase
MIE	Molecular initiating event
MOA	Mode of action
MRE	Metal response element
MSA	Multiple-species alignment
MTF1	Metal transcription factor 1
NFkB	Nuclear-factor Kappa B
Nrf2	Nuclear factor (erythroid-derived 2)-like 2
ODE	Ordinary differential equation
OS	Oxidative stress
P53	Tumour protein P53
PFM	Positional-factor matrix
PWM	Positional weight matrix

QSAR	Quantitative structural activity relationship
RF	Random forest
RNA	ribonucleic acid
ROC	receiver-operator characteristics curve
RT-PCR	real-time polymerase-chain reaction
SNP	Single nucleotide polymorphism
SQSTM1	Sequestosome 1
tBHQ	Tert-butyl hydroquinone
TF	Transcription factor
TFBS	Transcription factor binding site
WISH	Whole-mount in situ hybridization
XRE	Xenobiotic response element

List of Species Names:

Antarctic icefish	*Chionodraco hamatus*
Stone Loach	*Barbatula barbatula*
African green monkey	*C. aethiops sabaeus*
Zebrafish	*Danio rerio*
Northern pike	*Esox lucius*
Atlantic killifish	*Fundulus heteroclitus*
Threespined stickleback	*Gasterosteus aculeatus*
Human	*Homo sapiens*
Mouse	*Mus musculus*
Coho Salmon	*Oncorhynchus kisutch*
Rainbow trout	*Oncorhynchus mykiss*
Nile Tilapia	*Oreochromis niloticus*
Medaka	*Oryzias latipes*
Fathead minnow	*Pimephales promelas*
Rat	*Rattus rattus*
Roach	*Rutilus rutilus*
Atlantic Salmon	*Salmo salar*
Brown trout	*Salmo trutta*
Arctic charr	*Salvelinus alpinus*
Fugu	*Takifugu rubripes*

Chapter 1

General Introduction

1.1 Introduction

Freshwater environments are considered a major sink for pollutants derived from agricultural, industrial and domestic sources[1]. The advanced engineering of novel pharmaceuticals and materials such as nanoparticles[2] has created an increasing risk of pollution from substances which have unknown toxicities in environmental systems. Additional pressures, including changes in temperature[3] and increases in hypoxia[4] have added to the growing concern for adverse consequences in aquatic systems, particularly for organisms that are of significant ecological or economical importance. As such, freshwater environments are estimated to have had a 76% decline in biodiversity from 1970 to 2012, the largest decline of any ecosystem[1].

Whilst it is not feasible to test the toxic potential of every chemical or pollutant combination, there is a clear acknowledgement at the international level for the need to develop robust and high-throughput (HT) environmental risk assessment methods in accordance with the growing complexity of pollutant exposures[5]. It has been widely established that progress in the field requires the development of predictive platforms which test the toxic potential of multiple chemical groups at various levels of biological organisation, an aim only likely to be achieved by combining advances in computational modeling techniques with experimental evidence derived from field-realistic exposure scenarios[5,6].

Currently, regulatory organisations including the Environmental Protection Agency (EPA) and the Organisation for Economic growth and Development (OECD) have adopted frameworks that rely on integrated approaches to testing and assessment (IATAs), combining the available knowledge on chemical exposures from molecular to population level effects [7] . IATAs, including Adverse outcome pathway (AOP) and mode of action (MoA) frameworks, have been established with the long-term aim of conducting environmental risk assessments across species *in silico* [7], reducing both the ethical concerns associated with toxicity testing on multiple organisms and

aiding *in silico* predictions of adverse outcomes (AOs) across chemical groups [7].

At the basis of both frameworks are pollutant-induced molecular initiating events (MIEs), defined as "the initial interaction between a molecule and a biomolecule or biosystem that can be causally linked to an outcome via a pathway" [8]. MIEs include processes such as receptor binding which initiates the transcription of specific cohorts of genes, an example being toxicants that target gene expression in the endocrine system through the oestrogen receptor. Such processes lead to changes at the cellular and tissue level, so called key events (KEs), and subsequently lead to observable AOs, be it chronic, such as DNA damage, or acute, such as a general decreases in biological activity (narcosis)[9], to the whole organism and population (Figure 1.1). Though both frameworks are conceptually similar, MoAs include outcomes up to the individual level whereas AOPs include dose-responses and outcomes at the population level [10] (Figure 1.1).
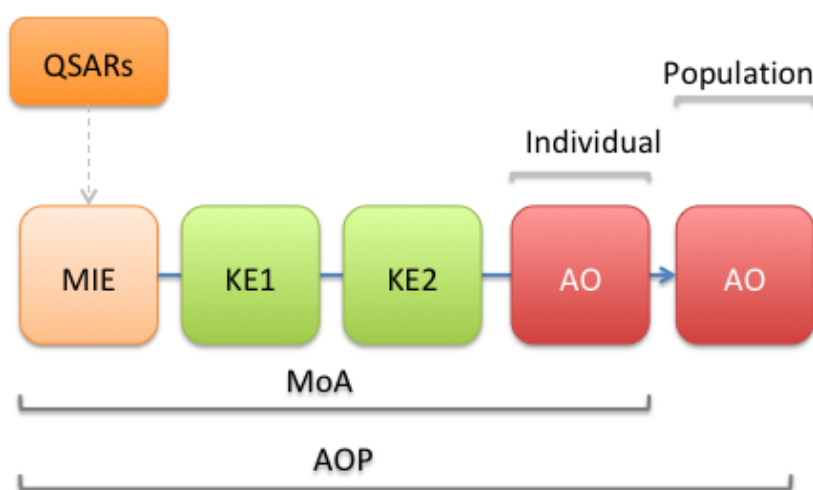
Figure 1.1: Adverse outcome pathway/ Mode of Action frameworks.
adapted from Perkins et al. 2011. Chemical toxicants activate molecular initiating events (MIEs) that cause measurable key events (KE) processes. This can lead to an adverse outcome (AO) at the whole organism or population level. MIEs can be predicted through quantitative structural activity relationships (QSARs).

Given the large number of potential pollutants, predictive toxicology for chemicals relies on quantitative structural activity relationships (QSARs), a "read-across" method which predict that chemicals with similar structures will cause toxic effects through the same MIEs [7]. Although this has been shown to be accurate in some instances, the method is limited by assuming that expected molecular interactions, such as receptor-mediated gene expression, act on only one downstream process, whereas in reality, multiple or different events may be triggered for a novel compound compared with a structurally similar counterpart [7]. In addition, the actions of metabolites, which could cause different biological interactions to the original compound, are as yet not well understood and even where known, are rarely considered within QSAR predictions.

The adoption of IATAs in toxicity testing, where knowledge is assembled for a range of chemicals and outcome processes, are likely to inform and improve on the outcomes of QSARs models[7]. However, the weight of evidence (WOE) between the compound and the observed outcome(s) as well as the transferability of the response across species needs to be considered. WOE for AOP/MoAs are assigned by 5 factors adopted from the Bradford-Hill considerations and comprise of the consistency, essentiality of key events, temporal, dose response and biological concordance as described in Table 1.1 [11]. The inclusion of evidence scores acknowledges the complexity of toxicity responses where, particularly at the molecular level, it can be challenging to directly determine the events that lead to observable adverse outcomes. The OECD's AOP Knowledge Base (AOP-KB) has been developed as a tool to compile evidence across research organisations and allows for the input of MIEs, KEs and AOs for chemical toxicants across vertebrates with respective WOE scores [12].

However, despite these developments in predictive toxicology, the integration of novel experimental methods widely used in other fields is lacking. In particular, the last decade has seen significant advances in the knowledge of gene regulation, which form the basis of MIEs in all AOP and MoA frameworks. Developments in sequencing technologies, epigenetics and

transcription factor interactions has led to a more robust understanding of gene-expression dynamics and regulatory cascades which could provide significant insight into toxic effects for both singular and mixture effects.

Bradford-Hill Considerations	Consideration
Consistency	Pattern of effects across species/strains/organisms/test systems as expected.
Essentiality of Key Events	Reversibility of events if dosing is stopped or a KE prevented.
Temporal Concordance	KE observed in hypothesized order.
Dose-Response Concordance	KE observed in doses bellow/similar to the adverse effect.
Biological Concordance	MoA in agreement with broader biological knowledge.

Table 1.1: Bradford-Hill considerations for MoA and AOP frameworks (*adapted from Meek et. al. 2014*)

This thesis investigates a novel approach to addressing the understanding of gene-regulatory cascades in a toxicology setting focusing on the adaptive stress response; an evolutionary conserved process across aerobic organisms, describing pathways which mitigate against adverse effects such as DNA damage to maintain cellular homeostasis [13]. Biomarkers of the stress-response have become standpoint measures in ecotoxicology testing, with processes such as DNA-damage, lipid peroxidation and in severe cases, cell death determined under chemical exposures *in vitro* and *in vivo* [14]. Oxidative stress is a primary activator of the adaptive stress-response and arguably comprises the most common outcome to pollutant exposures in vertebrates [13].

1.2 The oxygen paradigm

All aerobic life forms rely on the mitochondrial electron transport chain to produce adenosine 5'-triphosphate (ATP) through a series of steps that utilise O_2 [15]. However, aerobic respiration produces the bi-product superoxide (O_2) from 1-2% of electrons 'leaking out' of the electron transport chain and binding to oxygen[15]. The conversion of O_2 to O_2^- is the first of a series of 4 reduction steps in the production of reactive oxygen species (ROS) [15]. ROS are defined as molecules containing oxygen that are either oxidising agents and/or are easily converted into free radicals that contain one or more unpaired electrons[16]. On leaving the mitochondria, the antioxidant superoxide dismutase (SOD) reduces O_2^- to produce hydrogen peroxide H_2O_2, which can be converted to the hydroxyl radical $HO^·$ through the Haber-Wiess and Fenton reaction (Figure 2).

Although ROS can have beneficial roles in processes ranging from cell signaling to host defence they can also cause damaging effects to molecules through interactions that add (oxidising) or subtract (reducing) electrons [3]. Excess ROS can result in DNA damage and lipid peroxidation, which can be largely attributed to the OH^- radical, and are responsible for diseases such as neurodegenerative disorders and cancer [17]. This creates the paradox that O_2 is required for survival but can itself be a toxic compound[18].

The evolutionary conserved antioxidant defence system, broadly defined as "any substance that delays, prevents or removes oxidative damage to a target molecule" [15], counteracts the damaging effects mediated by free radicals and maintains ROS at low-levels. However, an imbalance in the levels of antioxidants to ROS in favor of the latter, causes a condition termed oxidative stress[18]. Antioxidants are required to neutralise reactive molecules, largely by initiating reduction reactions or acting as electron acceptors, and in turn maintain the redox balance through keeping a stable ratio of oxidised to reduced molecules[19]. Central to this is the antioxidant and scavenger molecule glutathione that exists in either a reduced form as GSH or oxidised as GSSG. The ratio of GSH:GSSG is widely used as an indicator of the severity of oxidative stress and is termed the redox status[19], a measure of the level of antioxidants to ROS. The basal ratio of GSH:GSSG can in-part be

controlled by the cofactors nicotinamide adenine dinucleotide phosphate (NADPH:NADP), which reduces GSSG in the presence of glutathione reductase (GR) to GSH[19] (Figure 2). In turn, NADP is reduced back to NADPH by glucose-6-phosphate dehydrogenase (G6PD) derived from glycolysis[20] (Figure 2). In addition to ROS from aerobic respiration, nitric oxide NO, formed from L-Arginine (L-Arg), an amino acid resulting from metabolic processes, can bind to O_2^- to produce $ONOO^-$, a highly reactive ROS[21]. The dietary intake of antioxidants, including vitamin C and bilirubin, can support antioxidant defences in order to maintain redox homeostasis where production of ROS occurs due to general biological activity[21].
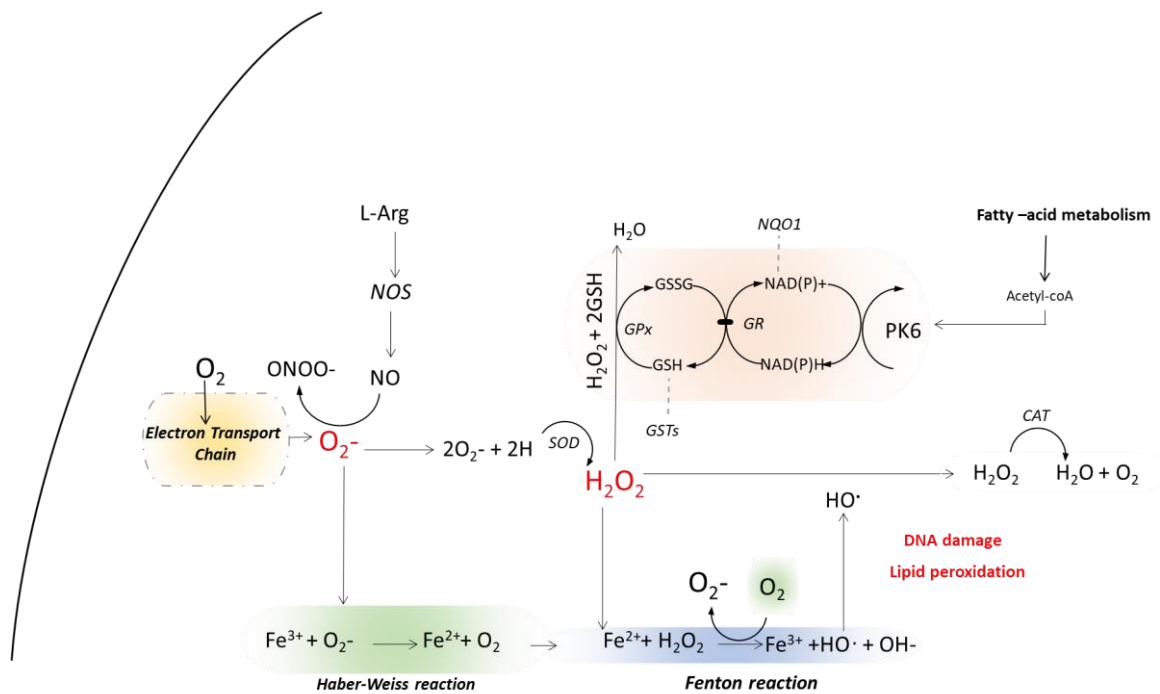
Figure 1.2: Biological production of ROS. Superoxide O_2^- is released from the electron transport chain and is converted to hydrogen peroxide H_2O_2 through superoxide dismutase (SOD). O_2^- also reacts with the Fe^{3+} to produce Fe^{2+} in the Haber-Weiss reaction[22]. H_2O_2 is then bound to Fe^{2+} to produce the OH^- and $OH^.$ in the Fenton reaction[22]. H_2O_2 is reduced by glutathione peroxidase (GPx) to produce H_2O and GSSG. GSSG is reduced to GSH by glutathione reductase (GR), converting NAD(P)H to NADP[19]. NADP is reduced by glucose-6-phosphate dehydrogenase (G6PD) [20]. L-Argenine (L-Arg) and nitric oxide synthase (NOS) produces ON[21]. ON reacts with O_2^- to produce $ONOO^-$ [21]. Figure created from information derived from the literature as cited.

1.3 Fish model organisms in ecotoxicology

An estimated 90% of all fish species belong to the teleost subfamily which comprise 50% of all vertebrates [23]. The divergence of teleost fish from mammals occurred with a whole genome duplication (WGD) event leading to genes conserved in mammals often exist as two copies, or paralogs, within fish species [24]. These duplicated genes are evolutionary preserved through processes such as sub-functional partitioning (where gene paralogs each have specific functions that are conserved in the mammalian counterpart [24]), neofunctionalization (where genes obtain novel functions) and dosage selection (where genes are preserved to maintain balance between other interconnected components) [25]. Evidence suggests additional gene duplication events have occurred within fish species, through further WGD or processes such as unequal cross-over and retrotransposition [26]. Copy number variants (CNVs) can exist for genes between species for example, the common carp (*Cyprinus carpio*) is predicted to have undergone a further *genome duplication event compared to the zebrafish [26]. This is also reflected in genome size, for example being 342 Mb in the fugu but 1.5 Gb in the zebrafish genome [26]. WGD event caused in the divergence of teleost fish has allowed for larger evolutionary plasticity by providing a greater amount of genetic information for adaption and this is believed to be at least partially responsible for the high levels of diversity seen within the vertebrate group[25].There are huge variations in niches filled by teleost fish and this is reflected by differences in factors such as lifespan and reproductive strategy (REF). Given the evolutionary distances between teleosts, it is therefore questionable whether the findings derived from one fish species under chemical toxicity can be applied to all fish species.*

A number of model organisms have been adopted to assess chemical toxicity in aquatic environments and include the zebrafish (Danio rerio), rainbow trout (Oncorhynchus mykiss), three-spined stickleback (Gasterosteus aculeatus), Japanese medaka (Oryzias latipes), fathead minnow (Pimephales promelas), Atlantic killifish (Fundulus heteroclitus) and brown trout (Salmo trutta). These belong to distinct taxonomic orders representing belinoforms (medaka),

cypriniforms (zebrafish and fathead minnow), salmoniforms (brown trout) and percomorpha (stickleback)[25]. However, given that the adaptive stress response evolved prior to the divergence of prokaryotes and eukaroytes, these processes are highly conserved across vertebrates [27].

A large proportion of research has been conducted on laboratory strains of zebrafish, which have the advantage of translucent embryos and rapid developmental lifespans[28]. The use of the zebrafish model not only in ecotoxicology but also across the life sciences has provided a wealth of information regarding the genetics and characterization of developmental stages as well as the development of a number of model lines that incorporate gene reporter systems[28]. From a toxicology perspective, experimental evidence derived from zebrafish studies have provided an in-depth understanding of exposure effects at the molecular level. However, laboratory strains of zebrafish, which suffer from some level of inbreeding, can have different levels of susceptibility in comparison to wild-type strains [29]. Zebrafish models are therefore used as a platform to identify mechanisms of toxicity, which can, once established, be supported by environmentally relevant exposures in either wild-type strains or in species that are more challenging to maintain under laboratory conditions when necessary.

1.4 The transferability of experimental evidence from fish-specific exposures to mammals.

Despite the evolutionary distances within teleost species and across to mammals, experimental evidence derived from studies on zebrafish have provided advances in biomedical science and toxicology. Comparisons between exposure results between zebrafish and rodents has shown high levels of conservation in response phenotypes [30]. 71 % of human genes have atleast one zebrafish ortholog and zebrafish disease models are used as a reliable indicator of the molecular processes behind diseases across vertebrates [31]. The similarities between zebrafish and mammals have

allowed it to be a successful model to study oestotoxicity [32], blood cell development and immune responses [33].

1.4 Pollutants and environmental stressors as a source of ROS in aquatic systems

While the production of ROS is a natural byproduct of cellular processes, a range of pollutants and environmental stressors can cause oxidative stress. Metals[22], chemical toxicants and/or their metabolites in addition to environmental changes in temperature[3] and increases in hypoxia[34] have been shown to induce oxidative stress. Biomarkers of stress-response processes are widely used in ecotoxicology to assess for oxidative damage. This section discusses the general MoAs for selected pollutant groups that are of most concern in freshwater environments. However, it should be noted that the severity of adverse outcomes discussed in this section is dependent on parameters including the dose and duration of the inducer as well as the route of uptake and life-stage of the targeted organism. In addition, responses to infection where macrophages induce the "oxidative burst" and release ROS to destroy pathogens, is a major endogenous source of ROS [35] but is beyond the scope of this thesis.

1.4.1 Biomakers for adverse outcomes induced by ROS

A range of biomarkers have been adopted to monitor the occurrence of oxidative stress in aquatic organisms and these provide representative measures of outcomes from antioxidant defence processes to cell death (Table 1.2). In some cases, markers can be used *in vivo,* such as staining dyes indicating apoptosis by binding to DNA including propidium iodine (PI), acridine orange and the Terminal deoxynucleotidyl transferase (dUTP) nick end labeling (TUNEL) assay. In other cases, biomarkers can only be used on cellular homogenates such as RT-qPCR assays. As oxidative stress is widely conserved, the same markers are often used across both vertebrates and invertebrates [14].

Process	Biomarker
Antioxidant defence	Gene expression analysis of antioxidants (e.g. SOD, GST, NQO1, HMOX1) using techniques such as **RT-qPCR.**
Lipid Peroxidation	Measurements of **isoprostanes** produced from peroxidation of polyunsaturated fatty acids can also be measured using gas-chromatography mass spectrometry (**GC-MS**)[36].
DNA-damage	**Comet** assay measures DNA strand breaks by electrophoresis. DNA from individual nuclei is unwound and run on a gel. Damaged segments form a tail that is proportional to the amount of damaged DNA. Alternatively, 8-hydroxy-2-deoxyguanosine (**8-OHdG**) can be measured using GC-MS as a maker for DNA oxidation[37].
Hydrogen Peroxide	Hydrogen peroxide probes such as **HyPer** where H_2O_2 is bound to a fluorescent protein to visualise responses *in vivo.* Alternatively, dichloroflourescin diacete (**DCFDA**) can be used which is oxidised by ROS to produce a fluorescent compound[38].
Cell death (apoptosis and necrosis)	**Acradine Orange** stains for pre-apoptotic and apoptotic cells in-vivo by binding to nucleated DNA. The fluorescent maker propidium iodine (**PI**) can be used which enters permeable cells due to necrosis and apoptosis and binds to nucleated DNA. Terminal deoxynucleotidyl transferase dUTP nick end labeling (**TUNEL assay**) can stain DNA strained breaks formed as a result of apoptosis.[36].
Inflammation	Gene expression analysis of inflammatory response genes e.g. (IL6, IL8) using techniques such as **RT-qPCR**.

Table 1.2 Biomarkers of oxidative stress responses. Oxidative damage can have wide ranging outcomes in cell systems that are measured using representative biomarkers of response processes as shown.

1.4.2 Metal pollution

Metal pollution is ubiquitous in the aquatic environment and caused by industrial activities such as mining [22]. Although essential metals are required for normal cellular functioning, pollutant exposures can lead to detrimental effects, causing mutagenesis and tetragenesis[22], widely associated with the formation of ROS through the action of metal ions. Underlying toxicity to some metals is an ability for ions to act as catalysts in the Fenton reaction[39]. Chromium, copper, titanium, cobalt and vanadium have all been identified as producing OH^- through reactions with H_2O_2 [39]

(Figure 1.3). In addition, metal ions can also act as catalysts in the Haber-Weiss reaction, inducing HO^- production through reducing O_2^- and reacting with H_2O_2 [39] (Figure 1.3).

Silver ions have been identified as highly toxic and the release of silver has been tightly controlled in wastewater management following high levels of contamination from the photography industry in the 1970s[40]. $AgNO_3^+$ inhibits the ion exchange of Na^+ and Cl^- in the gills of fish leading to a decrease in osmoregulation, a process coupled with the induction of a stress response[41]. In addition, silver and other redox-unreactive metals can react with sulfurhydl groups on cellular proteins to produce thiol radicals[41]. Damaging effects of metals can also be removed through interactions with sulfhydryl groups of metallothioniens (MTs), low-molecular weight proteins which have metal binding capacity[22].

More recently, engineered metal nanomaterials, which have distinct properties to their bulk counterparts, such as an increased surface-area to volume ratio[42], have been shown to induce toxic effects via oxidative stress pathways. Metal nanoparticles are increasingly being used in domestic applications, such as anti-microbial silver nanoparticles in socks, and in the pharmaceutical industry[42]. Silver[43,44] and zinc-oxide[45] nanoparticle exposures in zebrafish embryos initiated the expression of antioxidant defence genes and the production of ROS is considered a likely MOA for nanoparticle toxicity[2].

1.4.3 Chemical toxicants

The toxicity of chemical pollutants can be explained using the Verharr classification scheme which associates MOA with chemical structure[9]. Chemical MOAs are broadly grouped as narcotics, polar narcotics, electrophiles or as acting through specific molecular targets (specifically-acting). In some cases, a chemical can belong to multiple groups for example oestrogenic chemicals are considered specifically-acting by causing receptor-mediated toxicity but their metabolites are also electrophilic.

27

1.4.4 Electrophiles

 Electrophilic molecules are defined as molecules that have one or more electron poor atoms which can accept electrons from nucleophiles, molecules that are electron rich to form covalent bounds[46]. Electrophiles have carcinogenic potential and can bind to DNA causing mutagenesis[47]. They fall into two distinct categories based on the selectivity of nucleophilic reactions; either soft, where a soft electrophile reacts with soft nucleophile or hard, where hard electrophiles react with hard nucleophiles[47]. Soft electrophiles exert toxicity through Michael-addition reactions, where nucleophiles are added to conjugated alkenes/alkyenes [46]. Such interactions include the covalent bounding of xenobiotics with GSH, reducing the overall cellular GSH:GSSG ratio[48] (Figure 3).

Chemical toxicants can be metabolized to electrophiles that are often short-lived but can drive toxic responses. Metabolites formed as a result of phase I drug metabolism initiated by the enzyme Cytochrome-P450 (CYP450) are often electrophilic (Figure 3)[49]. Electrophiles derived from the biotransformation of aromatic compounds such as polyaromatic hydrocarbons (PAHs) and oestrogens are termed quinones. Quinones are Michael acceptors [48] and their reduction by CYP450 leads to the release of O_2^-, initiating redox cycling and the eventual production of $\cdot OH$[48]. PAHs and oestrogens including 17-beta estradiol (E2) and bisphenol-A (BPA), which have the potential to form quinone compounds, are associated with initiating oxidative stress indicated through the up-regulation of antioxidants [50,51].

1.4.5 Specifically-acting toxicants

The specifically acting group of toxicants describes toxicity where chemicals act on individual cellular components such receptors to initiate responses. Receptor mediated toxicants are widespread MOAs particularly for pollutants derived from the pharmaceutical and pesticide industry. Many drugs are

specifically designed to target individual receptor or protein complexes in-order to initiate responses. However, receptors can be highly conserved across the animal kingdom leading to indirect toxicity in non-target organisms if released into the environment [52]. The glucocorticoid, androgen, oestrogen and thyroid receptors are all targets for environmental toxicants in fish, initiating responses that include the production of the yolk precursor protein, vitellogenin, and feminisation in the case of oestrogen[53]. Despite specifically acting toxicants not primarily initiating toxicity through oxidative stress, there is an increasing body of evidence supporting antioxidant induction following exposures to such compounds. For example, the MOA for the herbicide glyphosate is plant specific, but has been shown to up regulate antioxidant defence genes in fish [34]. It is therefore highly likely that oxidative stress responses, which can result from xenobiotic metabolism, will be initiated alongside specifically targeted response pathways.

1.4.6 Temperature Change

Acute changes in temperature can lead to the formation of ROS through altering the metabolic activity of the organism in accordance with thermodynamic theory and thus increase the consumption of O_2 and the rate of aerobic respiration[34]. This response has been widely shown across fish species where increased temperature has been identified as producing ROS[3]. In addition, acute cold stress has been shown to induce antioxidant defences in zebrafish [54].

1.4.7 Hypoxia

Hypoxia, where the level of O_2 is below that necessary for aerobic respiration, is predicted to rise due to increases in eutrophication and greater numbers of microbial blooms resulting from climate change[4]. There are several hypotheses of how ROS are produced as a result of hypoxia. A decrease in O_2 in the electron transport chain is predicted to result in an increased rate of electrons leaking from the mitochondria [34]. In addition, the enzyme xanthine

reductase is converted to xanthaine oxidase under hypoxic conditions and is able to produce ROS (Figure 1.3)[34] . Studies of hypoxia have identified biomarkers of oxidative stress including the upregulation of antioxidants across exposures in fish species[55–57].

1.4.8 Mixed-pollutant effects

Oxidative stress is a widespread phenomenon to pollutant exposures and changes in environmental conditions. Organisms will therefore be exposed to multiple inducers of oxidative-stress in field-realistic exposure scenarios with evidence that stress response processes are altered under mixture effects. Chemical exposures under hypoxia/temperature-change have shown differing responses depending on toxicant group. For example, the fold increase in the antioxidant proteins CAT, SOD and MT in zebrafish larvae exposed to cadmium was greater under higher temperatures [58] and hypoxia has been shown to decrease the toxicity of copper in zebrafish embryos[55]. In addition, ROS is widely associated with disease burden as a result of infection or as a pre-requisite to its acquisition.
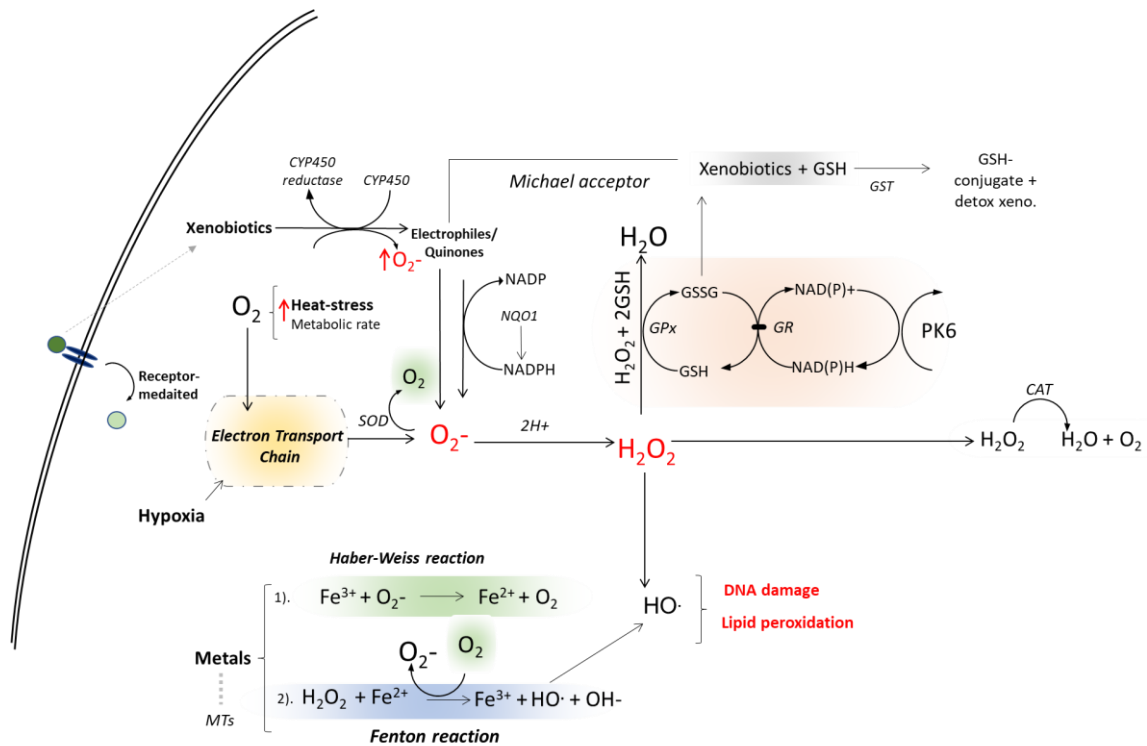
Figure 1.3. Production of ROS from pollutants and environmental change. Metals ions catalyse Fenton and Haber-Weiss reactions[22]. Higher temperatures increase the levels of free electrons causing increases in O_2^- [3]. Xenobiotics can be metabolised by cyptochrome-p450s (CYP450s) to produce electrophilic compounds which can bind to GSH and reduce overall GSH:GSSG concentration[48]. Alternatively, xenobiotics are reduced by NADP to NADPH using NAD(P)H dehydrogenase (NQO1). Metabolism of xenobiotics releases O_2^-, which is converted to H_2O_2 or reduced to O_2 using superoxide (SOD) [48]. H_2O_2 is converted to $2H_2O + O_2$ by catalase (CAT) or glutathione peroxidase (GPx) reduces GSH – GSSG to give H_2O[17]. GSSG is returned back to GSH via glutathione reductase (GR), reducing NADP to NADPH[17]. Figure created from literature sources cited.

1.5 Consequences of oxidative stress at the whole-organism and population level in fish

Oxidative stress is widely regarded as being influential in life-history strategy and is hypothesised to be a driving force in evolution[59]. Often the responses identified in fish are widely conserved across vertebrates and include both behavioral and physiological outcomes.

1.5.1 Development

ROS act as a key signaling molecule in developmental processes with oxidised conditions associated with cell death and differentiation and reduced conditions associated with cell proliferation[60]. High levels of ROS are associated with the metabolic burden of development due to rapid increase in cell proliferation[59]. Basal GSH:GSSG levels have been shown to fluctuate throughout organogenesis in zebrafish, becoming oxidised from 2 hpf to 2 dpf before returning to reduced levels [61]. The role of ROS as a signaling molecule in developmental processes creates the possibility that early development is highly susceptible to ROS. In support of this, a large number of teratogens, chemicals that cause developmental defects, have been associated with initiating oxidative stress[62]. Life stage specific effects of oxidative stress with whole-organism and population level consequences are shown in Figure 1.4.

1.5.2 Behaviour and neurodegenerative diseases

Antioxidant-defence processes are key in preventing ROS-induced neuronal dysfunction where factors such as H_2O_2 can mediate apoptosis in nerve cells [63]. In humans, levels of peroxidase-induced reduction of the myelin sheath has been correlated with antioxidant levels in multiple sclerosis[64]. Downstream target genes associated with Parkinson's disease have been shown to be upregulated under H_2O_2 exposures in zebrafish embryos [65]. At the individual level, exposures to metals and metal nanoparticles in zebrafish larvae caused the depletion of neuronal cells in the lateral

line[66,67] and olfactory bulb [68], a response elevated by the up-regulation of antioxidant defence processes[68]·[43]. Damage to lateral line cells has been directly related to a decrease in the startle response[69]. In addition, the knockdown of the antioxidant gene *sqstm1*, has been shown to cause a decrease in locomotive activity in zebrafish larvae, again supporting the role of sensory systems as key target for oxidative-stress responses to metal exposures[70].

Reductions in schooling and aggressive behavior as well as predator-avoidance across fresh-water fish species, including rainbow trout, fathead minnow (*Pimephales promelas*), guppy (*Poecilia reticulate*) and Japanese medaka have been identified following exposure to metals [71].

1.5.3 Fecundity

ROS are significant in reproductive success for example NO$^-$ being necessary in spermatogenesis [59]. The allocation of resources to dimorphic features has been identified as resulting in ROS[59] and male stickleback with diets rich in antioxidants had a greater reproductive success compared to males with poorer diets[72]. Maternal resource allocation has been identified in brown trout where increased deposits of carotenoids in eggs was correlated with higher survival rates under microbial burden in developing offspring[73]. The maternal deposits of antioxidants are also supported in zebrafish, where glutathione-s-transgerase- p (GSTp) has been identified as being maternally deposited [61].
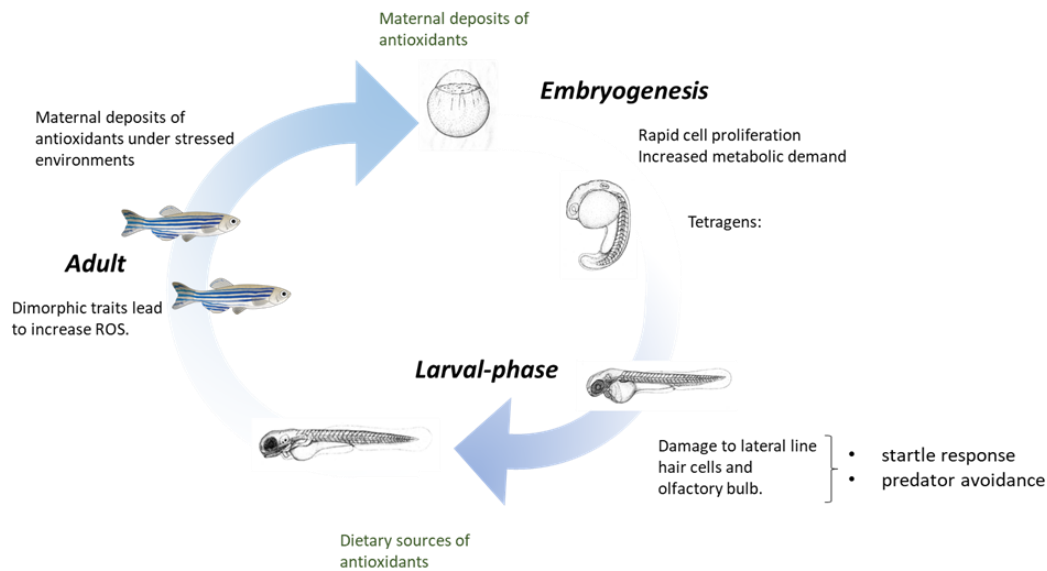
Figure 1.4: Life stage specific effects of oxidative stress with whole-organism and population level consequences. (*images obtained from Kimmel et al. 1995*). Developmental stages are under oxidative stress due to rapid cell proliferation events but have maternal deposits of antioxidants. Antioxidants can be derived from the diet at larval and adult stages. Larval stages have been identified as being susceptible to nerve damage in the lateral line. Male dimorphic traits and aggressive behavior leads to ROS production. Maternal deposits of antioxidants in embryos.

1.6 Molecular initiating events in stress responses pathways

Oxidative stress responses can have wide-ranging outcomes at both individual and population levels. Such outcomes are controlled by a series of stress-response pathways activated in MIE processes. This thesis focuses on some of the main targets of prooxidant chemical and environmental stressors of concern comprising of TFs belonging to the adaptive stress response pathways. These are activated in order to prevent or mitigate against cellular damage and maintain homeostasis [13]. Adaptive stress response factors are held in the nucleus by fast-acting response proteins with the exception of the Aryl hydrocarbon receptor (AhR) which is included due to its association with nuclear factor (erythroid-derived 2)-like 2 (Nrf2) as well as being targeted by

dioxins. In most cases, factors are named after the general responses they are associated with and comprise of Nrf2, nuclear factor kappa-light-chain-enhancer of activated B cells (NFkB), hypoxia inducible factor 1 (HIF-1α), metal-transcription factor 1 (MTF1), heat shock factor 1 (HSF1), tumor protein p53 (p53) and the AhR.

1.6.1 Overview of gene regulatory processes

The initiation of gene transcription requires a coordinated response between multiple regulatory proteins and the underlying DNA-sequence to recruit RNA polymerase II to the transcription start site (TSS) [74]. At the fundamental level, general transcription factors (TFs) bind to a core promoter such as the TATA box and assemble with ribonuclear-polymerase II (RNAPII) to form a transcriptional complex [74]. General TFs are required for all transcriptional processes and produce a basal level of transcription[74].

Inducible gene expression requires the association of TFs with unique sequences called transcription factor binding site (TFBS), traditionally viewed as being proximal to the gene coding region in promoters. These TFs can inhibit or enhance the levels of transcription and are themselves inhibited by receptors or binding proteins within the cell and released by specific inducers[13]. With some exceptions, the majority of stress-responsive factors fall into the latter case, which is considered to be a fast-acting response to stimuli compared to receptor-mediated inhibition[13].

On entering the nucleus, the accumulation of transcription factors increases the likelihood of binding to TFBS within the DNA sequence [75]. The number of TFs binding to DNA sequences influences the level of inducible gene transcription, which is correlated with the number of factors occupying TFBS [75]. In addition, functionally redundant TFBS can bind TFs, reducing the overall concentration of free TFs, suggested to act as a negative feedback mechanism [76]. The strength of association between TF and DNA can also influence the transcriptional rate and is dictated by the underlying DNA sequence[75]. TFs with a weak binding strength cause a lower transcriptional rate compared to those that are tightly associated with the DNA sequence[75].

In addition to regulatory regions in promoters, enhancers comprise of multiple regulatory elements and are considered to further increase the rate of transcription[77]. Enhancers can be located distal or proximal to the TSS as well as upstream, downstream or within the introns of target genes[77]. Shadow enhancers, duplicates of enhancer sequences, act even more remotely to the target gene and complement enhancer regions[78]. Although promoter and enhancers are traditionally viewed as being functionally distinct, the sequences have been shown to be interchangeable[77]. For the purpose of this thesis, the term regulatory regions will be used to define promoter and enhancer sequences.

As the DNA double helix is three-dimensional, features that influence the structure are also essential for regulatory function. DNA is held in the nucleus through associations with histone proteins within nucleosomes, where 8 histone proteins make one nucleosome[79]. TFs are not able to access DNA sequences that are tightly bound to nucleosomes whereas DNA that is not tightly associated can be bound by TFs[79]. The strength of DNA-nucleosome associations is dictated by the status of histone tails where methylation causes deactivation and acetylation, activation[79]. Sites that contain methylated regions are associated with CpG islands, areas that have a high GC% content in DNA sequences[80]. The influence of structural features on gene transcription is the basis of epigenetics where phenotypic outcomes cannot be explained by the underlying DNA-sequence alone[81].

Both the DNA-sequence composition and structural features therefore influence the regulatory potential of gene transcription. However, due to experimental costs, the influence of both TF-binding and changes in chromatin composition under specific conditions is rarely known. This thesis bases its research focus on the concepts surrounding DNA sequence and protein binding rather than chromatin composition as an indicator of regulatory function, though both factors are essential for interpreting gene-regulatory processes from experimental data. In this case, DNA sequence analysis gives a broad overview of regulatory interactions given that the sequence is identical within every tissue of the organism.

1.6.2 Stress-responsive transcription factors share a common regulatory architecture.

Due to the need to respond to rapid changes in internal cellular conditions, stress-response factors share a common regulatory architecture that at a basic level comprises of a sensor molecule, transcription factor and co-activator. Sensor molecules are able to prevent TFs entering the nuclease under basal conditions, such as inhibitory proteins that target TFs for ubiquitin-proteasome pathway, but the interaction between the sensor-TF is rapidly disrupted under the presence of an inducer (Figure 1.5) [13]. The TF is then able to associate with co-activators, a process necessary to either to enter the nucleus or to bind to TFBS in regulatory regions themselves [13].
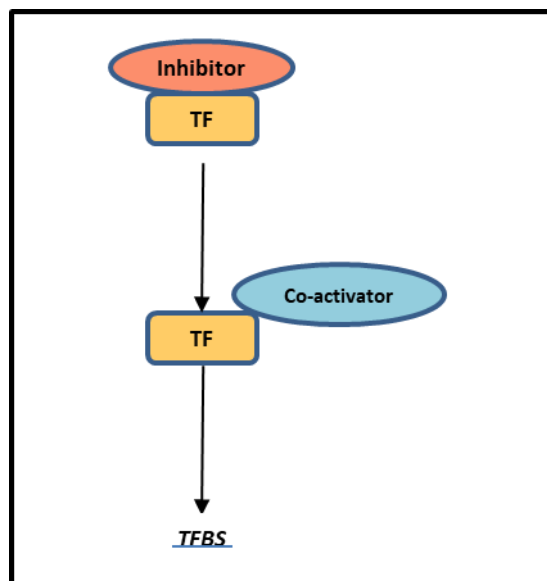
Figure 1.5: Schematic of the common architecture shared between stress-response TFs and TFBS. TFs are prevented from entering the nucleus by sensor molecules, inhibitory proteins that prevent TF activation by initiating processes such as the ubiquitin-proteasome pathway. Inducers disrupt the interaction between the TF and sensor molecule. TFs are then release and bind to co-activators that transport them into the nucleus to bind to TFBS.

1.6.3 Nuclear factor (erythroid-derived 2)-like 2 (Nrf2)

In 1991, *Rushmore et al*. [82] defined a conserved sequence which regulated the expression of the rat GSTya and NQO1 antioxidants. This consensus sequence of 5' TGACnnnGC 3' [82] was found to be evolutionarily conserved[83] and necessary for the expression of phase II detoxification enzymes [84] and over 200 antioxidant responsive genes including GST, GPx, heme oxygenase 1 (HMOX1) and NQO1 [85]. This conserved region was defined as the antioxidant response element (ARE) or electrophile response element (EpRE) and shown to be activated through binding of the transcription factor, Nrf2 [86] (Figure 5).

Nrf2 is a member of the Cap 'n' Collar (CNC) basic leucine zipper protein (bZip) family and under basal conditions is held by its sensor molecule, the inhibitory protein Kelch-like ECH associated protein 1 (Keap1), which favors degradation of Nrf2 by ubiquiation [87]. In response to oxidative stress, H_2O_2 and electrophiles, phosphorylation of cysteine molecules in Keap1 changes its conformational shape causing the release of Nrf2 in a hinge and latch mechanism [64,87] . The transcription factor then enters the nucleus where it accumulates and associates with its coactivators (small Maf proteins, an ARE-binding protein and a C-binding protein) before being able to bind to EpRE sequences [88]. Nrf2 and its downstream targets are evolutionarily conserved in fish species[83]. Paralogs of Nrf2 have been identified in zebrafish with *nrf2a* and *nrf2b* having divergent functions in antioxidant regulation and …. Respectively [89].

1.6.4 Nuclear factor kappa-light-chain-enhancer of activated B cells (NFkB):

Oxidative stress is just one factor acting on NFkB which is involved in regulation of immunity, inflammation and the cell cycle[90]. NFkB is a dimer formed from NFKB subunit 1/2 and proteins of the Rel family which have a Rel homology domain and include Rela, Relb and c-Rel [90]. Formation of the dimer allows binding to kB elements of target genes at the consensus sequence 5' GGGRNWYYCC 3' identified in mammals[90]. Activation of

NFkB can occur in a canonical or non-canonical response pathway (Figure 5). In the canonical response, which is fast acting, the IkB kinase (IKK) complex is activated acting as a sensor molecule and targets IkBa for phosphorylation, allowing the concentrations of NFkB dimers to increase and bind to kB elements on the DNA. In the non-canonical response, the NFkB-inducing kinase (NIK) phosphorylates p100 to p52, forming a dimer with RelB and binding to kB elements. NF-kB then triggers a ROS-induced inflammation response. Cell specific expression of NFkB has been recorded under H_2O_2, with responses dependent on cell type [91,92]. NFkB is conserved in zebrafish and shown to be essential for notochord development[93].

1.6.5 Heat shock factor 1 (HSF1)

Under oxidative stress, heat shock proteins (Hsp) are activated by heat shock transcription factor-1 (HSF1) following stimulation from the oxidative stress detectors, Hsp33 and HMOX1[94]. Hsp33/HMOX1 act as sensor molecules, containing cysteine residues which are oxidised, releasing zinc and causing the formation of disulphide bonds which activate the protein [94] (Figure 1.5). During oxidative stress, an increased expression of HSP occurs and is termed the heat shock response (HSR) [94]. The heat shock element (HSE) is composed of inverted repeats of 5' nGAAn 3' and the arrangement of 5'nGAAn 3' influences the binding of HSF[95]. HSF1 is also activated by oxidised proteins or DNA by activating Hsp70/90 [94]. Once transcribed, Hsp can prevent the accumulation of damaged proteins. HSF1 is conserved in zebrafish and necessary for eye development [96].

1.6.6 Metal transcription factor 1 (MTF1)

The evolutionarily conserved metal response element binding transcription factor-1 (MTF1) activates gene expression though the metal responsive elements (MRE) which share a core consensus sequence of 5' TGCRCNC 3' [86] . MREs control gene expression of metallothioneins, heavy metal binding proteins which reduce both the damaging effects of metals and are scavengers of hydroxyl radicals [86] (Figure 1.5). MTF1 is activated by heavy

metals such as zinc [97], which can be released from MT through the oxidation of cysteine[98].

MTF1 is also involved in the transcription of antioxidants such as glutamate cysteine ligase (GCL) [99]. MTs can act as zinc pools, releasing zinc under oxidative stress which leads to the binding of MTF1 to the MRE[99]. MTF1 is conserved in zebrafish where it is regulated through the release of zinc and binds to MRE sequences in MT gene promoters[100].

1.6.7 Hypoxia inducible factor 1 (HIF-1α)

All aerobic organisms require the ability to sense when internal and external oxygen supply is reduced. Hypoxia inducible factors (HIFs) are evolutionary conserved TFs across metazoans, activating the transcription to downstream targets in low oxygen conditions[101]. Three subtypes of HIFα exist in mammals, HIF-1α, HIF-2α and HIF-3α, which all respond to hypoxia and act on distinct target genes[101]. The most well studied factor is HIF-1α, bound in the cytoplasm to von hippel Linau protein (VHLp) after undergoing proline-degradation from the sensor molecules proyl-hydroxylase domain-containing proteins (PHDs) under basal conditions[101]. Hypoxia suppresses the action of PHDs allowing free HIF-1α to enter the nucleus, dimerise with the aryl hydrocarbon nuclear translocator (ARNT also known as HIF-1β) and bind to hypoxia-response elements in promoter regions[101] (Figure 1.5).

HIF-1α targets increase the oxygen supply and maintain energy production in affected tissues. Under hypoxia, HIF-1α regulates genes involved in increasing metabolic processes using glutamate for aerobic respiration as well as increasing glucose transport and mitochondrial turnover[102]. In addition, HIF-1α regulates the targets erythroprotien (EPO) and vascular endothelial growth factor (VEGF), which activate erythrocyte production and angiogenesis to increase the number of red blood cells and the oxygen supply to affected tissues respectively[102]. Divergent features of HIF-1α subtypes have been identified in zebrafish but both *HIF-1αa* and *HIF-1αb* were shown to be up

regulated throughout development [103]. In addition, in the atlantic killifish, lactase-dehydrogenase B was shown to be regulated through HREs, suggesting the transcription factor is conserved[103].

1.6.8 Aryl–hydrocarbon receptor (AhR)

The association with ligands including polyaromatic hydrocarbons (PAHs) and 2,3,7,8-tetrachlorodibenzo-*p*-dioxin (TCDD) make the aryl-hydrocarbon receptor (AhR) a crucial regulator of dioxin toxicity[104]. In an inactive state, the AhR forms a protein complex with Hsp90 in the cytoplasm[105]. Ligand binding cause a conformational change within the protein complex leading to the exposure of the nuclear localization signal on the AhR[105]. Within the nucleus, the AhR dissociates from Hsp90 and dimerizes with the ARNT and binds to dioxin response elements/xenobiotic response elements (DREs/XREs) within target genes (Figure 1.5) [105]. The AhR is the principle mediator of the canonical pathway of enzyme detoxification, initiating transcription of the cytochrome P450 family which act to metabolize toxic chemicals through oxidation reactions[106]. Like Nrf2, the AhR upregulates expression of GSTs and NQO1 and is considered the activator of phase 1 metabolism of xenobiotic-induced stressors[107].

Fish species are particularly sensitive to the toxic effects of dioxin-like chemicals which are of environmental concern. As a result, the characterization of AhR protiens has been documented in a range of fish species including rainbow trout[108], killifish[109] and zebrafish[109]. These species have three copies to the mammalian AhR; paralogs *ahr1a* and *ahr1b*, as well as the ortholog *ahr2* [109]. In this case, experimental evidence indicates that *ahr1a* has become functionally redundant over evolutionary time whereas *ahr1b* and *ahr2* have retained the function of the AhR gene in mammals [109]. In addition, the differential expression of *ahr2* target genes in the Arctic Charr (*Salvelinus alpinus*) has been associated with controlling phenotypes consisting of blunt snout and sub-terminal mouth of the benthic subspecies [110].

41

1.6.9 Tumor protein P53 (P53)

Tumor protein P53 (P53) is associated with regulating both cell survival and apoptotic pathways, the ability of which is not yet fully understood. P53 is held in the cytoplasm by the sensor molecule, mouse double minute 2 homolog (MDM2), which targets the TF for the ubiquitin-proteasome pathway. Release of P53 requires the disruption of the MDM2-P53 interaction by phosphorylation of residues on either protein, signaled by events including DNA damage and hypoxia (Figure 1.9) [111]. P53 is then acetylated before binding to the regulatory sites of target genes at consensus 5' RRRCWWGYYY 3' elements [111]. The MDM2-P53 pathway has been identified as being conserved in zebrafish, with the TF existing as one copy and activated in response to DNA damage[112].

1.6.10 Biomarkers of pathway activation.

Genes which are known targets of TFs are widely regarded as biomarkers for the activation of specific pathways and are therefore indicators of MIEs. These are subsequently measured in whole-genome sequencing such as RNA-seq, and gene-targeted experimental methods such as Reverse Transcriptase-quantitative polymerase chain reaction (RT-qPCR). Biomarker genes are associated with specific biological processes. Biomarkers for the adaptive stress response are shown in Table 1.3.

TF	Target Genes	Process
Nrf2	GSTP, HMOX1, NQO1, GSTA, SOD1	Phase 2 detoxification (Antioxidant Defence)
AhR	CYP450s	Phase 1 detoxification (Xenobiotic metabolism)
HSF1	HSP70, HSP90, HSP72	Heat Shock
HIF-1α	VEGF1, EPO, TIMP1	Angiogenesis, Erythropoiesis
MTF1	MT1, MT2	Metallothioneins
NFkB	IL8, IL6, IL4, COX2, BCL-2	Immune Response.
P53	BAX, BIM, APAF1, NIX	Apoptosis, Cell survival

Table 1.3 Downstream biomarkers and processes associated with stress-response factors.
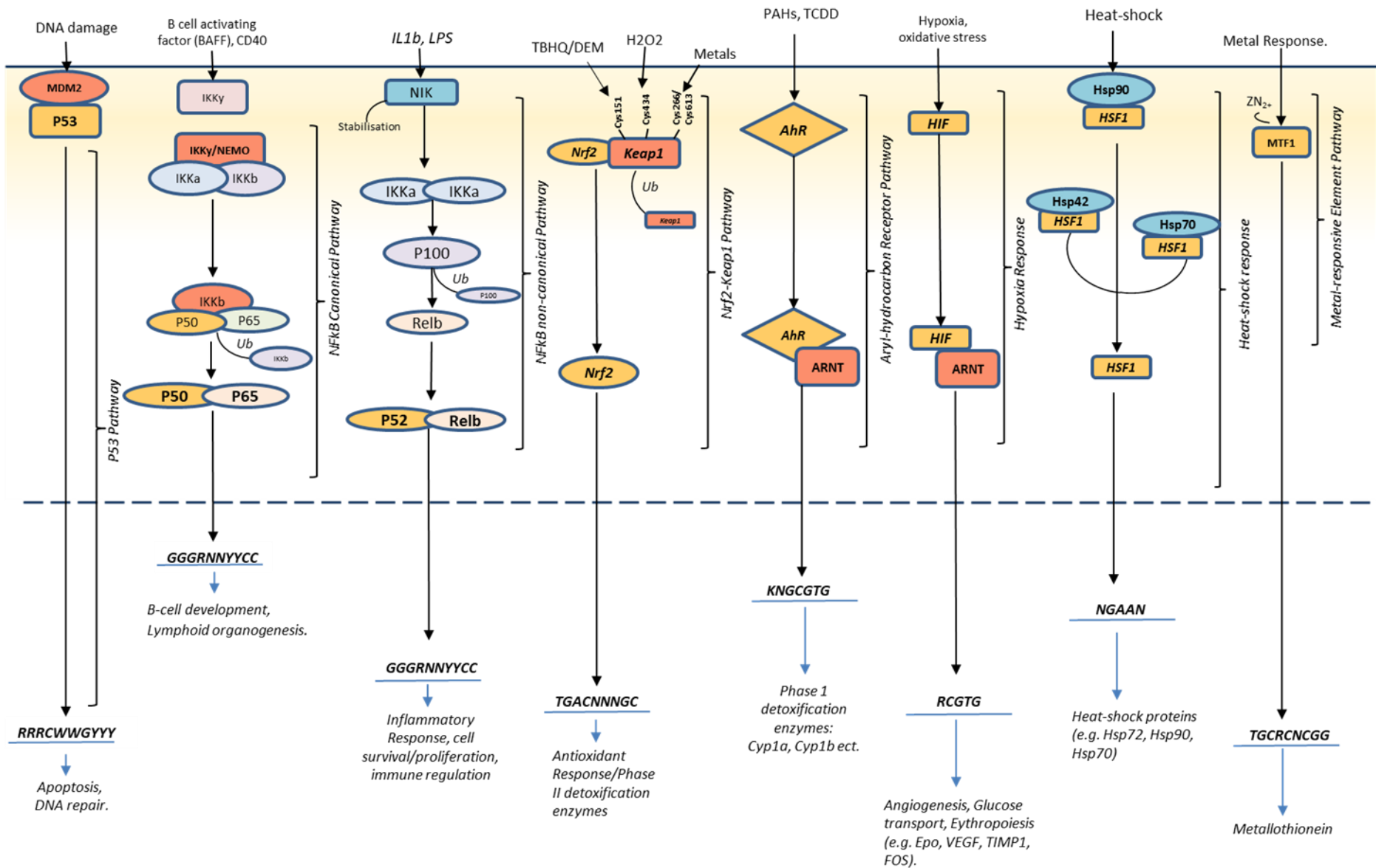
Figure 1.6: Activation pathways of selected stress-responsive transcription factors. Schematic representation of stress-response pathways as focused on in this thesis. TFs, sensor molecules and coactivators are coloured in accordance to the regulatory architecture components as shown in Figure 1.5. Consensus TFBS are indicated with the IUPAC alphabet as shown in Table 1.4. P53 is bound by the sensor molecule mouse double minute 2 homolog (MDM2) and released by signals such as DNA damage. P53 binds to consensus response elements (5'RRRCWWGYYY3'*).* In the canonical pathway, the sensor molecule subunit NFkB kinase subunit beta (IKKb) is activated and phosphorylates IkBa leading to the translocation of the p65/p50 complex into the nucleus where it binds to NFkB response elements. In the non-canonical pathway, NFkB kinase subunit alpha (IKKa) becomes activated through the sensor molecule NFkB-inducing kinase (NIK), leading to the phosphorylation of p100. This allows p52 and RelB to heterodimerise, enter the nucleus and bind to NFkB regions. Nrf2 is released from its sensor molecule Keap1 when electrophilic compounds react with cysteine residues on Keap1. Nrf2 enters the nucleus and binds to electrophile response elements (5'TGACNNNGC3') to initate the transcription of antioxidant genes. The AhR is activated by dioxin-like chemicals and binds to the coactivator ARNT before binding to xenobiotic response elements (5'KNGCGTG3') and initiating the transcription of CYP450s. HIF-1α binds to the coactivator ARNT before binding to hypoxia response elements (5'RCGTG 3') initiating genes involved in angiogenesis and erythropoiesis. HSF1 is activated and released from the sensor molecule Hsp90 and binds to heat-shock elements (5'NGAAN3') causing transcription of heat shock proteins. Metal transcription factor 1 (MTF1) is activated by the release of zinc and binds to metal response elements (5'TGCRCNCGC3').

1.7 Advances in ecotoxicology require the understanding of pathway connectivity

Although stress-response target genes and their TFs have been widely established, it is currently challenging to establish complete AOPs from MIE to phenotypes, especially as multiple phenotypes have been established to be life-stage and exposure specific. The identification of novel and shared target genes is a major factor influencing regulatory networks and multiple methods have been established to identify the downstream targets of specific TFs, largely through the identification of TFBS.

1.7.1 Transcription factor binding site identification

TFBS are short and degenerate, meaning that TFs bind to multiple combinations of sequences and bases in binding site patterns are interchangeable[113]. The identification of functional binding sites requires the use of a combination of methods based on *in vivo, in vitro* and *in silico* techniques. The method(s) selected depend on the research question, broadly falling into those methods which identify TFBS at the whole-genome level (Table 1.2) and those at the gene-specific level (Table 1.3). In combination with the methods discussed, whole genome sequencing methods assessing regulatory features such as the methylation-status of CpGs islands using bisulfite sequencing[108] and chromatin conformation using chromatin conformation capture sequencing (Hi-C)[108] are being widely employed to identify regulatory regions by indicating regulatory status.

1.7.2 Genome-specific methods

Genomic methods aim to establish either the downstream targets or functional binding regions by focusing on associations of individual TFs with the DNA region (Table 1.4). Where the results of inducible expression of a TF is investigated, gene expression analysis such as through RNA-sequencing is used in combination with ChIP-Seq, ATAC-Seq or DNAase1-hypersensitivity to establish changes in regulatory expression in combination with protein-DNA binding events.

1.7.3 Gene-specific identification

Where the identification of regulatory proteins controlling the expression of a specific gene is necessary, identification methods can be used at a smaller scale for example, ChIP-PCR (Table 1.5). Methods such as morpholino (MO) analysis are used in combination with knock-outs (KOs) or knock-downs (KDs) to establish regulatory interactions [114]. However, these methods are unable to

determine direct binding events and affects can be due to gene expression changes resulting from KO/KDs upstream in the regulatory network.

Method	Technique	Limitations
RNA-seq	Adapter molecules are bound to cDNA fragments and sequenced on a microplate. RNA-seq methods are integrated with ChIP-seq, ATAC-seq and DNAse-seq to identify TFs that are active under exposure conditions.	Variations in mapping TSS[115].
ChIP-Seq (Chromatin-immunoprecipitation)	*Used with RNA-Seq for inducer-specific TFs.* Active transcription factors are cross-linked to DNA, which is then sheared, and immunoprecipitated. Antibodies targeting the protein of interest are then used to isolate bound-DNA regions that can then be sequenced[116].	Requires the use of specific antibodies[116]. Functionally redundant binding of TFs.
ATAC-Seq (assay for transposase-accessible chormatin using sequencing)	*Used with RNA-Seq for inducer-specific TFs.* A highly sensitive transposase (Tn5) is used to cut DNA free of nucleosomes and preferentially enriched for proteins bound to DNA(open chromatin)[117]. The resulting regions are isolated and sequenced[117].	Can cause non-specific amplification of non-nuclear DNA.
DNAse-Seq (Dnase- I hypsersensitvitiy sequencing)	*Used with RNA-Seq for inducer-specific TFs.* Transcription factors are able to associate with DNA that is not bound by chromatin. DNAse fragments uncondensed regions and the resulting fragments can be compared to the background genome sequence to identify regulatory regions of DNA[118]. *In silico* regulatory analysis can then be used to determine associated transcription factors (open chromatin).	Large amounts of DNA are needed and can cause cleavage bias leading to the mis-identification of TFBS[119].
FAIRE-Seq (formaldehyde assisted isolation of regulatory element sequencing)	Chromatin is cross-linked onto DNA with formaldehyde (open chromatin). The remaining regions are isolated using phenol-chloroform and sequenced to identify regulatory regions of DNA. TFs are assigned based on DNA-sequences[120].	High levels of background in output data[119].
SELEX (systematic evolution of ligands by exponential enrichment)	Purified TF are incubated with DNA oligos. Degradation steps leave TF ligated to specific DNA oligo sequences[121]. DNA sequences are amplified and the process is repeated for increased specificity[121].	Only high affinity binding sites are amplified[121].

Table 1.4: Genome-specific TFBS identification. Whole-genome sequencing techniques and limitations for identifying TFBS. RNA-Seq is used in combination with genome sequencing assays ChIP-Seq, ATAC-Seq and DNAse-Seq to correlate changes in gene expression under certain inducers with increases in TF-binding.

Method	Technique	Limitations
ChIP-PCR	Same assay as ChIP-Seq but antibodies are incubated with only the DNA sequence of a specific gene. Can be used to validate the efficiency of ChIP-Seq assays[116].	Limitations in available antibodies[116].
Reporter-gene assay	Promoter attached to reporter gene (e.g. luciferase, green florescent protein (GFP). Promoter is truncated to identify functional binding sites. Can be used *in vitro* and *in vivo* to identify functional binding sites[28].	Mosaic-effects in model organisms[28].
Electrophoretic mobility shift assay (EMSA)	Using electrophoresis, DNA sequences associated with proteins cause a band shift in comparison to the free-nucleic acid sequence[122].	Samples are not at chemical equilibrium during electrophoresis and the exact DNA-sequence bound is unknown[122].
**Morpholino analysis*: Gene-specific knock-downs (KD)*	Morpholino oligos (MO) prevent the translation of target proteins by binding to the mRNA sequence of the gene. For regulatory gene analysis, MOs can be used in combination with WISH and qPCR to identify any changes in gene induction of downstream targets *in vitro*[123].	Off-target effects of MOs could alter the gene-response. The efficiency of MOs is reduced over time as the concentration is depleted[123].
Gene-specific knock-outs (KO)	Target TF is knocked-out through generation of mutant lines formed using techniques such as CRISPR-Cas9[124].	Successful development of KO lines can require multiple generations depending on technique used.
Whole mount in-situ hybirdisation: (WISH)	*Used in combination with MO/KOs in vivo.* RNA probes generated using the mRNA sequence of target gene tagged to BM-purple dye[125].	Does not identify direct binding events.
RT-qPCR	*Used in combination with MO/knock-outs in vivo.* RNA extracted and amplified using gene-specific primers.	Does not identify direct binding events.

Table 1.5: Gene-specific TFBS identification. Gene-specific methods for identifying binding sites used *in vivo* and *in vitro*. KO/KD methods are used in combination with RT-qPCR or WISH to correlate changes in gene expression with the presence/absence of a selected factor as an indicator of its role in gene-expression. However, this does not show direct binding events.

1.7.4 In silico identification of TFBS

In silico methods can be used to identify TFBS in combination with whole-genome and gene-specific methods. TFBS generated from *in vitro* or *in vivo* data are aligned to create overall consensus binding sequence. These binding

sequences are defined by the International Union of Pure and Applied Chemistry (IUPAC) alphabet[126] which can be used to represent the occurrence of 2 or more bases in a single position within a sequence (Table 1.6). Consensus motifs represent the binding sequence pattern but have no additional information such as the probability of a particular sequence occurring within the binding site.

More complex motifs represent DNA binding sequences in the form of positional weight matrices (PWMs) [127]. These matrices incorporate the probability that a specific base occurs in each position in the pattern following from the multiple sequence alignments of validated binding sites[127]. The original alignment file produced is known as a positional factor matrix (PFM) and is formed of the counts of each base within each position from the available alignment data[127]. The counts shown in a PFM are then converted into a position probability matrix (PPM) with counts converted to probabilities; the sum of probabilities in each position is equal to 1. In this case, each probability score across every position is independent of the surrounding sequence. For example, if in position 1, the probability of base A is 0.5 and in position 2, the probability of base T is 0.25, these scores are independent of each other even if in all cases where 2 =T, 1 = A .

Nucleotide code	Base
A	Adenine
C	Cytosine
G	Guanine
T	Thymine
U	Uracil
R	A or G
Y	C or T/U
S	G or C
W	A or T/U
K	G or T/U
M	A or C
B	C or G or T/U
D	A or G or T/U
H	A or C or T/U
V	A or C or G
N	any base

Table 1.6: IUPAC alphabet (*Johnson, 2010*.[126])

TFBS identification is highly sensitive to nucleotide bias and the identification of binding sites has the potential to be an artifact of higher GC% in the predicted sequences against the background readings, particularly if enrichment is used as a filtering measure as discussed below. Models must incorporate the genomic background of the species of interest, taking into account the GC% across the genome[128]. The position frequency matrix (PFM) can then be determined by calculating the log-likelihood of occurrence based on the genomic DNA background sequence[128]. *In silico* searches only give putative predictions of binding sites and produce a large number of false positives due to the high likelihood of degenerate sequences occurring in the genome by chance[128].

Databases containing matrixes are widespread and include JASPAR and Transfac (Table 1.7). The JASPAR database is the most widely cited open-source collection of TFBS, spanning six taxonomic groups from vertebrates to plants and regularly used in motif discovery methods in Table 1.7. It is

unsuprisng that the JASPAR database has increased since the development of sequencing-based technologies as shown in Figure 1.7 where it is widely used to characterise binding sites. Transfac remains the most extensively used database overall.

Motif Database	Description	Form	Refs:
JASPAR	Open source binding site discovery for vertebrates, invertebrates, plants and fungi. DNA binding sequences based on experimental evidence collected from PAZAR.	PFMs	[113]
TRANSFAC	Eukaryote transcriptional regulation. Identifies binding sites scores and PWM score. Uses tissue specific profiles.	PFMs	[129]
Hocomoco	Combines known TFBS models for each factor unless motifs associated with the same factor are significantly different. Based on Human, mouse, rat, fungi data.	PFMs	[130]
Pazar	Database of the annotations used to create PWMs in databases such as JASPAR from SELEX data.	PFMs	[131]

Table 1.7: In silico databases methods for TFBS identification. Open source and subscription databases containing TFBS derived from SELEX and ChIP-seq in the format of positional-frequency matrixes (PFMs).
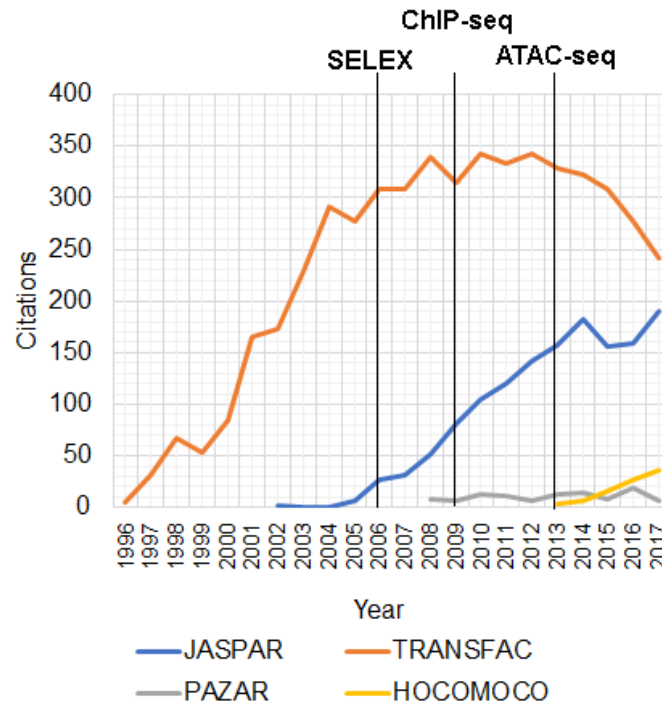
Figure 1.7: Citations by year of databases of TFBS motifs. Number of citations for JASPAR, TRANSFAC, PAZAR and HOCOMOCO databases by year according to web of science. Vertical lines indicate development of whole genome sequencing techniques for binding site identification as indicated.

Multiple software tools have been developed to search PWM within sequence reads based on algorithms including hidden marcov models (HMMs). The output comprises of a probability score and is often represented as the p-value, q-value or z-score to reflect the likelihood that the identified sequence is a true positive in relation to the background sequence composition[132]. The probability score is therefore used as a tool to reduce the number of false positive binding sequences and a stringent threshold measure is recommended to extract the top-scoring probabilities[132]. In the MEME suite, the most widely used search tool (Figure 1.8), a threshold q-value within the range of 0.01-0.05 is regarded as being a good indicator of positive hits.

Motif Discovery	Method	Refs
oPOSSUM	Searches for TFBS clusters with species covering human, mouse, rat and fungi.	[128]
Dcode within the ECR Browser.	Comparative genomics using whole genome alignments and phylogenetic foot printing. Conservation is used as a filter.	[133]
MatInspector	Assigns quality range to searches that match similarity, position and gene id but only searches in the first 1 kb of the selected gene.	[134]
ORCAtk	Phylogenetic footprinting aligning orthologous sequences by pairwise alignment methods. Regions which are significantly similar are scanned for TFBS matrices.	[135]
PhastCons	Multiple alignments generated using UCSC. Alternative plots to identify regions under selective pressure in mammals vs non-mammals using a HMM model. Species cover reptiles, mammal, bird and fish clades.	[136]
The MEME suite	Comprising of MEME (multiple Em for motif excitation) CentriMo (local motif enrichemnt), FIMO (individual motif occurance), SpaMo (spaced motif analysis tool), Mcast (motif cluster alignment), GLAM2 (gapped local alignment of motifs), GoMo (gene ontology for motifs), Tomtom (motif comparison tool). Enriched nucleotide sequences determined by a custom significance threshold. PWM need to be converted into the MEME format.	[137]
CENTIPEDE	Predicts TFBS based on TSS proximity, conservation score, PWM score as well as from experimental observations of ChIP-seq and histone markers. Searches motifs from JASPAR and TRANSFAC.	[138]
RAVEN (Regulatory analysis of variation in Enhancers)	Identifies SNPs from potential TFBS from the Human genome and searches polymorphic sites for overlap with potential TFBS.	[134]
TFBSshape	Predictions made from DNA shape features are derived from core motif sequence files.	[136]
ConTra	All transcripts for a gene of interest are searched for user-selected PWMs across mammalian species. Results are aligned to indicate levels of conservation.	[139]
HOMER (Hypergeometric Optimization of Motif EnRichment)	De-novo motif discovery from sequencing data for motifs 8-12 bps in length. Compares differential enrichment between two sequences. Uses JASPAR database as reference.	[140]

Table 1.8: In silico search programmes for TFBS identification. Open-source search tools used to identify putative TFBS in regulatory regions using additional information such as phylogenetic foot-printing, enrichment and DNA-shape features.

Databases form the basis of a number of searching tools that use various parameters to reduce the occurrence of false positive hits (Table 1.6). In this case, evolutionary conservation[141], levels of enrichment [128] and DNA-shape [136] have been shown to increase the predictive potential of TFBS searches. These parameters have informed on different search-based tools, the use of which depends on the target species and the type of sequencing data available (Table 1.6). Of these methods, MEME and HOMER programs are currently the most widely used TFBS search systems, both of which, as allowing for de-novo motif discovery, have increased in use with the development of ATAC-Seq and ChIP-Seq assays (Figure 1.8).
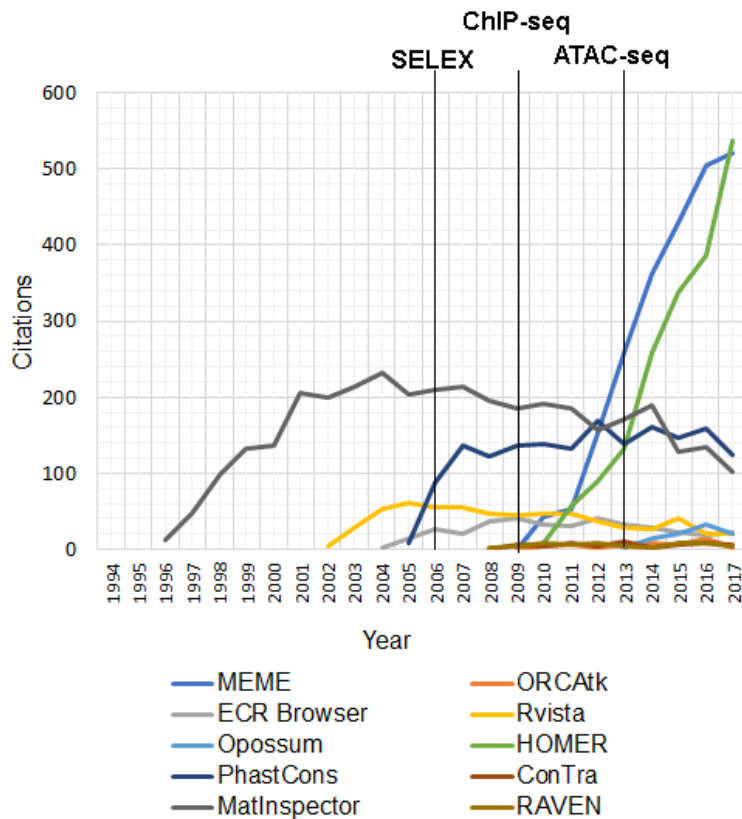
Figure 1.8: Citations by year of TFBS identification programmes. Citations from publications on search tools used to identify TFBS by year according to web of science. Vertical lines indicate development of whole genome sequencing techniques for binding site identification as indicated.

1.7.5 Methods used in combination

In silico methods are often used with genome-wide experimental methods by searching for the specific binding sequences of TFs of interest [137]. For ChIP-Seq, binding peaks are searched for the related TFBS using software programs such as the MEME suite in combination with binding motifs derived from databases such as JASPAR[137]. For methods which identify open chromatin for example, DNAse-1 hypersensitivity assays and ATAC-Seq, returned sequences are compared to databases containing motif files to identify expected TFs [128,137]. In addition, *in silico* methods can be used to identify putative TFBS in RNA-Seq data, where the promoters of genes that have altered expression patterns following exposures are searched using matrix files [142]. This technique has been widely used in ecotoxicology to predict active TFs, especially as the genomic methods and suitable TF antibodies are not

71

always available for the selected species. In this case, there is a reliance on combined *in silico* and gene-specific identification methods. Experimental evidence can inform on more robust PWMs and *in silico* identification methods (Figure 1.8).

Figure 1.9: Flow-diagram of TFBS identification. *In vitro/in vivo* methods inform on PFMs. *In silico*, PFMs identify putative TFBS in novel gene targets. Predicted sites can then be validated *in vitro/in vivo*.

1.8 Transcription factors act in gene regulatory networks.

Identifying downstream targets of TFs allows for the development of gene regulatory networks (GRNs) which describe the molecular cascades leading to outcome processes. So far, regulatory networks have been used in a developmental context to establish the progression of cell-fate in sea-urchins, zebrafish and drosophila. The regulatory events that underpin biological responses to chemical perturbation at MIEs follow the same principles, where TFs act in a holistic framework to regulate genes involved in causing adverse outcomes.

1.8.1 Network Motifs

Motifs are regulatory circuits that underpin the structure of regulatory networks and include positive feedback, negative feedback, coherent and incoherent feedforward loops (Figure 1.9) [143]. Network motifs annotate the regulatory interactions and can be modeled to indicate the processes activated in a time and in more complex models, threshold specific context (Figure 1.9). Networks that are inferred from published data on interactions (prior knowledge networks) can identify key regulatory motifs which can be experimentally validated.

The timing of each process in the cascade is dependent on network motifs (Figure 1.10), the rate of transcription influenced by the number of accessible binding sequences and binding site composition. In addition, the availability of DNA polymerase, the rates of elongation and mRNA clearance all have roles in the final transcript abundance.
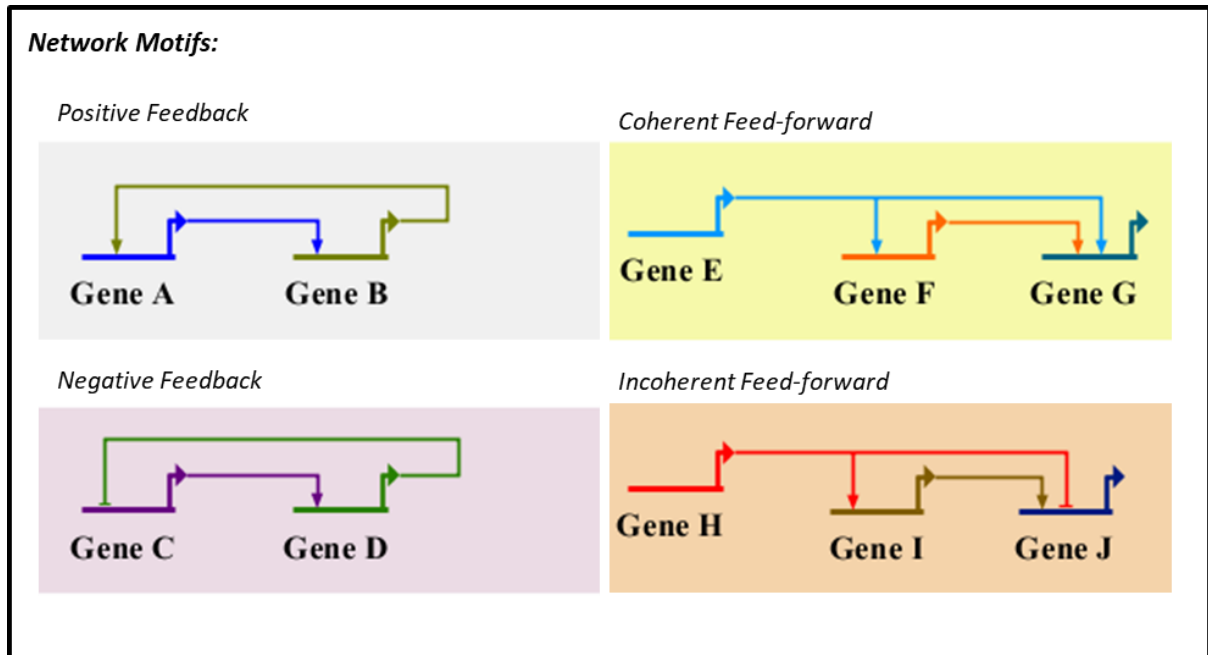
Figure 1.10: Schematic of regulatory motifs created in Biotapestry [144]. A) Positive feedback loop where A activates the transcription of B and vice versa. B) Negative Feedback where C activates D but D inhibits C. C) Coherent Feed-forward where E activates F and G and F also activates G. D) Incoherent Feed-forward where H activates I but inhibits J but I activates J.

1.8.2   Statistical Methods for modeling regulatory networks

The interpretation of regulatory networks depends on the available experimental evidence explaining interactions for example, protein abundance and enzyme kinetics[145]. Often, the level of information is a limiting factor in creating biological models that can quantitatively predict outcome events. As such, statistical methods used to describe networks are selected based on the available information for the system of interest and the number of nodes in the network (Table 1.9).

Method	Explanation
Boolean	Qualitative models which explain gene expression in binary terms with activity described as being active (1) or inactive (0). Boolean models can exist as synchronous, asynchronous or probabilistic and define gene expression occurring over a set period of time[146].
Fuzzy Logic	Incorporate the principles of Boolean logic but regulatory states can fall within a range of 1 or 0. Interaction levels are estimated from biological data[145].
Ordinary differential equations (ODEs)	Produce quantitative outputs by incorporating the rate-dependant steps in regulation including the rate of transcription, mRNA degradation and enzyme kinetics[145].
Pharmacokinetics and Pharmacodynamics (PK-PD)	Describe the actions of drugs on biological processes. PDs is the action of the body on a drug whereas PKs in the action of the drug on the body[147].

Table 1.9 Statistical methods used to explain regulatory processes.
Regulatory networks can be modeled simplistically using Boolean and fuzzy logic algorithms. More complex modeling platforms can be used where there is a greater amount of data available and include ordinary differential equations and pharmacokinetic and pharmacodynamics models.

1.9 Thesis Aims and objectives:

This thesis explores the potential for the use of a GRN approach to predict regulatory cascades in the adaptive stress-response which is activated by a range of pollutants and environmental stressors (Figure 1.7). To achieve this aim, the work in this thesis is separated into 3 main sections: 1.) the development of a mammalian GRN model, 2.) assessing the transferability of the generated mammalian GRN model to fish species through TFBS conservation and, 3.) evaluating the modeling outcomes using the literature and experimental assays. By viewing pathway activation in a holistic framework rather than through independently activated molecular events, the model generated in this thesis identifies the potential biological processes initiated

under a range of chemical exposures and suggests a novel approach to identifying MIEs at the root of AOP and MoA frameworks.

Objective 1 (Chapter 2): *To establish a theoretical model of regulatory connections between stress-response pathways in vertebrates using experimental evidence from mammalian cell lines.*

Chapter 2 proposes a GRN model of the adaptive stress-response using compiled experimental evidence from studies using mammalian cell lines. This network identifies the regulatory links between Nrf2, NFkB, HIF-1α, HSF1, MTF, P53 and AhR pathways through shared downstream targets and protein-protein interactions. The GRN was modeled using Boolean logic (as described in section 1.8) to determine the regulatory cascades and activation of biological processes though the initiation of each TF in the network. Perturbation analysis was conducted in the model to identify the significance of TFs in the progression of response outcomes. Furthermore, interactions with the oestrogen receptor were included as a case-study for determining the influence of receptor-mediated interactions with the adaptive stress-response.

Objective 2 (Chapter 3): *Determine the evolutionary conservation of the mammalian stress-response GRN in fish species through assessing the conservation of TFBS.*

To establish the transferability of the mammalian GRN model (derived in Chapter 2) to fish, the ability for mammalian-derived binding site matrices to predict putative regulatory regions for known downstream target genes across sequenced fish species was investigated. Firstly, consensus-binding sequences derived from mammals were compared to experimentally validated binding sites in fish species. In cases where the consensus sequence did not match validated sites, the efficiency of mammalian PWMs to identify the downstream target genes in fish was conducted using random forest models. Experimentally validated binding sites in fish were aligned and used to generate PWMs. The identification of sites using the fish-specific PWMs in comparison to mammalian

PWMs was conducted for the target genes *hmox1* and *hsp70*, which are known to be targeted by multiple stress-response transcription factors.

Objective 3 (Chapter 4): *To validate the GRN model proposed in Chapter 2 through in-silico and literature derived studies.*

Validation of the GRN model presented in Chapter 2 was undertaken using modified Bradford-Hill considerations (Table 1.1) to establish concordance between gene-expression profiles from pollutant exposures in the literature with the outcomes of the GRN model. Further support was provided by *in vivo* exposures to the Nrf2 inducer, tert-butylhydroquinone (tBHQ) using RT-qPCR in zebrafish developmental stages. The expression of *nrf2*, *mtf1*, *HIF-1α* and *nfkb* was assessed at 2 and 4 dpf following 6 and 12 hr exposures respectively. Acradine orange staining was used as an indicator of apoptosis as an adverse outcome.

An overview of the results addressing each objective, the implications for the field and necessary future research approaches are critically discussed in Chapter 5.

1.10 References:

1. OECD. 2017 .Diffuse Pollution, Degraded Waters. Emerging Policy Solutions. *OECD Policy Highlights* (doi:10.1787/9789264269064-en)

2. Handy RD, Henry TB, Scown TM, Johnston BD, Tyler CR. 2008 .Manufactured nanoparticles: their uptake and effects on fish--a mechanistic analysis. *Ecotoxicology* 17, 396–409. (doi:10.1007/s10646-008-0205-1)

3. Banh S, Wiens L, Sotiri E, Treberg JR. 2016 .Mitochondrial reactive oxygen species production by fish muscle mitochondria: Potential role in acute heat-induced oxidative stress. *Comp. Biochem. Physiol. Part B Biochem. Mol. Biol.* 191, 99–107. (doi:10.1016/J.CBPB.2015.10.001)

4. Gewin V. 2010 .Dead in the Water. *Nature* 466.

5. Krewski D, Acosta D, Andersen M, Bailar III J, Boekelhedie K, Brent R, Charneley G, Cheung V, Green S, Kelsey K, Kerkvliet N, Li A, McCray L, Meyer O, Patterson RD, Pennie W, Scala R, Soloman G, Stephens M, Yager J, Zeise L. 2007 .Toxicity Testing in the 21St Century: a Vision and a Strategy. *Natl. Acad. Sci.* 13, 51–138. (doi:10.1080/10937404.2010.483176.)

6. Wittwehr C, Aladjov H, Ankley G, Byrne HJ, Knecht JD, Heinzle E, Landesmann B, Luitjen M, Mackay C, Maxwell G, Meek MEB, Paini A, Perkins E, Sobanski T, Vileneuve D, Waters KM., Whelan M. 2017 .How Adverse Outcome Pathways Can Aid the Development and Use of Computational Prediction Models for Regulatory Toxicology. 155, 326–336. (doi:10.1093/toxsci/kfw207)

7. Tollefsen KE, Scholz S, Cronin MT., Edwards SW, de Knetch J, Crofton K, Gracia-Reyro N, Hartung T, Worth A, Patlewicz G. 2014 .Applying Adverse Outcome Pathways (AOPs) to support Integrated Approaches to Testing and Assessment (IATA). *Regul. Toxicol. Pharmacol.* 70, 629–640. (doi:10.1016/J.YRTPH.2014.09.009)

8. Allen TEH, Goodman JM, Gutsell S, Russel, PJ. 2014 .Defining Molecular Initiating Events in the Adverse Outcome Pathway Framework for Risk Assessment. *Chem. Res. Toxicol.* 27, 2100–2112. (doi:10.1021/tx500345j)

9. Vrartaar HJM, Lrruwen CJ Van, Hramens JLM. 1992 .Classifying environmental pollutants. 1: structure-activity relationships for prediction

of aqautic toxicity. *Chemosphere* 25, 471–491.

10. OECD., 2012. Appendix 1 - Collection of working definitions. [Online]. [Accessed 7 December 2019]. Avaliable from: http://www.oecd.org/chemicalsaftey/testing/49963576.pdf/

11. Meek B, Palermo CM, Bachman AN, North CM, Lewis RJ. 2014 .Mode of action human relevance (species concordance) framework: Evolution of the Bradford Hill considerations and comparative analysis of weight of evidence. *Applied Toxicology.,* 34, 595-606. (doi:10.1002/jat.2984)

12. OECD., 2018. Users' Handbook supplement to the Guidance Document for developing and assessing Adverse Outcome Pathways., *OECD Series on Adverse Outcome Pathways,* No.1 (doi:10.1787/5jlv1m9d1g32-en)

13. Simmons SO, Fan CY, Ramabhadran R. 2009 .Cellular stress response pathway system as a sentinel ensemble in toxicological screening. *Toxicol. Sci.* 111, 202–225. (doi:10.1093/toxsci/kfp140)

14. Regoli F, Giuliani ME. 2014 .Oxidative pathways of chemical toxicity and oxidative stress biomarkers in marine organisms. *Mar. Environ. Res.* 93, 106–17. (doi:10.1016/j.marenvres.2013.07.006)

15. Halliwell B. 2007 .Biochemistry of oxidative stress. *Biochem. Soc. Trans.* 35, 1147–50. (doi:10.1042/BST0351147)

16. Halliwell B, Whiteman M. 2004 .Measuring reactive species and oxidative damage in vivo and in cell culture: how should you do it and what do the results mean? *Br. J. Pharmacol.* 142, 231–55. (doi:10.1038/sj.bjp.0705776)

17. Cooke MS, Evans MD, Dizdaroglu M, Lunec J. 2003 .Oxidative DNA damage: mechanisms, mutation, and disease. *FASEB J.* 17, 1195–214. (doi:10.1096/fj.02-0752rev)

18. Davies KJA. 2016 .The Oxygen Paradox, oxidative stress, and ageing. *Arch. Biochem. Biophys.* 595, 28–32. (doi:10.1016/j.abb.2015.11.015)

19. McCord JM. 2000 .The evolution of free radicals and oxidative stress. *Am. J. Med.* 108, 652–9.

20. Cho ES, Cha YH, Kim HS, Kim NH, Yook JI. 2018 .The Pentose Phosphate Pathway as a Potential Target for Cancer Therapy. *Biomol. Ther. (Seoul).* 26, 29–38. (doi:10.4062/biomolther.2017.179)

21. Liu J, Litt L, Segal MR, Kelly MJS, Pelton JG, Kim M. 2011 .Metabolomics of oxidative stress in recent studies of endogenous and exogenously

administered intermediate metabolites. *Int. J. Mol. Sci.* 12, 6469–501. (doi:10.3390/ijms12106469)

22. Valko M, Morris H, Cronin MTD. 2005 .Metals, toxicity and oxidative stress. *Curr. Med. Chem.* 12, 1161–208.

23. Volff JN. 2005 .Genome evolution and biodiversity in teleost fish. *Heredity (Edinb).* 94, 280–294. (doi:10.1038/sj.hdy.6800635)

24. Postlethwait J, Amores A, Cresko W, Singer A, Yan Y. 2004 .Subfunction partitioning , the teleost radiation and the annotation of the human genome. *TRENDS Genet.* 20. (doi:10.1016/j.tig.2004.08.001)

25. Glasauer SMK, Neuhauss SCF. 2014 .Whole-genome duplication in teleost fishes and its evolutionary consequences. *Mol. Genet. Genomics* 289, 1045–1060. (doi:10.1007/s00438-014-0889-2)

26. Lu J, Peatman E, Tang H, Lewis J, Liu Z. 2012 .Profiling of gene duplication patterns of sequenced teleost genomes: evidence for rapid lineage-specific genome expansion mediated by recent tandem duplications. *BMC Genomics* 13, 246. (doi:10.1186/1471-2164-13-246)

27. Leung MCK, Procter AC, Goldstone J V, Foox J, Desalle R, Mattingly CJ, Siddall ME, Timme-laragy AR. 2017 .Applying evolutionary genetics to developmental toxicology and risk assessment. *Reprod. Toxicol.* 69, 174–186. (doi:10.1016/j.reprotox.2017.03.003)

28. Lee O, Green JM, Tyler CR. 2015 .Transgenic fish systems and their application in ecotoxicology. *Crit. Rev. Toxicol.* 45, 124–141. (doi:10.3109/10408444.2014.965805)

29. Brown AR, Hosken DJ, Balloux F, Bickley LK, LePage G, Owen SF, Hetheridge MJ, Tyler CR. 2009 .Genetic variation, inbreeding and chemical exposure--combined effects in wildlife and critical considerations for ecotoxicology. *Philos. Trans. R. Soc. B Biol. Sci.* (doi:10.1098/rstb.2009.0126)

30. Ducharme NA, Reif DM, Gustafsson J-A, Bondesson M. 2015 .Comparison of toxicity values across zebrafish early life stages and mammalian studies: Implications for chemical testing. *Reprod. Toxicol.* 55, 3–10. (doi:10.1016/J.REPROTOX.2014.09.005)

31. Howe DG., Bradford YM, Eagle A, Fashena D, Frazer K, Kalita P, Mani P, Martin R, Moxon ST, Paddock H, Pich C, Ramachandran S, Ruzicka L, Schaper K, Shao X, Singer A, Toro S, Westerfield M. 2017 .The Zebrafish

Model Organism Database: new support for human disease models, mutation details, gene expression phenotypes and searching. *Nucleic Acids Res.* 45. 758 - 758(doi:10.1093/nar/gkw1116)

32. Fernández I, Gavaia PJ, Laizé V, Cancela ML. 2018 .Fish as a model to assess chemical toxicity in bone. *Aquat. Toxicol.* 194, 208–226. (doi:10.1016/J.AQUATOX.2017.11.015)

33. Planchart A, Mattingly CJ, Allen D, Ceger P, Casey W, Hinton D, Kullmen SW, Tal T, Bondesson M, Burgess SM, Sullivan C, Kim C, Behl M, Padilla S, Reig DM, Tanguay RL, Hamm J. 2016 .Advancing toxicology research using in vivo high throughput toxicology with small fish models. *ALTEX* 33, 435–452. (doi:10.14573/altex.1601281)

34. Lushchak VI. 2011 .Environmentally induced oxidative stress in aquatic animals. *Aquat. Toxicol.* 101, 13–30. (doi:10.1016/j.aquatox.2010.10.006)

35. Slauch JM. 2011 .How does the oxidative burst of macrophages kill bacteria? Still an open question. *Mol. Microbiol.* 80, 580–3. (doi:10.1111/j.1365-2958.2011.07612.x)

36. van Ham TJ, Mapes J, Kokel D, Peterson RT. 2010 .Live imaging of apoptotic cells in zebrafish. *FASEB J.* 24, 4336–42. (doi:10.1096/fj.10-161018)

37. Yahia D, Haruka I, Kagashi Y, Tsuda S. 2016 .8-Hydroxy-2′-deoxyguanosine as a biomarker of oxidative DNA damage induced by perfluorinated compounds in TK6 cells. *Environ. Toxicol.* 31, 192–200. (doi:10.1002/tox.22034)

38. Oparka M, Walczak J, Malinska D, van Oppen LMPE, Szczepanowska J, Koopman WJH, Wieckowski MR. 2016 .Quantifying ROS levels using CM-H2DCFDA and HyPer. *Methods* 109, 3–11. (doi:10.1016/J.YMETH.2016.06.008)

39. Sevcikova M, Modra H, Slaninova A, Svobodova Z. 2011 .Metals as a cause of oxidative stress in fish : a review., *Veterinarni Medicina,* 56 (11), 537–546.

40. Fabrega J, Luoma SN, Tyler CR, Galloway TS, Lead JR. 2011 .Silver nanoparticles: Behaviour and effects in the aquatic environment. *Environ. Int.* 37, 517–531. (doi:10.1016/j.envint.2010.10.012)

41. Wood C. Farrel AP, Brauner CJ, 2002 .*Homeostasis and toxicity of non-essential metals*; *Fish Physiology Volume 31B, Academic Press,*

Elsievier, USA.

42. Depledge MH, Pleasants LJ, Lawton JH, 2010 .Nanomaterials and the Environment : The Views of the Royal Commission on Environmental Pollution (UK). *Environ. Toxicol. Chem.* 29, 1–4. (doi:10.1002/etc.35)

43. Osborne O, Mukaigasa K, Nakajima H, Stolpe B, Romer I, Phillips U, Lynch I, Mourabit S, Hirose S, Lead JR, Kobyashi M, Kodoh T, Tyler CR. 2016. Sensory systems and ionocytes are among the potential target tissues for silver nanoparticles in fish. *Nanotoxicity,* 10, 9, 1276-1286, (doi:10.1080/17435390.2016.1206147)

44. Aerle R Van, Lange A, Moorhouse A, Paszkiewicz K, Ball K, Johnston BD, Booth T, Tyler CR, Santos EM. 2013 .Molecular Mechanisms of Toxicity of Silver Nanoparticles in Zebrafish Embryos. *Environ. Sci. Technol.* 16, 47, 8005-8014. (doi:10.1021/es401758d)

45. Zhao X, Wang S, Wu Y, You H, Lv L. 2013 .Acute ZnO nanoparticles exposure induces developmental toxicity, oxidative stress and DNA damage in embryo-larval zebrafish. *Aquat. Toxicol.* 136–137, 49–59. (doi:10.1016/j.aquatox.2013.03.019)

46. Groeger AL, Freeman BA. 2010 .Signaling actions of electrophiles: anti-inflammatory therapeutic candidates. *Mol. Interv.* 10, 39–50. (doi:10.1124/mi.10.1.7)

47. Zhang J, Wang C, Ji L, Liu W. 2016 .Modeling of Toxicity-Relevant Electrophilic Reactivity for Guanine with Epoxides: Estimating the Hard and Soft Acids and Bases (HSAB) Parameter as a Predictor. *Chem. Res. Toxicol.* 29, 841–850. (doi:10.1021/acs.chemrestox.6b00018)

48. Trush M a, Penning TM, Dryhurst G, Monks TJ. 2000 .Role of Quinones in Toxicology, *Chem. Res. Toxicol.* 13, 135–160. (doi:10.1021/tx9902082)

49. Banerjee S, Ghosh J. 2016 .Drug Metabolism and Oxidative Stress: Cellular Mechanism and New Therapeutic Insights. *Biochem. Anal. Biochem.* 3, 1–11. (doi:10.4172/2161-1009.1000255)

50. Niethammer P, Grabher C, Look a T, Mitchison TJ. 2009 .A tissue-scale gradient of hydrogen peroxide mediates rapid wound detection in zebrafish. *Nature* 459, 996–999. (doi:10.1038/nature08119)

51. Ansell PJ, Espinosa-Nicholas C, Curran EM, Judy BM, Philips BJ, Hannink M, Lubahn DB. 2004 .In Vitro and in Vivo Regulation of Antioxidant Response Element-Dependent Gene Expression by

Estrogens. *Endocrinology* 145, 311–317. (doi:10.1210/en.2003-0817)

52. Verbruggen B, Gunnarsson L, Kristiansson E, Osterlund T, Owen SF, Snape JR, Tyler CR. 2018 .ECOdrug: a database connecting drugs and conservation of their targets across species. *Nucleic Acids Res.* 46., 930-936 (doi:10.1093/nar/gkx1024)

53. Tyler CR, Jobling S. 2012 .Roach , Sex , and Gender-Bending Chemicals : The Feminization of Wild Fish in English Rivers. *Am. Inst. Biol. Sci.* 58, 1051–1059.

54. Wu SM, Liu J-H, Shu L-H, Chen CH. 2015 .Anti-oxidative responses of zebrafish (Danio rerio) gill, liver and brain tissues upon acute cold shock. *Comp. Biochem. Physiol. Part A Mol. Integr. Physiol.* 187, 202–213. (doi:10.1016/j.cbpa.2015.05.016)

55. Fitzgerald JA, Jameson HM, Dewar Fowler VH, Bond GL, Bickley LK, Uren Webster TM, Bury NR, Wilson RJ, Santos EM. 2016 .Hypoxia Suppressed Copper Toxicity during Early Development in Zebrafish Embryos in a Process Mediated by the Activation of the HIF Signaling Pathway. *Environ. Sci. Technol.* 50, 4502–4512. (doi:10.1021/acs.est.6b01472)

56. Tiedke J, Thiel R, Burmester T. 2014 .Molecular response of estuarine fish to hypoxia: a comparative study with ruffe and flounder from field and laboratory. *PLoS One*, 9, 3, (doi:10.1371/journal.pone.0090778)

57. Chen N, Wu M, Tang G-P, Wang H-J, Huang C-X, Wu X, He Y, Zhang B, Haung C-H, Liu H, Wang W-M, Wang H-L. 2017 .Effects of Acute Hypoxia and Reoxygenation on Physiological and Immune Responses and Redox Balance of Wuchang Bream (Megalobrama amblycephala Yih, 1955). *Front. Physiol.* 8, 375. (doi:10.3389/fphys.2017.00375)

58. Zhu QL, Guo SN, Yuan SS, Lv ZM, Zheng JL, Xia H. 2017. Heat indicators of oxidative stress, inflammation and metal transport show dependence of cadmium pollution history in the liver of female zebrafish. *Aquat. Toxicol.* 191, 1–9. (doi:10.1016/j.aquatox.2017.07.010)

59. Metcalfe NB, Alonso-Alvarez C. 2010 .Oxidative stress as a life-history constraint: the role of reactive oxygen species in shaping phenotypes from conception to death. *Funct. Ecol.* 24, 984–996. (doi:10.1111/j.1365-2435.2010.01750.x)

60. Sant KE, Hansen JM, Williams LM, Tran NL, Goldstone J V., Stegeman

JJ, Hahn ME, Timme-Laragy A. 2017 .The role of Nrf1 and Nrf2 in the regulation of glutathione and redox dynamics in the developing zebrafish embryo. *Redox Biol.* 13, 207–218. (doi:10.1016/j.redox.2017.05.023)

61. Timme-Laragy AR, Goldstone J V, Imhoff BR, Stegeman JJ, Hahn ME, Hansen JM. 2013 .Glutathione redox dynamics and expression of glutathione-related genes in the developing embryo. *Free Radic. Biol. Med.* 65, 89–101. (doi:10.1016/j.freeradbiomed.2013.06.011)

62. Wells PG, McCallum GP, Chen CS, Henderson JT, Lee CJJ, Perstin J, Preston TJ, Wiley MJ, Wong AW. 2009 .Oxidative Stress in Developmental Origins of Disease: Teratogenesis, Neurodevelopmental Deficits, and Cancer. *Toxicol. Sci.* 108, 4–18. (doi:10.1093/toxsci/kfn263)

63. Li J, Johnson D, Calkins M, Wright L, Svendsen C, Johnson J. 2005 .Stabilization of Nrf2 by tBHQ confers protection against oxidative stress-induced cell death in human neural stem cells. *Toxicol. Sci.* 83, 313–28. (doi:10.1093/toxsci/kfi027)

64. Sykiotis G, Bohmann D. 2010, Stress-Activated Cap'n'collar Transcription Factors in Aging and Human Disease. *Sci Signal* 3, 1–45. (doi:10.1126/scisignal.3112re3.Stress-Activated)

65. Priyadarshini M, Orosco L a, Panula PJ. 2013 .Oxidative stress and regulation of Pink1 in zebrafish (Danio rerio). *PLoS One* 8, e81851. (doi:10.1371/journal.pone.0081851)

66. Hernández PP, Moreno V, Olivari F a., Allende ML. 2006. Sub-lethal concentrations of waterborne copper are toxic to lateral line neuromasts in zebrafish (Danio rerio). *Hear. Res.* 213, 1–10. (doi:10.1016/j.heares.2005.10.015)

67. McNeil PL, Boyle D, Henry TB, Handy RD, Sloman K. 2014 .Effects of metal nanoparticles on the lateral line system and behaviour in early life stages of zebrafish (Danio rerio). *Aquat. Toxicol.* 152, 318–23. (doi:10.1016/j.aquatox.2014.04.022)

68. Wang L, Gallagher EP. 2013 Role of Nrf2 antioxidant defense in mitigating cadmium-induced oxidative stress in the olfactory system of zebrafish. *Toxicol. Appl. Pharmacol.* 266, 177–186. (doi:10.1016/j.taap.2012.11.010)

69. Buck, Lauren M. J., Winter, Matthew J., Redfern, William S., Whitfield TT. 2012 .Ototoxin-induced cellular damage in neuromasts disrupts lateral

line function in larval zebrafish. *Hearing Research.,* 284 (1-2), 67-81. (doi: 10.1016/j.heares.2011.12.001.)

70. Lattante S, de Calbiac H, Le Ber I, Brice A, Ciura S, Kabashi E. 2015 .Sqstm1 knock-down causes a locomotor phenotype ameliorated by rapamycin in a zebrafish model of ALS/FTLD. *Hum. Mol. Genet.* 24, 1682–1690. (doi:10.1093/hmg/ddu580)

71. Scott GR, Sloman KA. 2004 .The effects of environmental pollutants on complex fish behaviour : integrating behavioural and physiological indicators of toxicity. *Aquat. Toxicol.* 68, 369–392. (doi:10.1016/j.aquatox.2004.03.016)

72. Pike TW, Blount JD, Lindström J, Metcalfe NB. 2007 .Availability of non-carotenoid antioxidants affects the expression of a carotenoid-based sexual ornament. *Biol. Lett.* 3, 353–6. (doi:10.1098/rsbl.2007.0072)

73. Wilkins LGE, Marques da Cunha L, Glauser G, Vallat A, Wedekind C. 2017 .Environmental stress linked to consumption of maternally derived carotenoids in brown trout embryos (Salmo trutta). *Ecol. Evol.* 7, 5082–5093. (doi:10.1002/ece3.3076)

74. Juven-Gershon T, Kadonaga JT. 2010 .Regulation of gene expression via the core promoter and the basal transcriptional machinery. *Dev. Biol.* 339, 225–229. (doi:10.1016/j.ydbio.2009.08.009)

75. Spivakov M. 2014. Overviews Spurious transcription factor binding: Non-functional or genetically redundant? *Bioessays,* 36, 798-806 (doi:10.1002/bies.201400036)

76. Wu W-S, Lai F-J. 2015 .Functional redundancy of transcription factors explains why most binding targets of a transcription factor are not affected when the transcription factor is knocked out. *BMC Syst. Biol.* 9, (doi:10.1186/1752-0509-9-S6-S2)

77. Andersson R, Sandelin A, Danko CG. 2015 .A unified architecture of transcriptional regulatory elements. *Trends Genet.* 31, 426–433. (doi:10.1016/j.tig.2015.05.007)

78. Barolo S, 2012 .Shadow enhancers: Frequently asked quations about distriuted cis-regulatory information and enhancer redundancy. *Bioessays* 32, 135–141. (doi:10.1016/j.jacc.2007.01.076.White)

79. Allis CD, Jenuwein T. 2016 .The molecular hallmarks of epigenetic control. *Nat. Rev. Genet.* 17, 487–500. (doi:10.1038/nrg.2016.59)

80. Bernstein BE, Birney E, Dunham I, Green ED, Gunter C, Snyder M. 2012 .An integrated encyclopedia of DNA elements in the human genome. *Nature* 489, 57–74. (doi:10.1038/nature11247)

81. Vandegehuchte MB, Janssen CR. 2014 .Epigenetics in an ecotoxicological context. *Mutat. Res. Toxicol. Environ. Mutagen.* 764–765, 36–45. (doi:10.1016/j.mrgentox.2013.08.008)

82. Rushmore TH, Morton MR, Pickett CB. 1991 .The antioxidant responsive element. Activation by oxidative stress and identification of the DNA consensus sequence required for functional activity. *J. Biol. Chem.* 266, 11632–9.

83. Mukaigasa K, Nguyen LTP, Li L, Nakajima H, Yamamoto M, Kobayashi M. 2012 .Genetic evidence of an evolutionarily conserved role for Nrf2 in the protection against oxidative stress. *Mol. Cell. Biol.* 32, 4455–61. (doi:10.1128/MCB.00481-12)

84. Nguyen T, Sherratt P, Pickett C. 2003 .Regulatory mechanisms controlling gene expression mediated by the antioxidant response element. *Annu. Rev. Pharmacol. Toxicol.* 43, 233–260. (doi:10.1146/annurev.pharmtox.43.100901.140229)

85. Xia T, Kovochich M, Brant J, Hotz M, Sempf J, Oberley T, Sioutas C, Yeh JI, Weisner MR, Nel AE. 2006 .Comparison of the abilities of ambient and manufactured nanoparticles to induce cellular toxicity according to an oxidative stress paradigm. *Nano Lett.* 6, 1794–807. (doi:10.1021/nl061025k)

86. Dalton TP, Shertzer HG, Puga A. 1999 .Regulation of gene expression by reactive oxygen. *Annu. Rev. Pharmacol. Toxicol.* 39, 67–101. (doi:10.1146/annurev.pharmtox.39.1.67)

87. Tebay LE, Robertson H, Durant ST, Vitale SR, Penning TM, Dinkova-Kostova AT, Hayes JD. 2015 .Mechanisms of activation of the transcription factor Nrf2 by redox stressors, nutrient cues, and energy status and the pathways through which it attenuates degenerative disease. *Free Radic. Biol. Med.* 88, 108–146. (doi:10.1016/j.freeradbiomed.2015.06.021)

88. Zhu M, Fahl WE. 2001 .Functional characterization of transcription regulators that interact with the electrophile response element. *Biochem. Biophys. Res. Commun.* 289, 212–219. (doi:10.1006/bbrc.2001.5944)

89. Timme-Laragy AR, Karchner SI, Franks DG, Jenny MJ, Harbeitner RC, Goldstone J V, McArthur AG, Hahn ME. 2012 .Nrf2b, novel zebrafish paralog of oxidant-responsive transcription factor NF-E2-related factor 2 (NRF2). *J. Biol. Chem.* 287, 4609–27. (doi:10.1074/jbc.M111.260125)

90. Siomek A. 2012 .NF-κB signaling pathway and free radical impact. *Acta Biochim. Pol.* 59, 323–31.

91. Li N, Karin M. 1999. Is NF- kB the sensor of oxidative stress? , 1137–1143. *FASEB,* 13, 1138-1143.

92. Brzóska K, Sochanowicz B, Siomek A, Oliński R, Kruszewski M. 2011. Alterations in the expression of genes related to NF-κB signaling in liver and kidney of CuZnSOD-deficient mice. *Mol. Cell. Biochem.* 353, 151–7. (doi:10.1007/s11010-011-0781-x)

93. Correa RG, Tergaonkar V, Ng JK, Dubova I, Izpisua-Belmonte JC, Verma IM. 2004 .Characterization of NF-kappa B/I kappa B proteins in zebra fish and their involvement in notochord development. *Mol. Cell. Biol.* 24, 5257–68. (doi:10.1128/MCB.24.12.5257-5268.2004)

94. Kalmar B, Greensmith L. 2009 .Induction of heat shock proteins for protection against oxidative stress. *Adv. Drug Deliv. Rev.* 61, 310–8. (doi:10.1016/j.addr.2009.02.003)

95. Yamamoto N, Takemori Y, Sakurai M, Sugiyama K, Sakurai H. 2009 .Differential recognition of heat shock elements by members of the heat shock transcription factor family. *FEBS J.* 276, 1962–74. (doi:10.1111/j.1742-4658.2009.06923.x)

96. Evans TG, Belak Z, Ovsenek N, Krone PH. 2007 .Heat shock factor 1 is required for constitutive Hsp70 expression and normal lens development in embryonic zebrafish. *Comp. Biochem. Physiol. A. Mol. Integr. Physiol.* 146, 131–140. (doi:10.1016/j.cbpa.2006.09.023)

97. Andrews GK. 2000 .Regulation of metallothionein gene expression by oxidative stress and metal ions. *Biochem. Pharmacol.* 59, 95–104.

98. Palmiter RD. 1998. The elusive function of metallothioneins. *Proc. Natl. Acad. Sci.,* 95, 8428–8430.

99. Günther V, Lindert U, Schaffner W. 2012 .The taste of heavy metals: Gene regulation by MTF-1. *Biochim. Biophys. Acta - Mol. Cell Res.* 1823, 1416–1425. (doi:10.1016/j.bbamcr.2012.01.005)

100. Hogstrand C, Zheng D, Feeney G, Cunningham P, Kille P. 2008 .Zinc-

controlled gene expression by metal-regulatory transcription factor 1 (MTF1) in a model vertebrate, the zebrafish. *Biochem. Soc. Trans.* 36, 1252–7. (doi:10.1042/BST0361252)

101. Koh MY, Powis G. 2012 .Passing the baton: the HIF switch. *Trends Biochem. Sci.* 37, 364–72. (doi:10.1016/j.tibs.2012.06.004)

102. Sharp FR, Bernaudin M. 2004 .HIF1 and oxygen sensing in the brain. *Nat. Rev. Neurosci.* 5, 437–448. (doi:10.1038/nrn1408)

103. Kajimura S, Aida K, Duan C. 2006 .Understanding Hypoxia-Induced Gene Expression in Early Development : In Vitro and In Vivo Analysis of Hypoxia-Inducible Factor 1-Regulated Zebra Fish Insulin-Like Growth Factor Binding Protein 1 Gene Expression Understanding Hypoxia-Induced Gene Expressio. *Mol. Cell. Biol.* 26, 1142–1155. (doi:10.1128/MCB.26.3.1142)

104. Miao W, Hu L, Scrivens PJ, Batist G. 2005 .Transcriptional regulation of NF-E2 p45-related factor (NRF2) expression by the aryl hydrocarbon receptor-xenobiotic response element signaling pathway: Direct cross-talk between phase I and II drug-metabolizing enzymes. *J. Biol. Chem.* 280, 20340–20348. (doi:10.1074/jbc.M412081200)

105. Pollenz RS. 2002 .The mechanism of AH receptor protein down-regulation (degradation) and its impact on AH receptor-mediated gene regulation. *Chem. Biol. Interact.* 141, 41–61. (doi:10.1016/S0009-2797(02)00065-0)

106. Li W, Harper PA, Tang B-K, Okey AB. 1998 .Regulation of cytochrome P450 enzymes by aryl hydrocarbon receptor in human cells. *Biochem. Pharmacol.* 56, 599–612. (doi:10.1016/S0006-2952(98)00208-1)

107. Lo R, Matthews J. 2012. High-Resolution Genome-wide Mapping of AHR and ARNT Binding Sites by ChIP-Seq. *Toxicological Sciences,* 130(2), 349-361 (doi:10.1093/toxsci/kfs253)

108. Abnet CC, Tanguay RL, Hahn ME, Heideman W, Peterson RE. 1999 .Two forms of aryl hydrocarbon receptor type 2 in rainbow trout (Oncorhynchus mykiss). Evidence for differential expression and enhancer specificity. *J. Biol. Chem.* 274, 15159–66. (doi:10.1074/JBC.274.21.15159)

109. Karchner SI, Franks DG, Hahn ME. 2005 .AHR1B, a new functional aryl hydrocarbon receptor in zebrafish: tandem arrangement of ahr1b and

ahr2 genes. *Biochem. J.* 392, 153–61. (doi:10.1042/BJ20050713)

110. Ahi EP, Steinhäuser SS, Pálsson A, Franzdóttir SR, Snorrason SS, Maier VH, Jónsson ZO. 2015 .Differential expression of the aryl hydrocarbon receptor pathway associates with craniofacial polymorphism in sympatric Arctic charr. *Evodevo* 6, 27. (doi:10.1186/s13227-015-0022-6)

111. Brooks C, Gu W. 2010 .New insights into P53 activation. *Cell Res.* 20, 614–621. (doi:10.1038/cr.2010.53)

112. Storer NY, Zon LI. 2010 .Zebrafish models of p53 Functions. *Cold Spring Harb. Perspect. Biol.* 2, 1–12. (doi:10.1101/cshperspect.a001123)

113. Sandelin A, Alkema W, Engström P, Wasserman WW, Lenhard B. 2004. JASPAR: an open-access database for eukaryotic transcription factor binding profiles. *Nucleic Acids Res.* 32, D91–D94. (doi:10.1093/nar/gkh012)

114. Martik ML, Lyons DC, Mcclay DR. 2016 .Developmental gene regulatory networks in sea urchins and what we can learn from them. *F1000 Fac. Rev.* 5, 1–8. (doi:10.12688/f1000research.7381.1)

115. Ozsolak F, Milos PM. 2011. RNA sequencing: advances, challenges and opportunities. *Nat. Rev. Genet.* 12, 87–98. (doi:10.1038/nrg2934)

116. Wardle FC, Tan H. 2015 .A ChIP on the shoulder? Chromatin immunoprecipitation and validation strategies for ChIP antibodies. *F1000Research* 235, 1–10. (doi:10.12688/f1000research.6719.1)

117. Buenrostro JD, Wu B, Chang HY, Greenleaf WJ. 2015 .ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. *Curr. Protoc. Mol. Biol.* 109, 21.29.1-9. (doi:10.1002/0471142727.mb2129s109)

118. Sullivan AM, Bubb KL, Sandstrom R, Stamatoyannopoulos JA, Queitsch C. 2015 .DNase I hypersensitivity mapping, genomic footprinting, and transcription factor networks in plants. *Curr. Plant Biol.* 3–4, 40–47. (doi:10.1016/J.CPB.2015.10.001)

119. Tsompana M, Buck MJ. 2014 .Chromatin accessibility: a window into the genome. *Epigenetics Chromatin* 7, 33. (doi:10.1186/1756-8935-7-33)

120. Giresi PG, Kim J, McDaniell RM, Iyer VR, Lieb JD. 2007 .FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. *Genome Res.* 17, 877–85. (doi:10.1101/gr.5533506)

121. Chai C, Xie Z, Grotewold E. 2011. SELEX (Systematic Evolution of

Ligands by EXponential Enrichment), as a Powerful Tool for Deciphering the Protein–DNA Interaction Space. In *Methods in molecular biology (Clifton, N.J.)*, pp. 249–258. (doi:10.1007/978-1-61779-154-3_14)

122. Hellman LM, Fried MG. 2007 .Electrophoretic mobility shift assay (EMSA) for detecting protein-nucleic acid interactions. *Nat. Protoc.* 2, 1849–61. (doi:10.1038/nprot.2007.249)

123. Summerton JE. 2007. Morpholino, siRNA, and S-DNA compared: impact of structure and mechanism of action on off-target effects and sequence specificity. *Curr. Top. Med. Chem.* 7, 651–660. (doi:10.2174/156802607780487740)

124. Auer TO, Duroure K, De Cian A, Concordet J-P, Del Bene F. 2014 .Highly efficient CRISPR/Cas9-mediated knock-in in zebrafish by homology-independent DNA repair. *Genome Res.* 24, 142–53. (doi:10.1101/gr.161638.113)

125. Thisse, Christine, Thisse B. 2007 .High resolution in situ hybridization to whole-mount zebrafish embryos. *Nat. Protoc.* 3, 59–69. (doi:20.1038/nport.2007.514)

126. Johnson AD. 2010. An extended IUPAC nomenclature code for polymorphic nucleic acids. *Bioinformatics* 26, 1386–9. (doi:10.1093/bioinformatics/btq098)

127. Mathelier A, Zhao X,Zhang AW, Parcy F, Worsley-Hunt R, Arenillas DJ, Buchman S, Chen CY, Chou A, Ienasescu H, Lim J, Shyr C, Tan G, Zhou G, Lenhard B, Sadelin A, Wasserman WW. 2014 .JASPAR 2014: An extensively expanded and updated open-access database of transcription factor binding profiles. *Nucleic Acids Res.* 42, 142–147. (doi:10.1093/nar/gkt997)

128. Kwon AT, Arenillas DJ, Worsley Hunt R, Wasserman WW. 2012 .oPOSSUM-3: advanced analysis of regulatory motif over-representation across genes or ChIP-Seq datasets. *G3 (Bethesda).* 2, 987–1002. (doi:10.1534/g3.112.003202)

129. Matys V, Fricke E, Geffers R, GÖβling E, Haubrock M, Hehl R, Hornischer K, Karas D, Kel a.E., Kel-Margoulis OV, Kloos DU, Land S, Lewicki-Potapov B, Michael H, Munch R, REuter I, Robert S, Saxel H, Scheer M, Thiele S, Wingender E. 2003 .TRANSFAC: Transcriptional regulation, from patterns to profiles. *Nucleic Acids Res.* 31, 374–378.

(doi:10.1093/nar/gkg108)

130. Kulakovskiy I V., Medvedeva YA, Schaefer U, Kasianov AS, Vorontsov IE, Bajic VB, Makeev VJ. 2013 .HOCOMOCO: a comprehensive collection of human transcription factor binding sites models. *Nucleic Acids Res.* 41, D195–D202. (doi:10.1093/nar/gks1089)

131. Portales-Casamar E, Kirov S, Lim J, Lithwick S, Swanson MI, Ticoll A, Snoddy J, Wasserman WW. 2007 .PAZAR: a framework for collection and dissemination of cis-regulatory sequence annotation. *Genome Biol.* 8, 10, R207.1- R207.12 (doi:10.1186/gb-2007-8-10-r207)

132. Grant CE, Bailey TL, Noble WS. 2011 .FIMO: scanning for occurrences of a given motif. *Bioinforma. Appl. NOTE* 27, 1017–101810. (doi:10.1093/bioinformatics/btr064)

133. Ovcharenko I, Nobrega M a, Loots GG, Stubbs L. 2004 .ECR Browser: a tool for visualizing and accessing data from comparisons of multiple vertebrate genomes. *Nucleic Acids Res.* 32, W280-6. (doi:10.1093/nar/gkh355)

134. Andersen MC, Engström PG, Lithwick S, Arenillas D, Eriksson P, Lenhard B, Wasserman WW, Odeberg J. 2008 .In Silico Detection of Sequence Variations Modifying Transcriptional Regulation. *PLoS Comput. Biol.* 4, e5. (doi:10.1371/journal.pcbi.0040005)

135. Portales-Casamar E, Arenillas D, Lim J, Swanson MI, Jiang S, McCallum A, Kirov S, Wasserman WW. 2009 .The PAZAR database of gene regulatory information coupled to the ORCA toolkit for the study of regulatory sequences. *Nucleic Acids Res.* 37, D54–D60. (doi:10.1093/nar/gkn783)

136. Yang L, Zhou T, Dror I, Mathelier A, Wasserman WW, Gordân R, Rohs R. 2014 .TFBSshape: a motif database for DNA shape features of transcription factor binding sites. *Nucleic Acids Res.* 42, D148–D155. (doi:10.1093/nar/gkt1087)

137. Bailey TL. 2011 .DREME: Motif discovery in transcription factor ChIP-seq data. *Bioinformatics* 27, 1653–1659. (doi:10.1093/bioinformatics/btr261)

138. Pique-Regi R, Degner JF, Pai AA, Gaffney DJ, Gilad Y, Pritchard JK. 2011 .Accurate inference of transcription factor binding from DNA sequence and chromatin accessibility data. *Genome Res.* 21, 447–55. (doi:10.1101/gr.112623.110)

139. Broos S, Hulpiau P, Galle J, Hooghe B, Van Roy F, De Bleser P. 2011 .ConTra v2: a tool to identify transcription factor binding sites across species, update 2011. *Nucleic Acids Res.* 39, W74-8. (doi:10.1093/nar/gkr355)

140. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. 2010 .Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* 38, 576–89. (doi:10.1016/j.molcel.2010.05.004)

141. Taher L, McGaughey DM, Maragh S, Aneas I, Bessling SL, Miller W, Nobrega M a, McCallion AS, Ovcharenko I. 2011 .Genome-wide identification of conserved regulatory function in diverged sequences. *Genome Res.* 21, 1139–49. (doi:10.1101/gr.119016.110)

142. Hahn ME, McArthur AG, Karchnew SI, Franks DG, Jenny MJ, Timme-Laragy AR, Stegeman JJ, Woodin BR, Cipriano MJ, Linney E. 2014.The Transcriptional Response to Oxidative Stress during Vertebrate Development: Effects of tert-Butylhydroquinone and 2,3,7,8-Tetrachlorodibenzo-p-Dioxin. *PLoS One*, 9, 11, e113158 (doi:10.1371/journal.pone.0113158)

143. Alon U. 2007 .Network motifs: Theory and experimental approaches. *Nat. Rev. Genet.* 8, 450–461. (doi:10.1038/nrg2102)

144. Paquette SM, Leinonen K, Longabaugh WJR. 2016 .BioTapestry now provides a web application and improved drawing and layout tools. *F1000Research* 5, 39. (doi:10.12688/f1000research.7620.1)

145. Parmar K, Blyuss KB, Kyrychko YN, Hogan SJ. 2015 .Time-Delayed Models of Gene Regulatory Networks. *Comput. Math. Methods Med.* 2015, 347273. (doi:10.1155/2015/347273)

146. Müssel C, Hopfensitz M, Kestler HA. 2010 .BoolNet—an R package for generation, reconstruction and analysis of Boolean networks. *Bioinformatics* 26, 1378–1380. (doi:10.1093/bioinformatics/btq124)

147. Hosseini I, Gajjala A, Bumbaca Yadav D, Sukumaran S, Ramanujan S, Paxson R, Gadkar K. 2018 .gPKPDSim: a SimBiology®-based GUI application for PKPD modeling in drug development. *J. Pharmacokinet. Pharmacodyn.* 45, 259–275. (doi:10.1007/s10928-017-9562-9)

Chapter 2

A gene regulatory network approach to investigating connectivity in vertebrate adaptive stress-response pathways.

2.1 Abstract

Many pollutants in aquatic environments activate the adaptive stress-response driven by evolutionary conserved transcription factors traditionally associated with regulating distinct sets of target genes. However, research has identified that transcription factors are highly interconnected both in terms of their downstream targets and in their ability to regulate the transcriptional activity of one another through direct and indirect protein interactions. Despite growing evidence of cross-talk between pathways, there has been no thorough analysis of how factors in the adaptive stress-response act within an interconnected network and signaling pathways are still widely seen as operating through independent processes. In this study, a gene regulatory network (GRN) for adaptive stress-response factors which are common responders of toxicological insult was constructed using prior knowledge from previously published experimental evidence on protein-DNA and protein-protein binding events in mammalian cell lines. The network was modeled using boolean logic and analysed following systematic perturbations to transcription factors to assess the level of pathway connectivity and its resulting influence on outcome processes. These results identified that the activation of the aryl-hydrocarbon receptor (AhR), hypoxia inducible factor 1 (HIF-1α), metal transcription factor 1 (MTF1), heat shock factor 1 (HSF1) and nuclear factor erythroid-derived 2-like 2 (Nrf2) resulted in the activation of the same adverse outcome processes. Nrf2 and the AhR were shown to be essential mediators of stress-response processes, the absence of which lead to inflammatory responses caused by NFkB. Using oestrogen receptors as a case study, the model also demonstrated that receptor-mediated pathways were able to initiate adaptive-stress response processes in a regulatory network setting. The GRN generated in this study provides an insight into the stress response as an integrated system and, providing validation, presents a novel method for establishing molecular initiating events and regulatory cascades activated by a range of toxicants. The model presents an approach for establishing molecular initiating events (MIEs) at the basis of adverse outcome pathway frameworks (AOPs), widely used in predictive toxicology.

2.2 Introduction

The freshwater environment is a major sink for pollutant exposures and there is a growing need to generate reliable methods to predict adverse consequences in affected organisms. The adaptive stress response underpins a series of molecular processes that are activated under a broad range of pollutant groups including oestrogenic chemicals such as bisphenol-a (BPA)[1], fluorosurfactants such as perflourinated compounds (PFCs) [2], poly-aromatic hydrocarbons (PAHs) [3] and metals[4]. Endpoints, such as the upregulation of antioxidant defence genes, are widely regarded as being controlled by key transcription factors such as nuclear factor erythroid-like 2 (Nrf2), hypoxia-inducible factor 1 alpha (HIF-1α), heat shock factor 1 (HSF1), metal transcription factor 1 (MTF1), nuclear factor kappa-light-chain-enhancer of activated B cells (NFkB) and tumor protein p53 (P53) which mediate the regulation of distinct sets of downstream target genes (Figure 1.6). At the molecular level, the induction of specific pathways leads to a range of outcomes from the protective antioxidant defense response and xenobiotic metabolism to DNA repair, cell proliferation and apoptosis [5]. Targets of stress-response pathways have become biomarkers indicative of both adverse outcomes (AOs) and molecular initiating events (MIEs). In the case of the latter for example, the up-regulation of antioxidant defense genes is commonly associated with the activation of the transcription factor (TF) Nrf2[6] (Table 1.2).

However, it has been widely established that the initiation of gene expression is mediated by the binding of multiple TFs[7] as well as processes such as methylation status and chromatin composition [8]. It is therefore unsurprising that whilst adaptive stress-response TFs are seen as acting on discrete biological processes, there is increasing evidence that suggests connectivity exists between factors in terms of sharing downstream target genes and through protein-protein interactions. Advances in genomic technologies such as chromatin-immunoprecipitation sequencing (ChIP-Seq) and assay for transposase-accessible chromatin using sequencing (ATAC-seq) have identified that stress-response factors can regulate a broader range of biological processes than previously identified; for example, HSF1, the primary TF activated under heat-stress has been associated with apoptosis [9] and Nrf2, a

regulator of antioxidants, with glucose metabolism [10]. In addition, increasing evidence exists for adaptive-stress response factors sharing downstream targets, for example, both Nrf2 and NFkB, traditionally associated with the inflammatory response, have both been shown to regulate the antioxidant gene, glutathione-s -transferase pi (GSTP) [11].

This creates a risk that assumptions based on the interpretation of gene-expression data and in the use of biomarkers as indicators of specific pathways could be inherently biased in defining molecular processes initiated under chemical and pollutant exposures. It is increasingly necessary for the interpretation of gene expression datasets to be considered as a result of interactions between multiple regulatory factors rather than single processes alone. The suggestion that adaptive stress-response pathways act within an integrated network has broad implications for toxicity testing; shared downstream gene targets imply that compensatory or canalized response processes occur where regardless of the inducer, there exists the potential for the same outcome to be reached. Evidence for this has been suggested by connectivity mapping approaches (Cmap), which identify connections between transcriptomic profiles and adverse outcomes, showing correlations between gene expression under different chemical exposures in the *Pimephales promelas* (fathead minnow) [12]. Given that adverse outcome pathway (AOP) and mode of action (MoA) frameworks are becoming widely used to identify toxicity processes from molecular initiating events (MIEs) to adverse outcomes (AOs), it is essential that the regulatory cascades that drive toxicity processes are better understood[13]. In this regard, binary interactions between adaptive stress-response pathways have been extensively reviewed [16–19] but the analysis of the response as an integrated network at the DNA-binding and protein-protein level is limited.

One of the key interactions that has been studied in the stress-response pathway is that between Nrf2 and NFkB in the context of oxidative stress, a mediator of adaptive stress-response processes[14]. Oxidative stress is defined by redox status, the ratio of oxidised and reduced glutathione (GSSG:GSH) and results from multiple mechanisms including the metabolism of xenobiotics[15] (Figure 1.3). Low levels of oxidative stress are associated with the activation of

Nrf2 and antioxidant defense processes that neutralise free-radicals to maintain homeostasis[16]. At higher levels of OS, the Nrf2 response is considered to be "overwhelmed" leading to NFkB and later P53 mediated gene transcription, resulting in inflammation and apoptosis respectively[17]. Nrf2 is considered to suppress NFkB activity and it is predicted that the absence of Nrf2 would lead to the activation of NFkB under inducers of adaptive-stress response processes. However, although interactions between Nrf2 and NFkB with other stress-response pathways are known, how this influences the response process has not been investigated.

Gene regulatory networks (GRNs) define the interactions between transcription factors and their downstream target genes. Comprising of key regulatory motifs, (circuits connecting nodes in the network such as positive and negative feedback loops (Figure 1.9), GRNs can inform on the relationship between observed patterns in gene-expression and the regulatory cascades that control biological processes [9]. Network inference in an ecotoxicology setting has largely been based on reverse engineering, considered a *top-down* approach where time-series omics data informs on the regulatory dynamics from the gene to metabolite level[18]. The method has been used to generate two models on the effects of oestrogenic chemicals on the hypothalamic-pituitary-gonadal axis in the fathead minnow[19] and *Oncorhynchus mykiss* (rainbow trout) [20].

Bottom-up approaches to GRN formation where prior knowledge on gene-regulatory interactions is collated from the literature to generate networks provide an alternative approach. This is an appropriate method where data on time-course interactions is limited or where the topography of underlying regulatory interactions between factors is not well defined. Whilst the *a priori* formation of GRNs has been widely adopted in medical and developmental biology, defining disease progression[21] and developmental patterning[22] respectively, the method has yet to be widely adopted in ecotoxicology. The call for reverse engineering approaches to inform on AOPs has been widespread[18] but GRNs based prior knowledge formed from robust experimental evidence on TF interactions would provide a novel platform for establishing networks at the basis of AOs.

Such analysis based on prior knowledge would be particularly beneficial for the adaptive stress-responses as chemical pollutants associated with targeting receptor mediated pathways at molecular initiating events (MIEs) are known to cause the upregulation of downstream targets of adaptive stress-response factors. For example, vertebrate exposures to pharmaceuticals which are specifically designed to act on individual pathways such as the oestrogenic chemical ethinyl-estradiol (EE2) and the estrogen receptor (ER), are widely shown to cause the upregulation of adaptive stress-response genes such as those involved in antioxidant defence[23]. Whilst the metabolism of chemical pollutants can initiate oxidative stress, the role of regulatory interactions between adaptive stress and receptor-mediated responses and the effect the activation of multiple pathways has on the course of adverse responses is unknown.

GRN analysis benefits from the ability to identify the sequential changes in gene activation (termed the regulatory states) in a systems-wide setting using mathematical principles to generate network-based model. The complexity of the mathematical model depends on the information available on network components as well as the size of the network itself. Small networks, where factors such as protein concentration, the rate of transcription, translation and messenger RNA (mRNA) degradation are known, are modeled using ordinary differential equations (ODEs) to provide quantitative predictions of gene expression under different exposure thresholds[24]. Where only knowledge of gene-regulatory interactions exists, boolean modeling approaches provide a qualitative estimate of network dynamics by representing gene regulation as either active or inactive[25]. These models give a predictive overview of network dynamics as a series of activation states which lead to a final *attractor,* comprising of state(s) which are repeated and indicate the model has become stabilized[26]. Attractors are largely associated with the observed phenotype[26]. Boolean modeling is an appropriate method for generalised models where little is known about biochemical dynamics. For adaptive stress-response pathways, boolean models of interactions between Nrf2 and the phosphoinositide 3-kinase (Pi3K) pathway, which regulates the cell cycle, have successfully predicted apoptotic responses, providing support for the adoption of this modeling technique in the context of the stress-response[25].

It has been widely demonstrated that adaptive stress-response pathways are evolutionary conserved across vertebrate groups with transcription factors and their downstream targets activated by the same chemical inducers across species[27,28]. Given that conserved regulatory interactions are widely demonstrated to be at the basis of robust regulatory networks[29], it is expected that the same systems controlling regulatory dynamics in mammals are likely to be conserved across distantly related species such as fish. Networks generated using mammalian data should therefore be transferable across vertebrate species, which is beneficial from an aquatic toxicology perspective where there is a lack of empirical data and limitations in genomic analysis that prevent the identification of species-specific regulatory networks.

This study employs a GRN approach to address how regulatory network connectivity between stress-response factors influences the response process through integrating the available knowledge on TF-DNA and protein-protein interactions in mammalian cell lines from the literature. DNA-binding events were included in the model if they were supported by direct validation of TF-binding through luciferase assays, enzyme-linked-immunosorbent assays (EMSA) and ChIP-Seq. The resulting GRN was modeled using boolean logic to identify the cascades of regulatory events and downstream response processes that are initiated through the activation of each transcription factor in the network.

The connectivity between oestrogenic pathways and the stress response was also explored as a case study to investigate the potential for receptor mediated pathways to activate adaptive stress-response processes. Simulations of the model identified that activation of Nrf2, AhR, MTF1, HIF-1α and HSF1 led to a canalized response with the same processes activated in attractor states. Knock-out analysis of Nrf2 and AhR caused the activation of NFkB and inflammatory response, largely predicted from the literature. Oestrogen pathways were able to activate stress-response processes in regulatory cascades.

2.3 Methods
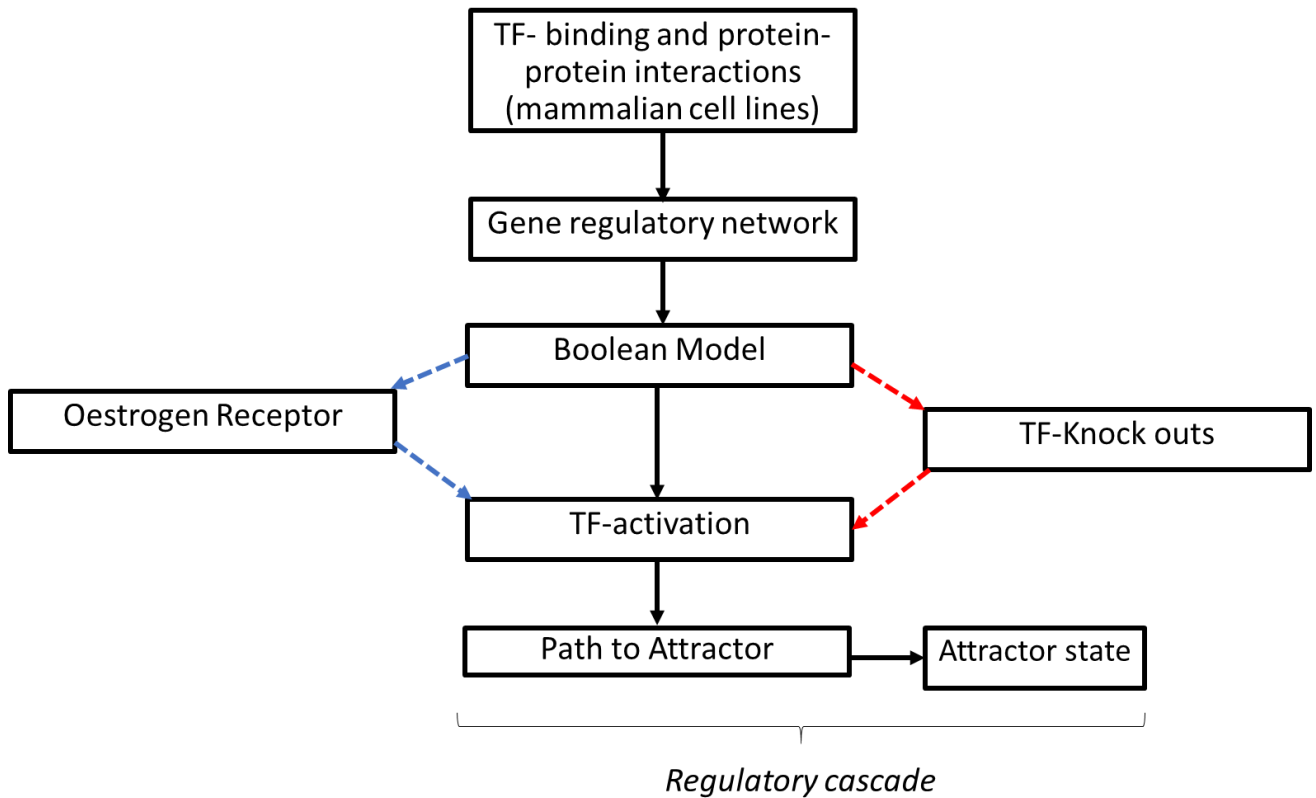
Figure 2.1: Method for identifying connectivity between stress-response pathways. The GRN was constructed based on evidence for TF DNA-binding and protein-protein interactions between adaptive stress response TFs (Figure 1.6) in mammalian cell lines based on the evidence from the literature (Table S2.2). The GRN was modeled using boolean logic with each stress-response TF activated independently in order to hypothesize which outcome processes and regulatory cascades would be reached under different exposure scenarios. Following this, perturbation analysis was conducted to identify the significance of each TF in regulating attractors by setting their activation to "0". Finally, additional network links between adaptive stress-response TFs and the oestrogen receptor where derived from the literature to establish if receptor-mediated pathways could initiate an adaptive stress-response cascade.

2.3.1 Mammalian Stress Response Network:

To create an accurate and inclusive GRN of the stress-response, interactions between Nrf2, AhR, HIF-1α, HSF1, MTF1, NFkB, P53 and downstream target genes were identified from a literature search of established direct DNA-binding and protein-protein interactions across mammalian cell lines (Table S2.2). The experimental evidence supporting DNA-binding events in the model was considered robust when the expression of the target gene was shown to be dependent on the specific TF in question binding to regulatory regions, largely identified through either;

1. Reporter-gene assays, where promoter regions are truncated to identify functional TF binding sites.

2. ChIP-PCR where an antibody-tagged TF is identified as interacting with DNA-sequences for the gene of interest.

3. Enzyme-linked immunosorbent assays (EMSA) showing the association of TFs with DNA sequences.

Protein-protein interactions were identified where experimental evidence strongly indicated interactions through ELISA or western-blot analysis between key regulatory factors (Table S2.2). The experimental evidence for interactions that were considered robust given the above was collated in Table S2.2 with abbreviations for cell lines shown in Table S2.1 respectively. This experimental evidence was used to construct a mammalian cell adaptive stress-response network using the GRN program, BioTapestry V7.1. [30], showing a graphical representation of the interactions between stress-response factors. Experimental evidence that was derived from whole-genome sequencing assays such as ChIP-Seq, ATAC-seq and DNAse 1 hypersensitivity and were not supported by direct gene-specific binding site analysis, were treated as putative. This acknowledges variations in quality between sequenced data interpretation and the potential that the identified sites may be functionally redundant [31].

2.3.2 Modeling the mammalian cell stress-response network:

The mammalian cell stress-response network was modeled using boolean logic by importing the BioTapestry GRN into the BoolNet package in R.studio [26] (Figure 2.1). A boolean model defines each node (gene) as being ON ("1") or OFF ("0") depending on the previous activity of its input nodes (regulatory genes and TFs) in the network and is defined by a series of logic rules indicative of the type of regulatory interaction as shown in Table 2.1 .

Rule	Symbol	Type
OR	|	Activator
AND	&	Both factors needed for activation
NOT	!	Inhibitor
XOR	$\oplus$	Exclusive or

Table 2.1 Boolean logic rules. The activity of a gene is determined by a series of logic rules; OR (|), AND (&), NOT (!) and XOR (Gene A & !GeneB)| (GeneB & !GeneA). OR logic shows that a regulator is active independent of other factors. AND logic identifies that both factors are necessary. NOT identifies inhibitory activity. XOR identifies which inputs are required at a particular time point. OR logic statements are used to describe regulatory processes between the same factors whereas XOR logic statements are used where multiple activators or inhibitors are involved in regulating specific genes.

In the GRN for the mammalian cell adaptive stress-response, the logic rules of interactions in the model were extrapolated from the literature and are defined in Figure S2.1. TFs known to form heterodimers prior to binding are in AND logic. Any change in the activity of the genes in the network indicates a change in regulatory state. All transitions between states eventually lead to an attractor that represents the final stabilized outcome of the network. The transitions between states at the base of the network are called the *basin of attraction* [26].

Attractor states can be *single-point,* containing only a single state or *cyclic,* whereby the attractor is composed of multiple states which occur in a cyclic pattern[26]. Prior to modeling, redundant network nodes (interactions which have no impact on the network model) were removed using the "*simplify.model*" command in BoolNet; this removed most protein-protein interactions that lack directionality and which have no regulatory impact in the analysis.

2.3.4 Model simulations:

The mammalian cell adaptive stress-response network models were run synchronously so that each discrete change in state would be included in the analysis. A SAT-exhaustive search was conducted where all combinations of nodes (genes) in the network are activated in the start state to determine all attractor states that can be reached by the model regardless of the input. Following this, the attractor states and basins of attraction for every stress-responsive TF in the GRN was determined independently by activating each TF in the start state by setting its activation to "1" with all other nodes = "0". Simulations were run till an attractor was reached and the output shows the start state, basin of attraction and attractor state for theoretical scenarios where a single TF is activated by a chemical pollutant at a MIE.

2.3.4 Network perturbation analysis of TFs:

To investigate the extent of crosstalk between pathways, the number of original attractors that could be returned from the model following the knock out (KO) of each TF independently was determined. A SAT-exhaustive search was conducted following the KO of each TF independently and the returned attractors were compared to the original model.

2.3.5 Exploring Nrf2 and AhR knockdowns in modelling outcomes:

To determine the role of Nrf2 and AhR in militating against the inflammatory response mediated in terms of NFkB and P53, the model was simulated in the absence of Nrf2 and AhR with their expression fixed to "0". MTF1, HSF1, NFkB, P53 and HIF-1α were activated in the start state and the path to attractor and attractor state were recoded.

2.3.5 Oestrogen receptor interactions with the mammalian cell adaptive-stress
response GRN:

To assess the role of interactions between receptor mediated pathways and the adaptive stress-response, interactions between oestrogen receptor alpha (ERα) and oestrogen receptor beta (ERβ), responders to endocrine disrupting chemicals, were added to the GRN model (Table S2.2). Interactions between ERα and ERβ and adaptive stress-response pathways were established from experimental evidence in mammalian cell lines as described in a literature search (Section 2.3.1) and are shown in Table S2.2. A boolean model containing the new interactions with ERs was generated (Figure S2.2) and simulations of the model were conducted by activating each ER independently and together in the start state of the network following the same method as stated in section 2.3.3.

2.4 Results

2.4.1 Mammalian cell adaptive stress-response networks:

The literature search identified interactions between 59 nodes in the mammalian cell adaptive stress-response network (Table S2.1). The data was collated into a gene regulatory network (Figure 2.2) with key positive regulatory interactions between shared downstream target genes shown in Table 2.2.

Downstream target genes linked with specific adverse or protective response outcome processes are grouped accordingly (Figure 2.2). Pathways had differing levels of connectivity with Nrf2-AhR and HIF-1α-NFkB-P53 both enhancing the expression of shared downstream targets and were either directly or indirectly, able to activate one-another (Figure 2.2). Nrf2 and NFkB, which indirectly inhibited each other's expression, shared downstream target genes including those involved in antioxidant defense process (glutathione-s-transferase pi (GSTp), NAD(P)H dehydrogenase (NQO1) and sequestosome 1 (SQSTM1) as well as the inflammatory response mediated by interleukin 6 (IL6) and interleukin 8 (IL8) (Table 2.2). The shared targets within this group were also able to initiate a cascade of processes through protein-protein interactions with additional stress-response factors (Figure 2.2). Heme-oxygenase 1 (HMOX1) positively regulates the expression of HIF-1α by stabilizing the factor in a protein-protein interaction and therefore allowing the activation of downstream processes (Table S2.3). NQO1 positively regulated HSF1 through a similar mechanism (Table S2.3). SQSTM1 is a scaffold protein, able to positively regulate NFkB through phosphorylating IKKb [32] as well as being involved in a positive feedback mechanism with Nrf2, through competitively binding to Keap1 [33](Table S2.3) .

Metal toxicity was identified as being regulated by multiple transcription factors including Nrf2 [34], MTF1 [35], HIF-1α [36] as well as the downstream target IL6[37]. Metallothioniens were found in positive feedback mechanisms with other nodes in the network k. In this case, metallothionein 1 (MT1) and metallothionein 2 (MT2) have been identified as being able to positively regulate NFkB by releasing zinc, a process necessary for NFkB binding [38]. The links

104

identified a mechanism in which positive feedback cascades can operate via metallothionein expression (Figure 2.2).

In the literature analysis, vascular endothelial growth factor (VEGF), a common biomarker for HIF-1α, was also identified as being directly regulated by P53 [39] and NFkB as well as existing in an indirect positive feedback loop with Nrf2 [40]. As well as direct and indirect regulatory interactions, the literature analysis identified common responders between molecular chaperone proteins ARNT, heat-shock protein 70 (HSP70) and heat-shock protein 90 (HSP90) (Table S2.2). These proteins are able to bind to TFs in the cytoplasm to initiate the activation of pathways. Whilst the interaction between the ARNT and HIF-1α/HSF1 has been widely reported [41], it has been established that HSPs are regulated by multiple TFs including, NFkB and Nrf2 [42] as well as HSF1 [43].

Whilst the majority of downstream target genes had no regulatory influence on the network, some had regulatory roles on stress-responsive factors and are all considered biomarkers of antioxidant defense processes. These genes included HMOX1, SQSTM1 and NQO1, which interact with other factors at the protein-protein level (Table S2.2). In addition, the co-activators ARNT and Hsp70 had multiple roles in stress response processes as chaperones for TFs identifying a potential for competitive protein-protein interactions between regulators.
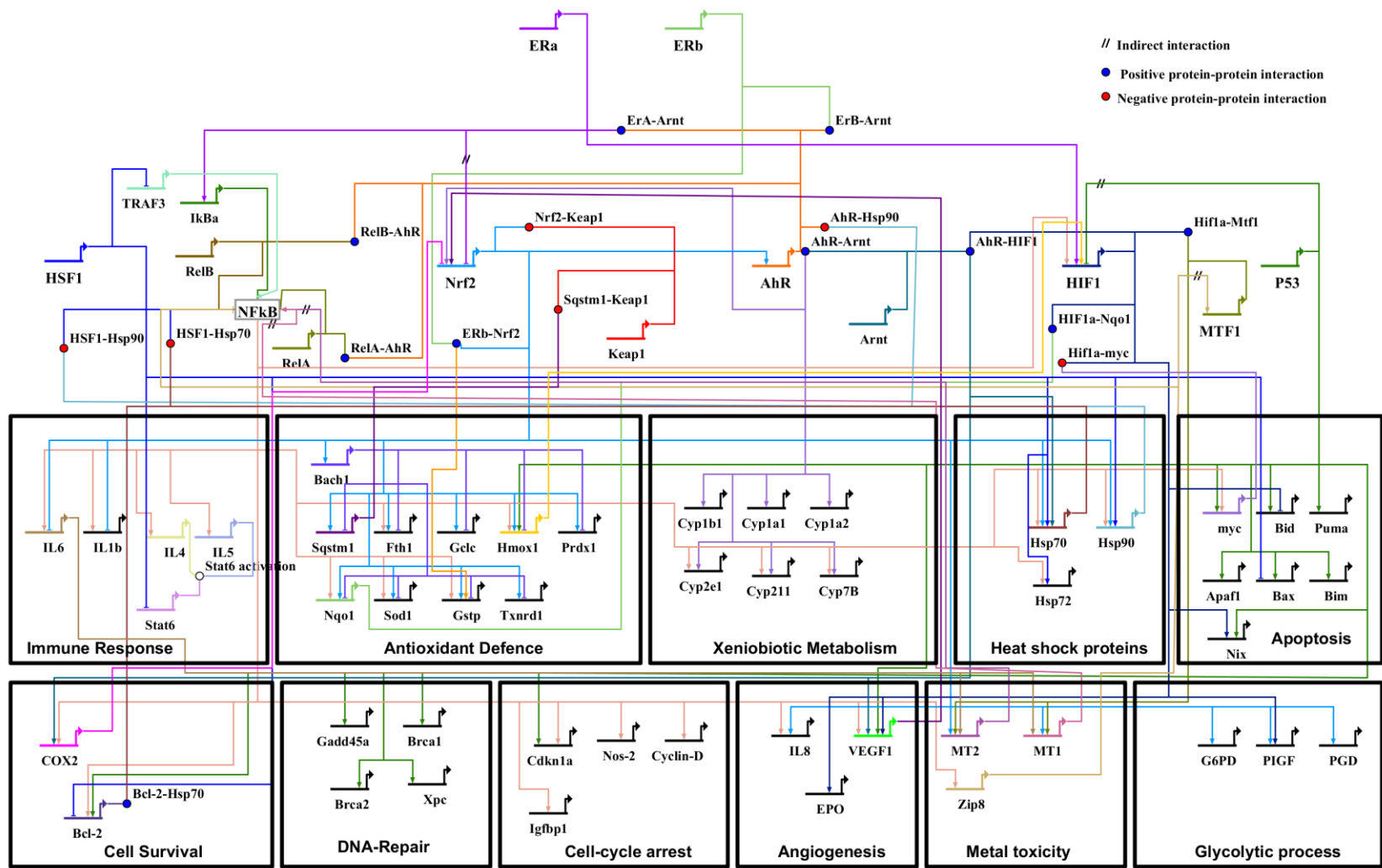
Figure 2.2: Mammalian cell adaptive stress-response network. Network interactions between stress response pathways (Figure 1.6) with transcription factors at the top of the model and downstream targets, grouped into modules relating to response processes below. Each module contains key biomarkers of response processes that are widely used target genes in toxicity testing. All nodes that directly bind to promoter regions are established based on validated DNA-binding events in mammalian cell lines as shown in Table S2.3. The BioTapestry model symbolises TF binding interactions as inhibitory = ⊥ or activating = ↓ with protein-protein interactions = ○. Indirect interactions are shown with a // through the interaction line and include evidence derived from whole-genome sequencing where the specificity of binding has yet to be validated further through direct protein-DNA binding analysis (Gene abbreviations are shown in Table S2.2).

Gene	Marker	Activating TFs/Genes
Transcription Factors		
HIF-1A	*Hypoxia*	NFKB, HMOX1
MTF1	*Metal tox.*	NFKB
AHR	*Xenobiotics*	NRF2
NRF2	*Oxidative Stress*	AHR, VEGF1, SQSTM1
NFKB	*Inflammation.*	TRAF2, MT2,
Antioxidants		
FTH1	NRF2	NRF2, NFKB
HMOX1	NRF2	NRF2,NFKB, p53,
GSTP	NRF2	NRF2, AHR, NFKB,
SOD1	NRF2	NRF2, NFKB
NQO1	NRF2	NRF2, NFKB
TXNRD2	NRF2	NRF2, MTF1
Cell survival		
COX2	HIF1	HIF-1A, NFKB, AHR-HIF-1A
BCL2	P53	P53, NFKB
Cell-cycle arrest		
CDKN1A	NFKB	NFKB, P53
Angiogenesis		
VEGF1	HIF1	NFKB, P53, AHR-HIF1
IL8	NFKB	NFKB, NRF2
Metallothioneins		
MT1	MTF1	MTF1, NRF2, IL6 (NFKB)
MT2	MTF1	MTF1, NRF2, IL6 (NFKB)
Glycolytic Processes		
PIGF	NRF2	NRF2, HIF-1A
Apoptosis		
MYC	P53	P53, NFKB
Heat-shock Chaperones		
HSP70	HSF1	HSF1, NRF2, AHR-HIF1
HSP90	HSF1	HSF1, NRF2, NFKB
HSP72	HSF1	HSF1, NFKB
Xenobiotic metabolism		
CYP1A1	AhR	NFKB

Table 2.2: Cross-talk between transcription factors and downstream targets in the mammalian cell adaptive stress-response network. Positive regulatory interactions between transcription factors and downstream targets as shown from evidence collated from the literature search on mammalian cell lines (Table S2.2) and represented in

Figure 2.2. For regulatory interactions with AhR-HIF-1α, evidence suggests a heterodimer is formed so both factors need to be active to initiate gene transcription.

2.4.2 The mammalian cell stress-response network produces four attractor states:

Modeling of the mammalian cell stress-response network identified that four potential attractor states could be reached by activating any combination of genes (a SAT-exhaustive search) in the start state of the model simulation. The four attractors comprised of two single point and two cyclic attractors as shown in Figure 2.3.

Genes involved in all outcome processes were activated in attractor 1. All genes involved in apoptosis are activated but only one gene in the immune response, IL8, and in angiogenesis, VEGF, were active. Active TFs were Nrf2, AhR and P53. In attractor 2, TFs active include Nrf2, AhR and HIF1α. Downstream targets included those involved in antioxidant defense, xenobiotic response, heat-shock chaperones, angiogenesis, metal toxicity and glycolytic processes. Only interleukin 8 (IL8) was active under immune response processes and B-cell lymphoma 2 (BCL2) in apoptosis (Figure 2.3).

Attractor 3 is a cyclic attractor comprising of two states (Figure 2.3). In state 1, Nrf2 and HIF-1α were both activated whereas in state 2, AhR is the only active TF. Downstream targets active in state 1 of the attractor include those involved in xenobiotic metabolism and the BCL2 gene. In state 2, antioxidant defence processes, xenobiotic metabolism, heat shock chaperones, metallothioneins and genes involved in glycolytic processes were all active. IL8 was the only gene involved in inflammatory responses to be activated (Figure 2.3).

Attractor 4 is a cyclic attractor comprising of three states. In state 1, NFKB, Nrf2 and the AhR were activated with metallothioneins, VEGF and the glycolytic gene, PIGF. In state 2, HIF1α and the AhR were active with downstream targets including antioxidant defence genes, xenobiotics, heat shock chaperones, immune response factors, cell-cycle arrest, metal toxicity and glycolytic processes. NIX involved in apoptosis was active as was Signal transducer and activator of transcription 6

(STAT6). In the final state of attractor 4, HIF1α and MTF1 were activated. Downstream targets included xenobiotics, STAT6, BCL2, MT2 and PIGF.

All genes in the antioxidant defense system, xenobiotic response processes and glycolytic gene pathways were activated in all attractor states. For antioxidant defense genes and xenobiotic responses, the activation of genes was state-dependent in attractors 3 and 4, highlighting that these processes are not continuously active. This was also the case for glycolytic processes in attractor 3, but not attractor 4, where PIGF was activated in all states but not Glucose-6-Phosphate Dehydrogenase (G6PD) or Phosphoglyconate Dehydrogenase (PGD). Genes involved in angiogenesis were also activated in all attractors; VEGF was activated in all attractor states with the exception of attractor 3.
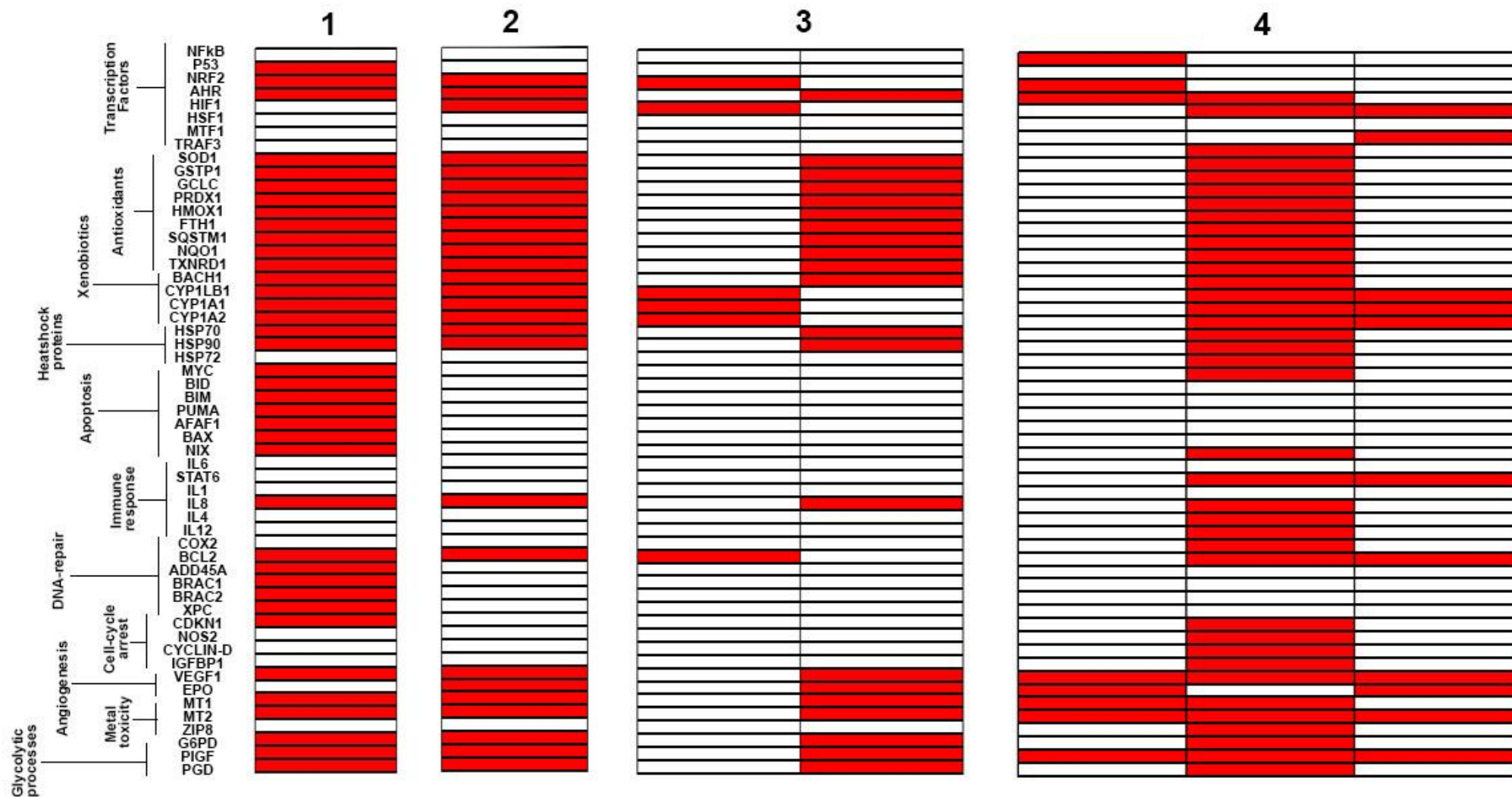
Figure 2.3: Attractors in the mammalian adaptive stress-response network. An exhaustive search of the mammalian stress-response network identified four potential attractor states could be reached by activating any combination of genes in the start state. Attractors

1 and 2 are *single-point* and 3 and 4 are *cyclic*. Active genes are shown in red and inactive white. Gene names and processes are shown in the left-hand column.

2.4.3 Simulations of transcription factor activation:

The single point attractor 1 (as shown in Figure 2.3) was the most common outcome of the TF-specific network simulations, comprising of biomarkers activated in the xenobiotic response, antioxidant defense, metallothioniens, heat-shock chaperones, glucose metabolism and angiogenesis. Simulations of the model with the AhR, HIF-1α and MTF1 activated independently in the start state each reached attractor 1 (Figure 2.5). Regulatory cascades initiated in the case of HIF-1α and MTF1 caused the initiation of the AhR receptor in state 2 from which the basin of attraction matched that of the AhR model until the attractor was reached (Figure 2.5).

Nrf2, P53 and the AhR were stabilised to attractors after 3 state transitions in the basin of attraction and went on to each reach a different attractor state (Figure 2.4). For Nrf2, activation lead to the induction of the AhR in state 2, the first state of the cyclic attractor and HIF-1α in state 3 (Figure 2.5) where the simulation reached attractor 3 (Figure 2.4). The model simulation with HSF1 activated in the start state resulted in the same attractor as Nrf2 but with an additional state in the basin of attraction (Figure 2.5). Like with MTF1 and HIF-1α simulations, in the HSF1 simulation, the AhR was activated in state 2. The activation of chaperone proteins (HSP72, HSP70, HSP90), in state 2 of the HSF1 simulation caused the network to reach the same attractor as Nrf2 and not the AhR despite all other active genes being the same as MTF1 and HIF-1α models.

The NFkB simulation induced a different attractor state, the three-point cyclic attractor 4 (Figure 2.3). Activation of NFkB in the start-state led to induction of HIF-1α in states 2 and 3 as well as MTF1 in state 3. The attractor state differed from the basin of attraction and the TFs NFkB, Nrf2 and the AhR were all active in the first state of the cyclic attractor. This was followed by the activation of HIF-1α in the second and third state (Figure 2.4). Target genes active in the final attractor state consisted of those involved in angiogenesis, immune response and DNA-repair with the activation of the BCL2 gene.

The p53 model simulation was the only attractor where genes involved in cell cycle arrest were activated. The p53 simulation led to the downstream activation of Nrf2 and the AhR, which were also active in the final state of the attractor where the simulation reached attractor 1.

Simulations of the model where HSF1, HIF-1α and MTF1 were activated in the start state were not active TFs in final attractors. However, for each TF, their predicted downstream target genes (heat shock chaperones for HSF1, angiogenesis genes for HIF-1α and metallothioneins for MTF1) were all active in the attractor state(s). In all cases, TFs active in the basin of attraction were also active in attractor state(s).
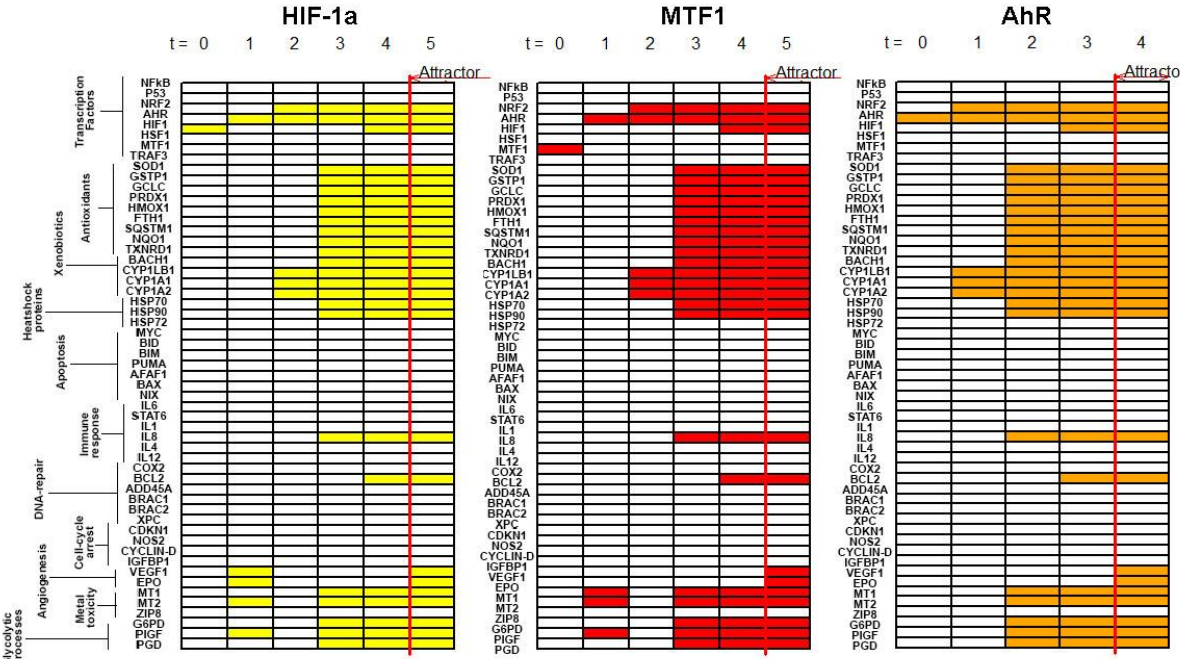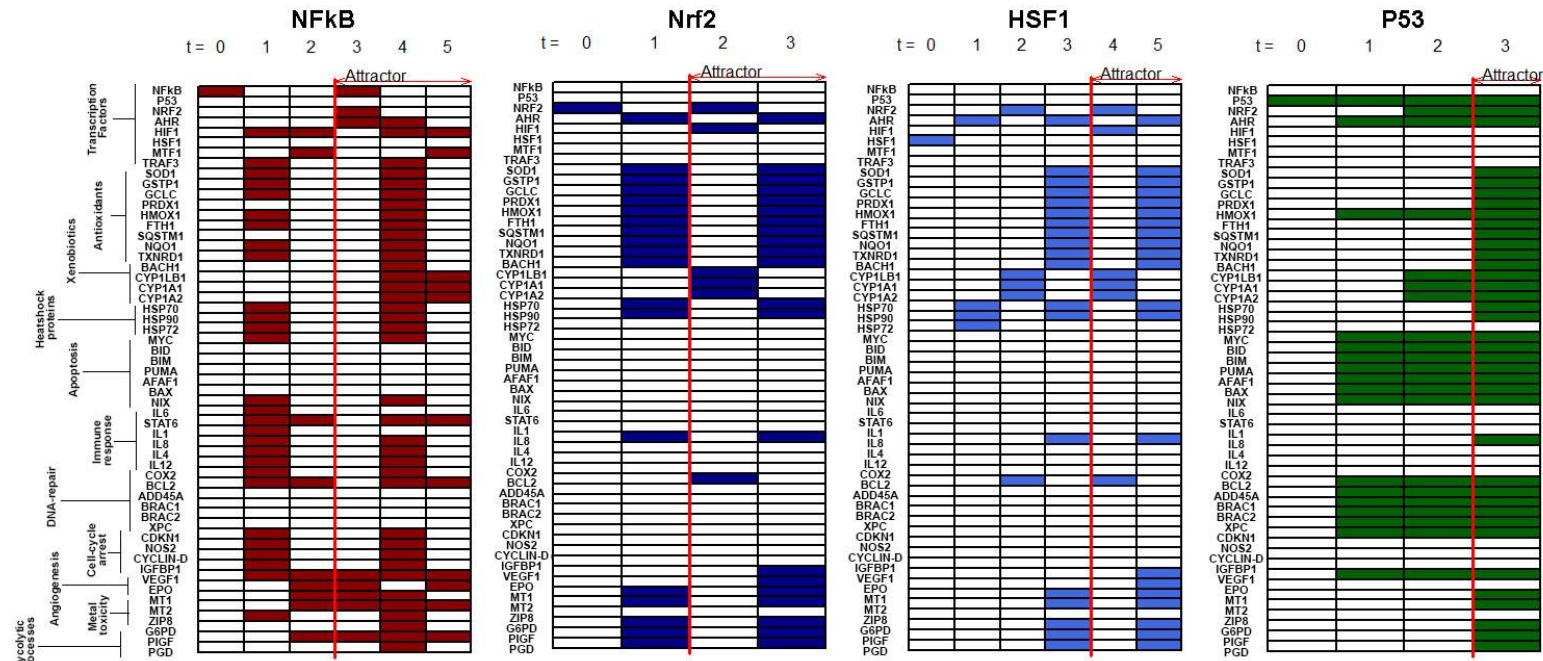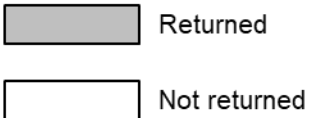
Figure 2.4: Basin of attraction in the mammalian adaptive stress-response network where each TF is activated independently in the start state. A). The basin of attraction for model simulations where each TF is independently activated in the start state. State transitions are indicated (t =), and colored boxes indicate genes are active. White boxes indicate inactive genes. Red arrows indicate the attractor state(s) with the columns after t = 0 showing the basin of attraction. Gene names and processes are shown in the left-hand column. B). Summary of the results shown in A with the attractors (Attr.) for each TF simulation based on the results presented in Figure 2.4. The TFs column in the table indicates the TFs that are active within the attractor state. In all cases, TFs active in the attractor state are also active in the basin of attraction and therefore essential components of the regulatory cascade initiated following the activation of the TF concerned.

2.4.6 Network perturbations on attractor state:

Knockout analysis of each key transcription factor was conducted to determine their influence on the ability of the model to reach the four potential attractor states (as shown in Figure 2.3). In each simulation, a TF was knocked out (KO) of the model and the ability to reach the original attractors was determined (Figure 2.5). This showed that all attractor states were reliant on Nrf2 and the AhR to reach the end state. Attractor 1 required the addition of HIF-1α and p53. Only attractor 4 required NFkB and MTF1. HSF1 was not necessary for the model to reach any of the original attractor states.

KO	Attractors 1	2	3	4		
NFkB	▓	▓	▓	☐	▓ Returned	
Nrf2	☐	☐	☐	☐		
HSF1	▓	▓	▓	▓	☐ Not returned	
P53	☐	▓	▓	▓		
HIF1a	☐	▓	▓	▓		
MTF1	▓	▓	▓	☐		
AhR	☐	☐	☐	☐		

Figure 2.5: Effect of TF perturbations on attractor states in the mammalian adaptive stress response network. Table representing the ability to reach the 4 attractor states in the original network following knockouts (KO) of each of the factors in the network and a sat.exhaustive search of the model. Shaded boxes indicate attractor was not returned and white boxes show the attractor was returned following perturbations and comparisons of attractors to the original model results.

2.4.6 Network perturbations to Nrf2 and AHR activation:

Considering that Nrf2 and the AHR were essential mediators of network models, the knockout of Nrf2 and AHR was conducted to determine how this changed the course of attractor. It is hypothesized that the removal of AhR and Nrf2, which are involved in phase 1 and phase 2 detoxification processes would lead to an increase in cell-death and inflammatory responses. Knock outs of AhR and Nrf2 where HIF-1α, NFkB and MTF1 were activated in the start state independently lead to the same attractor state (Figure 2.6). This cyclic attractor is comprised of 4 states where genes involved in cell-cycle arrest, immune responses, angiogenesis and antioxidant defense processes were activated. For HSF1, no gene processes were activated in the attractor state.
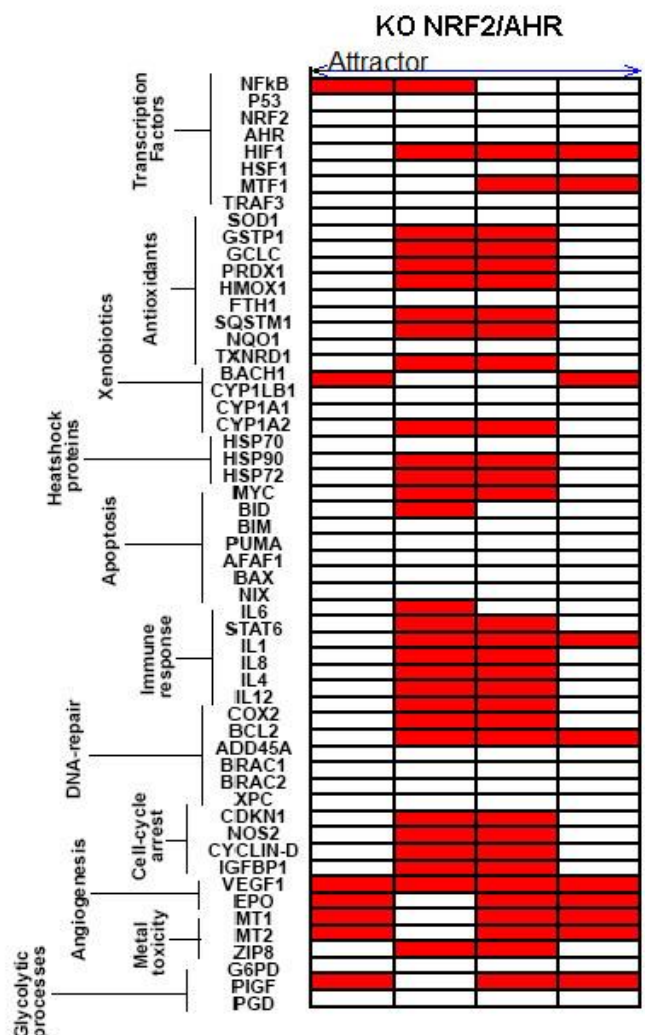
Figure 2.6: Conserved attractor of KO of NRF2 and AHR following the activation of MTF1, HIF-1α and NFkB in model simulations. Cyclic attractor reached through the KO of AhR and Nrf2 from the mammalian cell stress-response network following simulations of MTF1, HIF-1α and NFkB in the start state. State transitions are indicated (t =), and colored boxes indicate genes are active. White boxes indicate inactive genes. Gene names and processes are shown in the left-hand column. The attractor comprises of the activation of NFkB, HIF-1α and MTF1. Downstream target genes are associated with processes include cell-cycle arrest, immune responses, angiogenesis and antioxidant defence.

2.4.6 Interactions between oestrogen receptors and stress-response network
dynamics:

The addition of oestrogenic regulatory interactions with adaptive stress-response pathways produced a network with 14 potential attractor states. Boolean modeling identified a divergence in ERα and ERβ in their ability to regulate distinct outcomes (Figure 2.7).

Regulatory cascades initiated after the induction of ERα caused the activation of NFkB, AhR, HIF1α and MTF1 (Figure 2.6) and returned a novel attractor in comparison to Figure 2.3. No downstream target genes were activated until the attractor state. In state 1 of the attractor, antioxidant genes involved in xenobiotic metabolism, heat-shock chaperones, inflammatory genes, cell-cycle arrest and metal response processes were active. The glycolytic gene PIGF was active in all states. Targets active in state 2 comprised of only STAT6, Growth Arrest and DNA Damage Inducible Alpha (GADD45A), metal toxicity and angiogenesis. State 3 of the attractor was composed of metal toxicity, angiogenesis, PIGF and the antioxidant Therodioxin Reductase 2 (TXNRD2). In contrast, ERβ activated a different regulatory cascade with Nrf2, AhR and HIF1α all being activated in the basin of attraction and the attractor state (Figure 2.7). The simulation reached the single-point attractor 2 of the original model shown in Figure 2.3, which was shared by the AhR, HIF-1α and MTF1. The simulation with both ERα and ERβ activated in the start state resulted in the same attractor as reached by the activation of ERα independently (Figure 2.7).
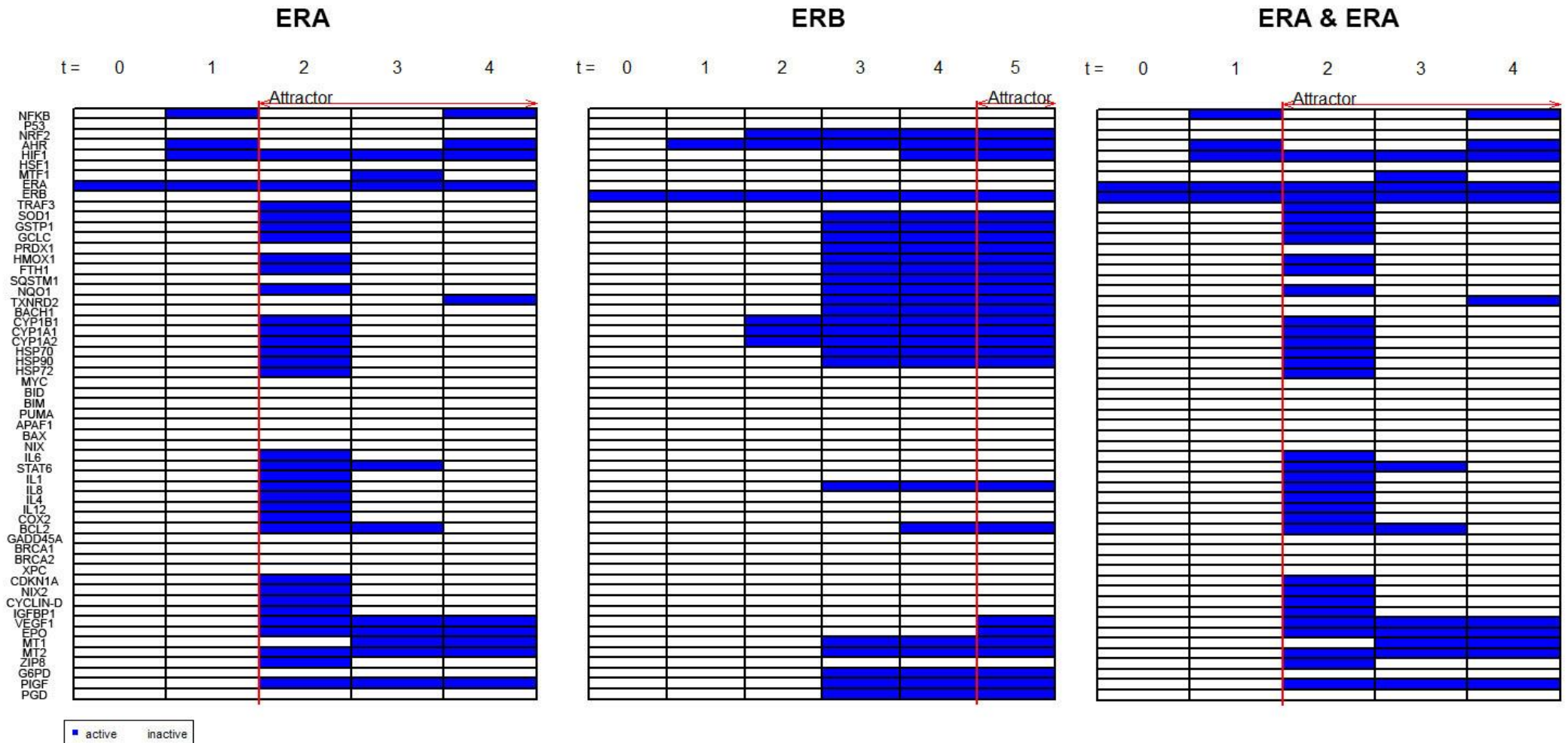
Figure 2.7: Path to attractor for the mammalian network following the addition of estrogen-receptor interactions. The basin of attraction for model simulations where ERα and ERβ are activated together and independently in the start state. State transitions are indicated (t =), and colored boxes indicate genes are active. White boxes indicate inactive genes. Red arrows indicate the attractor state(s) with columns after t = 0 showing the basin of attraction. Gene names and processes are shown in the left-hand column.

2.5 Discussion

The adaptive stress response comprises of an evolutionary conserved and highly interconnected series of molecular mechanisms that are rarely considered in a holistic system. This research sought to identify how connectivity in the regulatory landscape can inform predictions on potential adverse outcome processes for TFs that are widespread targets of environmental pollutants. Using a GRN approach, the boolean model presented in this study provides a theoretical framework for identifying outcomes following the activation of selected stress-response TFs based on characterised interactions at the DNA and protein level from mammalian studies.

2.5.2 Identification of canalized response outcomes in adaptive stress response processes.

Genetic canalization refers to a developmental biology concept proposed by Waddington that organisms develop the same phenotype despite environmental perturbations and therefore, regardless of the input, the same outcome can be reached[58]. The GRN of the mammalian cell-stress response network identified multiple levels of cross-talk between TFs and downstream target genes which indicated a level of canalization can occur in stress-response processes.

Model simulations of the mammalian cell stress-response network identified that the activation of the AhR, MTF1, HIF-1α, Nrf2, HSF1 and ERb caused the same processes to be activated despite TF pathways being traditionally associated with specific outcome events. Phase I and II detoxification, metal toxicity and glycolytic processes were all involved in the final attractors reached through model simulations with each factor activated in the start state. These processes related to sharing roles in detoxification and cellular homeostasis. The canalization of response processes was largely through the activation of the AhR in the basin of attraction initiating a positive feedback

loop with Nrf2, which has been shown to be conserved across vertebrate species [45,46–48]. KO analysis identified that the AhR and Nrf2 were essential for all final attractors in the model despite multiple factors having roles in regulating the antioxidant targets of Nrf2. This research suggests that AhR and Nrf2 are both keystone factors in stress-response processes and establishing their roles across chemical treatments could provide a vital indicator of pathway activation.

As attractor states show the steady-state response in the network[25], the results from model simulations suggest that inducers of stress-response processes cause the same outcome events. Whilst it is currently difficult to extrapolate the response trajectories from the literature alone, correlations exist between the up-regulation of metallothioneins, xenobiotic responses, and antioxidant defense processes and exposures to metals[48], heat-stress [49] and hypoxia[41]. In addition, correlations between gene-expression outcomes have been identified for inducers of HIF-1α and MTF1 [36] but further research is necessary to identify if outcomes are conserved across other traditional inducers of stress-response pathways.

2.5.2 Biomarkers of response pathways were activated in the absence of predicted TF regulators.

Model simulations demonstrated that downstream targets associated with specific TF could be activated in the absence of their traditional regulator. This was the case for MTF1 and HIF-1α where downstream targets were activated in attractor states despite these TFs being inactive. Similarly, where Nrf2 was KO from model simulations, antioxidant genes were still activated, largely through the regulatory role of NFkB. This creates a risk of false positive associations being made between downstream target genes and MIEs. Such a result can be challenging to validate from experimental datasets where only HTS data exists in the absence of regulatory evidence (e.g. RNA-seq data alone) and this suggests that molecular activators may be falsely assumed in reverse engineering approaches.

However, the fact that responses to Nrf2, MTF1, AhR, HIF-1α and HSF1 were canalized to the same outcome processes creates the argument that it may not be necessary to know the specific MIE at the start of the regulatory cascade where an identical outcome is reached. However, the model presented in this research is simplistic, with no indication of dose response, which is likely to have an impact on adverse outcomes reached. Additional parameters in the mammalian cell stress-response network, such as sub-setting the network by tissue-specific components, is likely to aid in the identification of response processes.

2.5.3 KO analysis of Nrf2 and the AhR provide support for the mammalian cell-stress response GRN.

NFkB and p53 are widely associated with higher levels of toxicity, initiating processes involved in the immune response, apoptosis and cell survival [50,51]. Model simulations with NFkB and p53 were the only simulations where cell-cycle arrest and cell-survival genes were activated. Oxidative stress, an inducer of adaptive stress-response pathways, is associated with activating Nrf2 at low and NFkB under high levels as indicated by internal GSH:GSSG, with regulatory responses changing as a factor of redox status[52]. Model simulations with HIF-1α and MTF1 in the absence of Nrf2 and AhR caused an attractor state comprising of NFkB activation and the expression of genes involved in angiogenesis, cell cycle arrest and the immune response. This provides support for the model and suggests that KOs of AhR and Nrf2 indicate an artificial insight into a dose-response. The literature provides some support of this outcome with the NFkB induction of pro-inflammatory responses induced by high-levels of zinc exposures identified in the rainbow trout [53]. However, in contrast, high levels of hypoxia have been shown to cause an increase in *hsp70* in flounder (*Platichthys flesus*), a gene that was not regulated in the KO simulation[54].

2.5.5 Activation of oestrogen receptors initiated an adaptive stress-response cascade at a gene regulatory level.

Exposures to oestrogenic chemicals are related to stress response process and the up-regulation of biomarkers correlating to antioxidant and xenobiotic response processes has been previously documented [55]. The literature analysis identified multiple connections between oestrogen receptors and stress response factors including P53 [56,57] and Nrf2[58]. Model simulations varied between ERa and ERb with ERa induction leading to inflammatory response processes in contrast with ERb which caused antioxidant defences. It has been widely demonstrated that the metabolism of oestrogenic chemicals causes the production of quinone containing compounds, potent activators of Nrf2 [59]. The activation of multiple pathways in the start state was not explored in this research as it is unknown whether the timing of such response processes, particularly when mediated by metabolites, would occur. Nevertheless, as the activation of antioxidant defense processes was not continuous in attractors under ERα/β expression alone, this suggests a mechanism whereby antioxidant defenses are activated under prolonged exposure periods. Therefore, whilst the adaptive stress response process is considered a general response to chemical exposure, the incorporation of additional network nodes is essential for predictions of AOPs.

2.5.7 Model development and considerations for GRN use in an ecotoxicology setting

The model presented provides an initial step in understanding the holistic response to stress-induced pathways from a bottom-up systems biology standpoint. However, there are multiple considerations that need to be made when interpreting data of this type in a systems-wide setting. Firstly, the difficulty in distinguishing between primary and secondary response processes, representing the start state and secondary responses that correlated to changes in regulatory state, mean that correctly simulating models to mimic environmentally relevant exposure scenarios is challenging.

As is the case, where outcomes are known, activating multiple components of a model in the start state will be necessary to infer MIE based on observed outcomes. Where there is a lack of time-series data on gene-expression, it may only be possible to infer regulatory state transitions rather than attractors. Such analysis should be conducted alongside current reverse engineering processes to determine how the activation of combinations of pathways causes specific response profiles.

As with all boolean models, the scientific plausibility of representing responses as binary factors needs to be considered. Where threshold responses have little influence on the network dynamics, it could be argued that modeling the response as a boolean system is sufficient to gain enough evidence to predict outcome processes. For stress-response pathways however, threshold responses where chemical dose impacts the level of gene expression, have been widely cited as having a significant influence on the response outcome. In this case, the internal levels of the redox regulator glutathione have been widely associated with the stress-response outcome and the ability of the cell to return to homeostatic conditions. The incorporation of glutathione-redox dynamics into a model of this type would greatly benefit the inference of AOP, particularly where the association of the levels of stress are linked to regulating key gene processes such as high levels of oxidative stress activating NFkB pathways whereas low, acute levels initiating Nrf2. The model presented here goes some way in interpreting this system by activating factors independently but requires clear knowledge of the MIE in relation to the level and duration of the inducer. Where the activation of pathways is less clear such as the dynamics of p53 outcomes, which can be apoptotic or protective, this would certainly add clarity to the modeling outcomes.

Sub-setting networks by cell or tissue type have widely improved the outcomes of system-level models[60]. Downstream gene targets of the stress-response have been shown to have tissue-restricted expression patterns, for example antioxidant genes in zebrafish [6], and the presence of key genes and factors within a cell will therefore influence the outcome. In addition, DNA-

binding and protein-protein interactions only partially account for the mechanisms in which genes are regulated. The influence of methylation status, chromatin composition, the presence of co-factors in TF-binding and post-translational processes all have roles in gene-expression[61]. Efforts to understand the influences of these systems is increasing but are not as well documented as DNA-binding events. Where shown to have a significant effect, data derived from these experimental systems should be incorporated into models of this type to help inform on outcome events.

Finally, systems biology approaches undertaken for ecotoxicology purposes would clearly benefit from a collated database of transcription-factor binding events supported by robust experimental evidence derived from multiple sources and across distantly related species. Such gene-regulatory information could then be easily incorporated into boolean and other modeling systems, particularly where novel functions of TFs on biological processes have been defined. This would add an additional source of information to AOP knowledge base (AOP-KB)[62] to create *in silico*-based predictions. In addition, the further characterisation of species-specific gene ontogeny profiles, such as those currently available for zebrafish on the ZFIN site, could also be incorporated in tissue-specific models, informing on future *in silico* research efforts.

2.5.1 Network interactions were established across multiple cell lines

The network presented in this study was inferred from multiple human, mouse and rat cell lines that largely characterised protein-DNA binding events through LUC-reporter gene assays. Data was therefore integrated from biologically distinct environments, creating potential challenges in network inference due to the high probability that lines will exhibit differences in response to inducers. Aspects such as the cellular-specific presence of co-factors, which impacts the ability of TFs to bind to their respective TFBS, can mediate these changes in cell-line specific responses [60]. The information on the cell lines selected is available in the supplementary information, providing

the potential for networks to be compartmentalized by line-type as has been done previously [60].

2.6 Conclusion

The model presented in this chapter provides evidence for the possibility that the activation of different TFs in the adaptive stress-response can initiate the same outcome, identifying the potential for canalization of responses in a toxicology setting. The model presented provides a simple, yet informative, qualitative assessment of connectivity between selected factors within the vertebrate adaptive stress-response network. It provides a platform for further rigorous and species-specific testing of the regulatory network motifs identified in this research and highlights a growing understanding of the integration of biological processes and outcome pathways. Whilst limited by a lack of empirical data on tissue-specificity, threshold concentrations and species specificity, it presents an analysis of potential interactions within biological systems based on the DNA-sequence and TF conservation. Further research into the conservation of regulatory regions across vertebrate species for adaptive stress-response factors is necessary to determine the extent of transferability of this model across species. In addition, validation of this model using data for exposures derived across vertebrates in accordance with the Bradford hill considerations[63] is necessary to determine the plausibility of using this model as a novel method for inferring regulatory cascades in AOPs.

Chapter 2: A Boolean model of connectivity in vertebrate stress response pathways.

2.7 Supplementary Information

This supplementary information contains:

Table S2.1: Cell-line abbreviations.

Table S2.2: Gene abbreviations.

Table S2.3: Chemical abbreviations.

Table S2.4: Validated of mammalian interactions between TFs and DNA sequences.

Figure S2.1: Network Logic Rules for BoolNet models.

Figure S2.2: Network Logic Rules for BoolNet models with oestrogen receptor links.

Table S2.1: Cell line abbreviations.

ABREV.	CELL LINE NAMES
7-m12	Human prostate cell line.
a549	Human Lung
hek293	Human embryonic kidney cells
hmler	Human breast cancer cell line
jeg-3	Human choriocarcinoma cells (placental)
k562	Human Erythroleukemia cell line
M3	Human recombinant Muscarinic Acetylcholine receptor
MDA-MB-435S	Human melanocytes
plb985	Human blood cells
PLF2	Proliferin-2
raw264.7	Mouse macrophage
Hepa-1c1c7	Mouse hepatoma
32d	Myeoblast-like cell line (mouse bone marrow)
A431	Epidermoid carcinoma
AGS	Human stomach gastric andenocarcinoma
ASM-SC	Human melanoma cells
beas-2b	Human bronchial epithelial cells
HMEC	Human mammary epithelial cells
cos-1	African green monkey kidney fibroblast
cos-7	African green monkey kidney fibroblast
dhl-4	Human diffuse large cell lymphoma
fls	Fibroblast-like synoviocytes
FS-4	Human fibroblast line
h1299	Human P53 deficient lung-cancer cell line.
NCI-H358	Human metastatic lung cells
hacat	Human immortalised keratinocytes
hct-116	human colon cells
hela	Human cervix cells
hela s3.	Human cervical carcinoma
hep3b	Human liver
hepa1	Hepatocytes
hepg2	Human liver
Hnmec-1	Human microvascular endothelial cells
Hs-578t	Human breast cell carcinoma epithelial cells
ht1080	Human fibrosarcoma cells
Ht-29	Human adenocarcinoma colon epithelial cells
huvec	Human umbilical vein endothelial cells
jr1	Human rhabdomyosarcoma cell line
jurkat	Human lymphocyte cells
LCLs	Human lymphoblastoid cell lines
LS-180	Human Caucasian colon adenocarcinoma

Hep-2 m4	Human epithelial type-2 cells
mc3t3-e1	Mouse osteoblastic cell line
mcf-10	Human non tumourigenic epithelial cell line
mcf-7	Human mammary cells
MDA-MB-231	Human adenocarcinoma breast cells.
Nih 3t3	Embryonic mouse fibroblast
Pc-12	Rat Adrenal phoechromocytom
Pc-3	Human prostatic adenocarcinoma.
pnt1a	Normal prostate epithelium immortalized with SV40
RASMCs	Rat aortic smooth muscle cells
rko	Colon carcinoma cell line
saos-2	Human bone cells (osteosarcoma)
tk6	Human spleen hereditary spheroc
u2os	Human bone cells
u-937	Human monocytic (pre-macrophage line).
Vm-10	Murine P53 temperature sensitive cells.
WEHI-213	Mouse B-cell lymphoma
wtk	Human P53-deficeint lymphoblastoma cell
RAW 264.7	Mouse macrophage.
HUVECS	Human umbilical vein endothelial cells
NIH/3T3	Mouse embryo fibroblast
C127	Mouse mammary tumour cell line.
BPLER	Human tumorigenic breast cell lines
JR8	Human cutaneous melanoma
JR1	Human rhabdomyosarcoma cell line.
RASMCs	Rat aortic smooth muscle cells.

Table S2.2: Gene name abbreviations.

ABREV.	GENE NAME
AHR	ARYL HYDROCARBON RECEPTOR
APAF1	APOPTOTIC PROTEASE ACTIVATING FACTOR 1
ARNT	ARYL HYDROCARBON RECEPTOR NUCLEAR TRANSLOCATOR
BACH1	BTB DOMAIN AND CNC HOMOLOG 1
BAX	BCL-2-ASSOCIATED X PROTEIN
BCL2	B-CELL LYMPHOMA 2
BID	BH3 INTERACTING-DOMAIN DEATH AGONIST
BNIP3	BCL2 INTERACTING PROTEIN 3
BRCA1	BREAST CANCER 1
BRCA2	BREAST CANCER 2
CCND1	CYCLIN D1
CDKN1A	CYCLIN DEPENDENT KINASE INHIBITOR 1A
COX2	CYCLOOXYGENASE
CTNNB1	CATENIN BETA 1
CXCL12	C-X-C MOTIF CHEMOKINE LIGAND 12
CXCL8	C-X-C MOTIF CHEMOKINE LIGAND 8
CYP1A1	CYTOCHROME P450 FAMILY 1 SUBFAMILY A MEMBER 1
CYP1A2	CYTOCHROME P450 FAMILY 1 SUBFAMILY A MEMBER 2
CYP1B1	CYTOCHROME P450 FAMILY 1 SUBFAMILY B MEMBER 1
CYP2E1	CYTOCHROME P450 FAMILY 2 SUBFAMILY E MEMBER 1
CYP7B	25-HYDROXYCHOLESTEROL 7-ALPHA-HYDROXYLASE
ELAVL1	ELAV LIKE RNA BINDING PROTEIN 1
EPO	ERYTHROPOIETIN
ERA	OESTROGEN RECEPTOR ALPHA
ERB	OESTROGEN RECEPTOR BETA
FTH1	FERRITIN 1
G6PD	GLUTATHIONE PEROXIDASE 6
GADD45A	GROWTH ARREST AND DNA DAMAGE-INDUCIBLE PROTEIN GADD45 ALPHA
GCLC	GLUTAMATE-CYSTEINE LIGASE CATALYTIC SUBUNIT
GSTP	GLUTATHIONE-S-TRANSFERASE PI
HIF1	HYPOXIA-INDUCIBLE FACTOR 1
HMOX1	HEME-OXYGENASE 1
HSF1	HEAT-SHOCK FACTOR 1
HSP70	HEAT-SHOCK PROTIEN 70
HSP72	HEAT SHOCK 70 KDA PROTEIN 1

HSP90	HEATSHOCK PROTEIN 90
IKBA	NUCLEAR FACTOR OF KAPPA LIGHT POLYPEPTIDE GENE ENHANCER IN B-CELLS INHIBITOR, ALPHA
IKKB	INHIBITOR OF NUCLEAR FACTOR KAPPA-B KINASE SUBUNIT BETA
IL-12	INTRALUEKIN 12
IL-4	INTRALUEKIN 4
IL1B	INTRALEUKIN 1B
IL6	INTRALEUKIN 6
IL8	INTRALEUKIN 8
KEAP1	KELCH-LIKE ECH DOMAIN CONTAINING PROTEIN 1
MT1	METALLOTHIONEIN 1
MT2	METALLOTHIONEIN 2
MTF1	METAL TRANSCRIPTION FACTOR 1
MYC	MYC PROTO-ONCOGENE PROTEIN
NFKB	NUCLEAR FACTOR KAPPA B SUBUNIT 1
NIX	BCL2 INTERACTING PROTEIN 3 LIKE
NOS2	NITRIC OXIDE SYNTHASE 2A
NQO1	NAD(P)H QUINONE DEHYDROGENASE 1
NRF2	NUCLEAR FACTOR (ERYTHRIOD LIKE) 2
P53	TUMOUR PROTEIN P53
PGD	6-PHOSPHOGLUCONATE DEHYDROGENASE
PIGF	PLACENTAL GROWTH FACTOR 1
PRDX1	PEROXIREDOXIN 1
PUMA	P53 UPREGULATED MODULATOR OF APOPTOSIS
RELA	RELA PROTO-ONCOGENE, NF-KB SUBUNIT
RELB	RELB PROTO-ONCOGENE, NF-KB SUBUNIT
SOD1	SUPEROXIDE DISMUTASE
SQSTM1	SEQUESTOSOME 1
STAT6	SIGNAL TRANSDUCER AND ACTIVATOR OF TRANSCRIPTION 6
TRAF3	TNF RECEPTOR ASSOCIATED FACTOR 3
TXNRD1	THIOREDOXIN REDUCTASE 1
TXNRD2	THIOREDOXIN REDUCTASE 2
VEGF1	VASCULAR ENDOTHELIAL GROWTH FACTOR
XPC	XERODERMA PIGMENTOSUM, COMPLEMENTATION GROUP C
ZIP8	ZINC TRANSPORTER ZIP8

Table S2.3: Chemical name abbreviations.

ABREV.	CHEMICAL NAME
SFN	Sulphurophane.
TCDD	2,3,7,8 – Tetrachlorodibenzodioxin
CEL	Celestrol
DEM	Diethylmaleate
E2	Ethinylestradiol
min	Minadione
1-NP	1-Nonylphenol
TAM	Tamoxifen
15D-PGJ2	15-Deoxy-Delta-12,13-prostagladin J2.
TOC.	Tocopherol

Table S2.4: Validated mammalian interactions between TFs and DNA sequences. DNA-binding interactions shown as direct. Protein-Protein interactions as stated. Indirect interactions are shown where the type of interaction is not known. Positive interactions (+), inhibitory interactions (-) and protein-protein interactions that have no directionality (~) are shown. Cell lines and experimental method is as indicated.

REGULATOR	Type		TARGET	INDUCER	SPECIES	ASSAY	TISSUE/ CELL LINE	REF.
NRF2	DIRECT	+	GSTP	SFN	human	ChIP-Seq	LCLS, beas-2b, A549	[10]
NRF2	DIRECT	+	GCLC	SFN	human	ChIP-Seq	LCLS, beas-2b, A549	[10]
NRF2	DIRECT	+	HMOX1	SFN	human	ChIP-Seq	LCLS, beas-2b, A549	[10]
NRF2	DIRECT	+	FTH1	SFN	human	ChIP-Seq	LCLS, beas-2b, A549	[10]
NRF2	DIRECT	+	PRDX1	SFN	human	ChIP-Seq	LCLS, beas-2b, A549	[10]
NRF2	DIRECT	+	SQSTM1	SFN	human	ChIP-Seq, EMSA	HELA, HEK293, P62 -/- MEFS	[33]
NRF2	DIRECT	+	NQO1	SFN	human	ChIP-Seq	LCLS, beas-2b, A549	[10]
NRF2	DIRECT	+	SOD1	TCDD	human	Luc-assay	HEPG2	[64]
NRF2	DIRECT	+	HSP70	CEL.	mouse	Western-Blot	MEFS/HEK293	[65]
NRF2	DIRECT	+	TXNRD1	NRF2 KO.	mouse	ChIP-Seq	MEFS	[10]
NRF2	DIRECT	+	IL8	MG-132	human	Luc-plasmid	MESANGIAL CELLS, HELA, HEK293	[66]
NRF2	DIRECT	+	BACH1	SIRNA (NRF2), TBHQ,	human	Luc-assay, EMSA, ChIP-PCR	HUVECs	[67]

144

				SFN, OA-NO$_2$				
NRF2	DIRECT	-	IL6	KEAP1 KD, DEM	mouse	LUC-ASSAY, ChIP-qPCR	INVIVO, RAW264.7	[68]
NRF2	DIRECT	-	IL1B	DEM	mouse	LUC-ASSAY, ChIP-qPCR	RAW264.7	[67]
NRF2	DIRECT	+	MT1	SFN	human	ChIP-Seq	LCLS, beas-2b, A549	[10]
NRF2	DIRECT	+	MT2	NRF2 KO	human	ChIP-Seq	LCLS, beas-2b, A549	[10]
NRF2	DIRECT	+	AHR	AHR siRNA, CDDO-IM	mouse	LUC-assay, ChIP-PCR,	MEFS	[47]
NRF2	DIRECT	+	HIF1	Min., sulphur ophane	human	ChIP-seq, Nrf2 KO.	HepG2, MCF7, Lymphoblastiod cells	[69]
AHR	DIRECT	+	NRF2	TCDD	mouse	Western blot, RT-qPCR, Luc-reporter	Hepa-1C1C7	[45]
AHR	DIRECT	+	CYP1B1	TCDD	human	ChIP-Seq, ChIP-qPCR, LUC-assay.	MCF-7	[70]
AHR	DIRECT	+	CYP1A1	TCDD	human	Immunoblot, Nothern blot	MCF-7, JEG-3, A431, LS180	[71]
AHR	DIRECT	+	CYP1A2	TCDD, MC	human	Immunoblot, northern blot.	MCF-7, JEG-3, A431, LS180	[71]
ARNT	O	+	AHR	TCDD	human	ChIP-seq, ChIP-PCR, Luc-assay.	MCF-7	[72]
ARNT	O	+	HIF1	Hypoxia	human	Gel-shift assay,	HELA	[73]
ARNT	O	+	ERB	E2	human	ChIP-PCR	COS-7, HELA	[74]
ARNT	O	+	ERA	E2	human	ChIP-PCR	COS-7, HELA	[74]

ERA	DIRECT	+	HIF1	hypoxia, E2, TAM.	human	ChIP-PCR, Luc-assay	MCF-7, MDA-MB-231	[75]
ERB	INDIRECT	-	HIF1	hypoxia, E2, TCDD, ERB siRNA	human, rat	ChIP-PCR, Western blot, Luc-assay	HEP3B, HEK293	[76]
ERB	INDIRECT	-	ARNT	hypoxia, E2, TCDD, ERB siRNA	human, rat	ChIP-assay, Western blot, Luc-assay	HEP3B, HEK293	[76]
RELA	DIRECT	+	NQO1	hypoxia, mitomycin C	human	EMSA, oligonucleotide labelling	HT29/HEPG2	[77]
RELA	DIRECT	+	GSTP	H_2O_2	human	Cat-reporter	HEPG2, MCF7, HELA	[11]
RELA	DIRECT	+	HMOX1	TGF-1B	human	Luc-assay, EMSA	A549	[76]
RELA	DIRECT	+	FTH1	TNF	mouse	EMSA, Mutation contructs	NIH3T3	[78]
RELA	DIRECT	+	SOD1	H_2O_2	human	Luc-assay, EMSA	PC12,	[79]
RELA	DIRECT	+	HSP90	PMA, TNF-A	human	Luc-assay, EMSA, ChIP-PCR	HEK293	[80]
RELA	DIRECT	+	IL8	IG/HIV, NFKB KD.	human	ELISA, Mobility shift assay.	u-937	[81]
RELA	DIRECT	+	IL6	NFKB KD, LPS, TNF-A, PMA.	human	EMSA	u-937, HELA	[82]
RELA	DIRECT	+	IL1B	PMA, POLY(RI-RC), Sandrai virus.	human	Luc-assay	U-937, plb985, JURKAT,	[83]
RELA	INDIRECT	-	NRF2		human	Luc-assay EMSA, RT-qPCR	HEK293	[84]
AHR	O	+	RELB	TCDD, siRNA	human	EMSA, western blot	MDA-MB 436, MCF-7	[85]

AHR-RELB	DIRECT	+	IL8	TCDD, siRNA	human	EMSA, Western blot	MDA-MB 436, MCF-7	[85]
AHR-RELA	DIRECT	+	IL6	TCDD, siRNA NFKB.	human	EMSA, Western blot, CHIP-assay, Luc-reporter	127 AD, beas-2b,	[86]
AHR-RELA	DIRECT	+	C-MYC	Transfection.	human	EMSA	HS578T, MCF-10F	[87]
MTF1	DIRECT	+	MT1	CDSO4	mouse	Luc-reporter	Mouse egg	[35]
MTF1	DIRECT	+	MT2	CDSO4	mouse	Luc-reporter	Mouse egg	[35]
MTF1	DIRECT	+	TXNRD2	MTF1 KO, ZN++	mouse	qPCR + Immunoprecipitation.	MEFS	[88]
AHR-RELA	DIRECT	-	CXCL8	1-NP, 1-AP, siRNA	human	Luc-assay, ChIP-PCR	beas-2b,	[89]
AHR	DIRECT	O	RELA	1-NP, 1-AP, siRNA	human	Luc-reporter, ChIP-PCR	beas-2b,	[89]
HIF1	DIRECT	+	MT1	Low O_2, $ZNCL_2$	mouse	Luc-reporter, ChIP-PCR	L CELLS	[36]
AHR	DIRECT	+	SOD1	TCDD	human	Luc-reporter, Mobility shift.	HEPG2	[64]
HIF1	DIRECT	+	IKKB	ERB-KO, Hypoxia	human	Luc-reporter	PNT1A	[90]
HIF1	DIRECT	+	VEGF1	Hypoxia	mouse	Luc-reporter, EMSA	HEP3B, HEPA1	[91]
HIF1	DIRECT	+	HSP70	$CoCl_2$	human/ mouse	Luc-Reporter	HepG2/MCF-7	[92]
HIF1	DIRECT	+	PIGF	EPO	human	Luc-reporter, ChIP-PCR	K562	[93]
HSF1	DIRECT	+	SQSTM1	siRNA transfection (HSF1)	human	Luc-reporter, Western blot	RKO, A549, MCF-7	[94]
HSF1	DIRECT	+	HSP90	Heat-shock.	human	ChIP-seq, RNA-seq.	MCF-7, HMLER, MCF10A	[95]

HSF1	O	+	MTF1	ZNSO$_4$	human	EMSA, Immunoblot	HELA	[14]
HMOX1	INDIRECT	+	HIF1	15D-PGJ2	human	siRNA (HMOX1), Western blot.	MCF-7	[96]
SQSTM1	PROTEO	-	KEAP1	NA	human	Luc-reporter, EMSA, RT-qPCR	HEK293	[33]
KEAP1	O	-	NRF2	NA.	human	Luc-reporter, siRNA transfection.	HEK293	[84]
VEGF	INDIRECT	+	NRF2	VEGF	mouse	Western blot.	BMEC	[40]
ERB	INDIRECT	+	NRF2	E$_2$	human	Western blot.	MCF-7, T47D	[97]
ERA	INDIRECT	-	NQO1	E$_2$, tBHQ	human	Western blot, Luc-reporter	MCF-7	[98]
NFKB	DIRECT	+	COX2	?		Luc-reporter, EMSA	AGS	[99]
ERB	O	+	NRF2	TNFA	mouse	Western blot, Luc-assay, EMSA	Mc3t3-e1	[97]
ERB	O	+	NRF2	TAM.	human	Gel-Shift assay	MCF-7	[58]
HIF1	DIRECT	+	COX2	E2	human	Western-blot, Luc-reporter	COS-1	[100]
CDKN1	O	+	KEAP1	DMOG	human	CHIP-seq	HEK293	[100]
NQO1	O	+	P53	KO P21 (siRNA), H$_2$O$_2$	mouse	Luc-reporter, RT-qPCR, Immunoblot	MEFS	[58]
HSF1	DIRECT	+	SQSTM1	H$_2$O$_2$	human mouse	Immunoblot	HCT116, 7-M12	[101]
HSF1	DIRECT	+	HSP70	siRNA HSF1, Hsp90 inhibition	human	GFP-reporter, Western blot, RT-qPCR	RKO, A549, MCF-7	[102]
ERb	INDIRECT	+	NFKB	TNF-a	rat	ChIP-seq,	RASMCs	[103]

P53	DIRECT	+	CDKN1	Expression vector.	human	Western blot.	HEK293,	[104]
P53	DIRECT	+	GADD45A	Short hairpin RNA KD	human	ChIP-seq, RNA-seq	MCF-7	[104]
P53	INDIRECT	-	VEGF	Short hairpin RNA KD	human	ChIP-seq, RNA-seq	MCF-7	[105]
P53	INDIRECT	+	BRCA1	DNA damage	human, mouse	Western blot, Northern blot	hNMEC	[106]
P53	DIRECT	+	XPC		human	Luc-assay, Western blot	MCF-7, MDA-MB435	[107]
P53	INDIRECT	-	HIF1	Quinacrane	mouse	RT-qPCR, Western blot, P53 KO	Cardiomyocytes	[108]
P53	DIRECT	+	HMOX1	COCL$_2$, P53 KO	mouse	Western blot	MCF-7, U-20S	[109]
P53	DIRECT	+	APAF1	Y-Irradiation	mouse, human	EMSA, Western blot, ChIP-PCR	MCF-7, U-20S.	[110]
P53	DIRECT	+	BAX	Ionizing radiation, DOXORUBICIN	human	Microarray, EMSA, Western blot.	TK6, WTK	[108]
P53	DIRECT	-	BCL-2	KO.	mouse	ChIP-assay, Western blot.	MEFS	[109]
NFKB	DIRECT	+	BCL-2	P53-vector.	human	Western-blot, Luc-assay.	DHL-4, K562	[111]
NRF2	DIRECT	+	PGD		human	Luc-assay, EMSA, DNASE footprinting	HELA S3, WEHI	[112]
HSF1	DIRECT	-	BAX		human	EMSA, ChIP-seq		[95]
ELAVL1	O	+	HSF1	NA	human	ChIP-seq	MCF7	[113]
ELAVL1	O	+	HIF1	Short hairpin RNA KD (ELAVL1)	human	RT-qPCR, KD OF HSF1	MCF7	[98]

HSF1	*DIRECT*	+	*CXCL12*	Short hairpin RNA KD	human	Western blot.	MDA-MB-231	[114]
HSF1	DIRECT	-	TRAF3	Retro-virus transfection.	mouse	Immonublot	MCF10A	[114]
HSF1	DIRECT	-	IL	NA	human	ChIP-seq	MCF7, BPLER	[115]
HSF1	INDIRECT	+	ELAVL1	NA	human	ChIP-seq	MCF7/BPLER	[115]
NFKB (IKK)	DIRECT	+	HIF1	Short hairpin RNA KD	human	KD, RT-qPCR	MDA-MB-231	[95]
HIF1	DIRECT	+	BNIP3	siRNA interference	human	ChIP-PCR, Luc-assay	HEK293, U2OS	[95]
HIF1	DIRECT	+	COX-2	Hypoxia	human	Luc-assay, EMSA	HEK293, MCF-7	[115]
BCL2	O	+	HIF1	Hypoxia	human	ChIP-PCR, Luc-assay	HT29	[113]
ERA	DIRECT	+	IKBA	Hypoxia	human	Western-blot, BCL-2 overexpression, Luc-reporter	JR8, JR1, PLF2, ASM-SC	[116]
RELA	DIRECT	+	IKBA	E2	rat	ChIP ERB	RASMCS	[117]
HIF1	DIRECT	+	IKKB	E2, DPN, THC	rat	ChIP P65	RASMCS	[118]
HIF1	INDIRECT	+	P53	KD-HIF1, hypoxia	human	Luc-reporter	PNT1A, PC3-M	[103]
P53	INDIRECT	-	C-MYC		human, mouse	Luc-assay, Immunoprecipitation.	H1299, MEFS	[103]
NFKB	DIRECT	+	TRAF3	DOXO	human	Luc-assay, ChIP-assay	HCT-116, MCF-7	[90]
P53	DIRECT	+	PUMA	Doxyclin	human	Northern-blot	SAOS-2, H1299	[119]
P53	DIRECT	+	BAX	Actin Mycin-D	human	Northern-blot, EMSA, flag-tag.	SAOS-2, U2OS, H1299	[120]
P53	DIRECT	+	BID	P53 transfection.	human	EMSA, CAT-reporter	H358	[121]

ERA	INDIRECT	-	NRF2	Y-irradiation	human/mouse	EMSA, Western blot, in-situ hybridization Luc-reporter, ChIP-assay	VM10, MEFS, M3	[59]
ERRB	INDIRECT	-	NRF2	E2	human	Luc-reporter, immunoprecipitation.	COS-1	[58]
IL4	INDIRECT	+	STAT6	TNF-a	human	Gel-shift assay	HepG2	[122]
COX2	INDIRECT	-	NRF2	15D-PGJ2	mouse	RNA-blot, Immunohistochemistry	MACROPHAGE HEPATOCYTES	[123]
HSF1	DIRECT	+	HSP72	17-B estradiol	rat	Western blot, EMSA, TF-decoy	MYOCYTES	[124]
NFKB	DIRECT	+	HSP72	17-B estradiol	rat	Western blot, EMSA, TF-decoy	MYOCYTES	[124]
NFKB	DIRECT	-	CYP1A1	UVB	human	Luc-reporter	HepG2	[125]
NFKB	INDIRECT	+	IL-4	TNF-A	Mouse	EMSA, Immunoblot	32D	[126]
NFKB	DIRECT	+	IL-12	TNF-A	human	Luc-reporter, ELISA	HEK293T	[126]
NFKB	DIRECT	+	ZIP8	TNF-A	human	Luc-reporter, ChIP-PCR	A549	[127]
P53	DIRECT	+	NIX	Hypoxia	human/mouse	Luc-reporter.	U2OS/MEFS	[128]
HIF1	INDIRECT	+	NIX	Cancer	human	Western-blot.	HT1080	[129]
NFKB	DIRECT	+	MYC	IL1	human	Luc-reporter	FS-4	[128]
HIF1	INDIRECT	-	MYC	Hypoxia	NA	NA	NA	[130]
P53	INDIRECT	+	P53		NA	NA	NA	[104]
NFKB	DIRECT	+	HSP70	TNF-α	human	Luc-assay	HEK-293	[80]
NQO1	INDIRECT	+	HIF1	Hypoxia	human	siRNA against NQO1	MDA-MB-231	[131]
HSP70	O	-	HSF1	Heat-shock	human	EMSA, Western blot	HELA	[132]

HSP90	O	-	HSF1	Heat-shock	human	EMSA, Western blot	HELA	[132]
HSP70	O	+	BCL2	H_2O_2	rat	Western blot.	BRL	[133]
HMOX1	O	-	NFkB	TOC.	human	EMSA, western blot, RT-qPCR	PC3.	[134]

Figure S2.1: Network Logic Rules for BoolNet models. Boolean modeling rules for adaptive-stress response network. Inhibitory results are set to AND/NOT as the presence of the factor will prevent gene expression (gene will not be expressed if inhibitor is present). Logic rules were selected based on the literature e.g. AND shows that binding is dependent on two factors e.g. MT1 needs MTF1 and HIF1.

```
targets, factors
NFKB, ( MT1 | MT2 | TRAF3 | SQSTM1) & !(ZIP8 | HMOX1 | AHR)
P53, P53
NRF2, (VEGF1 | AHR) & !(COX2 | NFKB)
AHR, NRF2 | !(HSP90 | NFKB)
HIF1, (HMOX1 | NFKB | NQO1) & !P53
HSF1, NQO1 & !(HSP70 | HSP90)
MTF1, ZIP8
TRAF3, NFKB & !HSF1
SOD1, NRF2 | NFKB
GSTP, NRF2 | NFKB
GCLC, NRF2 | NFKB
PRDX1, NRF2
HMOX1, NRF2 | P53 | NFKB
FTH1, (NRF2 | NFKB)
SQSTM1, NRF2
NQO1, NRF2 | NFKB
TXNRD2, NRF2 | MTF1
BACH1, NRF2
CYP1B1, AHR
CYP1A1, AHR | NFKB
CYP1A2, AHR
HSP70, NRF2 | HSF1
HSP90, NRF2 | HSF1 | NFKB
HSP72, HSF1 | NFKB
MYC, (P53 | NFKB) & !HIF1
BID, P53 & !HIF1
BIM, P53
PUMA, P53
APAF1, P53
BAX, P53 & !HSF1
NIX, (P53 | NFKB) & !HIF1
IL6, NFKB & !NRF2
STAT6, ( IL4 | NFKB) & !HSF1
IL1, NFKB & !NRF2
IL8, NRF2 | NFKB
IL4, NFKB
IL12, NFKB
COX2, NFKB
BCL_2, (P53 | HSP70 | NFKB) & !HSF1
GADD45A, P53
```

BRCA1, P53
XPC, P53
CDKN1A, P53 | NFKB
NOS_2, NFKB
CYCLIN_D, NFKB
IGFBP1, NFKB
VEGF1, P53 | HIF1 | NFKB
EPO, HIF1
MT1, NRF2 | IL6 | MTF1| (MTF1 & HIF1)
MT2, NRF2 | IL6 | MTF1 | HIF1
ZIP8, NFKB
G6PD, NRF2
PIGF, NRF2 | HIF1 | MTF1
PGD, NRF2

Figure S2.2: Network Logic Rules for BoolNet models with oestrogen receptor links. Boolen logic rules for oestrogenic models. Inhibitory results are set to AND/NOT as the presence of the factor will prevent gene expression (gene will not be expressed if inhibitor is present). Rules were select in and when the literature shows that binding is dependent on two factors e.g. MT1 needs MTF1 and HIF-1α.

```
NFKB, ( MT1 | MT2 | TRAF6 | TRAF3 | SQSTM1 | ERA) & !(ZIP8 | HMOX1 | AHR)
NRF2, (VEGF1 | AHR) & !(COX2 | NFKB | ERA )
SOD1, NRF2 | NFKB
MYC, (P53 | NFKB) & !HIF1
BID, P53 & !HIF1
BIM, P53
PUMA, P53
APAF1, P53
BAX, P53 & !HSF1
P53, P53
HSP70, NRF2 | HSF1 | NFKB
GSTP, NRF2 | NFKB
HSF1, NQO1 & !(HSP70 | HSP90)
COX2, NFKB
BCL_2, (P53 | HSP70 | NFKB) & !HSF1
GADD45A, P53
TXNRD2, NRF2 | NFKB
VEGF1, P53 | HIF1 | NFKB
CDKN1A, P53 | NFKB
BACH1, NRF2
IL6, NFKB & !NRF2
G6PD, NRF2
PIGF, NRF2 | HIF1
PGD, NRF2
NOS_2, NFKB
IGFBP1, NFKB
CYCLIN_D, NFKB
IL1B, NFKB & !NRF2
MTF1, ZIP8
BRCA1, P53
XPC, P53
MT1, NRF2 | IL6 | MTF1| (MTF1 & HIF1)
MT2, NRF2 | IL6 | MTF1 | HIF1
AHR, NRF2 | !(HSP90 | NFKB)
HIF1, (HMOX1 | NFKB | NQO1| ERA) & !(P53 | ERB)
EPO, HIF1
HSP70, NRF2 | HSF1
HSP90, NRF2 | HSF1 | NFKB
HSP72, HSF1 | NFKB
NIX, P53 | HIF1
ZIP8, NFKB
IL4, NFKB
IL5, NFKB
TRAF6, TRAF6
IL8, NRF2 | NFKB
GCLC, NRF2 | NFKB
CYP1B1, AHR
CYP1A1, AHR
CYP1A2, AHR
HMOX1, NRF2 | P53 | NFKB
```

FTH1, (NRF2 | NFKB)
TRAF3, NFKB & !HSF1
PRDX1, NRF2
CYP2A1, AHR | NFKB
CYP211, AHR | NFKB
CYP7B, AHR | NFKB
SQSTM1, NRF2
NQO1, NRF2 | NFKB
STAT6, (IL4 | NFKB) & !HSF1
ERA, ERA
ERB, ERB

2.8 References:

1. Pique-Regi R, Degner JF, Pai AA, Gaffney DJ, Gilad Y, Pritchard JK. 2011 .Accurate inference of transcription factor binding from DNA sequence and chromatin accessibility data. *Genome Res.* 21, 447–55. (doi:10.1101/gr.112623.110)

2. Huang Q, Zhang J, Martin FL, Peng S, Tian M, Mu X, Shen H. 2013. Perfluorooctanoic acid induces apoptosis through the p53-dependent mitochondrial pathway in human hepatic cells: A proteomic study. *Toxicol. Lett.* 223, 211–220. (doi:10.1016/J.TOXLET.2013.09.002)

3. Moorthy B, Chu C, Carlin DJ. 2015 .Polycyclic aromatic hydrocarbons: from metabolism to lung cancer. *Toxicol. Sci.* 145, 5–15. (doi:10.1093/toxsci/kfv040)

4. Valko M, Morris H, Cronin MTD. 2005 .Metals, toxicity and oxidative stress. *Curr. Med. Chem.* 12, 1161–208.

5. Leung MCK, Procter AC, Goldstone J V, Foox J, Desalle R, Mattingly CJ, Siddall ME, Timme-laragy AR. 2017 .Applying evolutionary genetics to developmental toxicology and risk assessment. *Reprod. Toxicol.* 69, 174–186. (doi:10.1016/j.reprotox.2017.03.003)

6. Nakajima H et al. 2011 .Tissue-restricted expression of Nrf2 and its target genes in zebrafish with gene-specific variations in the induction profiles. *PLoS One* 6, e26884. (doi:10.1371/journal.pone.0026884)

7. Ritter DI, Li Q, Kostka D, Pollard KS, Guo S, Chuang JH. 2010 .The importance of Being Cis: Evolution of Orthologous Fish and Mammalian enhancer activity. *Mol. Biol. Evol.* 27, 2322–2332. (doi:10.1093/molbev/msq128)

8. Nardone J, Lee DU, Ansel KM, Rao A. 2004 .Bioinformatics for the 'bench biologist': how to find regulatory regions in genomic DNA. *Nat. Immunol.* 5, 768–74. (doi:10.1038/ni0804-768)

9. Gomez-Pastor R, Burchfiel ET, Thiele DJ. 2017 .Regulation of heat shock transcription factors and their roles in physiology and disease. *Nat. Rev. Mol. Cell Biol.* (doi:10.1038/nrm.2017.73)

10. Chorley BN, Campbell MR, Wang X, Karaca M, Sambandan D, Bangura F, Xue P, Pi J, Kleeberger SR, Bell DA. 2012 .Identification of novel NRF2-regulated genes by ChIP-Seq: influence on retinoid X

receptor alpha. *Nucleic Acids Res.* 40, 7416–29. (doi:10.1093/nar/gks409)

11. Xia C, Hu J, Ketterer B, Taylor JB. 1996 .The organization of the human GSTP1-1 gene promoter and its response to retinoic acid and cellular redox status. *Biochem. J.* 313, Pt 1, 155–61.

12. Wang R-L, Biales AD, Garcia-Reyero N, Perkins EJ, Villeneuve DL, Ankley GT, Bencic DC. 2016 .Fish connectivity mapping: linking chemical stressors by their mechanisms of action-driven transcriptomic profiles. *BMC Genomics* 17, 84. (doi:10.1186/s12864-016-2406-y)

13. Wittwehr C, Aladjov H, Ankley G, Byrne HJ, Knecht JD, Heinzle E, LAndesmann B, Luijten M, MAckay C, Maxwell G, Meek MEB, Paini A, Perkins E, Sobanski T, Vileneuve D, Waters KM, Whelan M. 2017 .How Adverse Outcome Pathways Can Aid the Development and Use of Computational Prediction Models for Regulatory Toxicology. *Toxicol.l Sci.* 155, 326–336. (doi:10.1093/toxsci/kfw207)

14. Simmons SO, Fan CY, Ramabhadran R. 2009 .Cellular stress response pathway system as a sentinel ensemble in toxicological screening. *Toxicol. Sci.* 111, 202–225. (doi:10.1093/toxsci/kfp140)

15. Bolton JL, Trush MA, Penning TM, Dryhurst G, Monks TJ. 2000 .Role of Quinones in Toxicology *Chem. Res. Toxicol.* 13, 135–160. (doi:10.1021/tx9902082)

16. Nerland DE. 2007 .The antioxidant/electrophile response element motif. *Drug Metab. Rev.* 39, 235–48. (doi:10.1080/03602530601125000)

17. Wardyn JD, Ponsford AH, Sanderson CM. 2015 .Dissecting molecular cross-talk between Nrf2 and NF-κB response pathways. *Biochem. Soc. Trans.* 43, 621–6. (doi:10.1042/BST20150014)

18. Perkins EJ, Chipman JK, Edwards S, Habib T, Falciani F, Taylor R, Van Aggelen G, Vulpe C, Antczak P, Loguinov A. 2011 .Reverse engineering adverse outcome pathways. *Environ. Toxicol. Chem.* 30, 22–38. (doi:10.1002/etc.374)

19. Miller DH, Jensen KM, Villeneuve DL, Kahl MD, Makynen EA, Durhan EJ, Ankley GT. 2007 .Linkage of biochemical responses to population-level effects: A case study with vitellogenin in the fathead minnow (Pimephales promelas). *Environ. Toxicol. Chem.* 26, 521.

(doi:10.1897/06-318R.1)

20. Gillies K, Krone SM, Nagler JJ, Schultz IR. 2016 .A Computational Model of the Rainbow Trout Hypothalamus-Pituitary-Ovary-Liver Axis. *PLoS Comput. Biol.* 12, 1–27. (doi:10.1371/journal.pcbi.1004874)

21. Khan FM, Marquardt S, Gupta SK, Knoll S, Schmitz U, Spitschak A, Engelmann D, Vera J, Wolkenhauer O, Putzer B. 2017 .Unraveling a tumor type-specific regulatory core underlying E2F1-mediated epithelial-mesenchymal transition to predict receptor protein signatures. *Nat. Commun.* 8, 198. (doi:10.1038/s41467-017-00268-2)

22. Martik ML, Lyons DC, Mcclay DR. 2016 .Developmental gene regulatory networks in sea urchins and what we can learn from them. *F1000 Fac. Rev.* 5, 1–8. (doi:10.12688/f1000research.7381.1)

23. Ansell PJ, Espinosa-Nicholas C, Curran EM, Judy BM, Philips BJ, Hannink M, Lubahn DB. 2004 .In Vitro and in Vivo Regulation of Antioxidant Response Element-Dependent Gene Expression by Estrogens. *Endocrinology* 145, 311–317. (doi:10.1210/en.2003-0817)

24. Delile J, Herrmann M, Peyriéras N, Doursat R. 2017 .A cell-based computational model of early embryogenesis coupling mechanical behaviour and gene regulation. *Nat. Commun.* 8, 13929. (doi:10.1038/ncomms13929)

25. Sridharan S, Layek R, Datta A, Venkatraj J. 2012 .Boolean modeling and fault diagnosis in oxidative stress response. *BMC Genomics* 13, S4. (doi:10.1186/1471-2164-13-S6-S4)

26. Müssel C, Hopfensitz M, Kestler HA. 2010 .BoolNet—an R package for generation, reconstruction and analysis of Boolean networks. *Bioinformatics* 26, 1378–1380. (doi:10.1093/bioinformatics/btq124)

27. Howe DG, Bradford YM, Eagle A, Fashena D, Frazer K, Kalita P, Mani P, Martin R, Moxon ST, Paddock H, Pich C, Ramachandran S, Ruzicka L, Schaper K, Shao X, Singer A, Toro S, Wersterfield M. 2017 .The Zebrafish Model Organism Database: new support for human disease models, mutation details, gene expression phenotypes and searching. *Nucleic Acids Res.* 45, 758–768. (doi:10.1093/nar/gkw1116)

28. Ducharme NA, Reif DM, Gustafsson J-A, Bondesson M. 2015 .Comparison of toxicity values across zebrafish early life stages and

mammalian studies: Implications for chemical testing. *Reprod. Toxicol.* 55, 3–10. (doi:10.1016/J.REPROTOX.2014.09.005)

29. Siepel A, Bejerano G, Penderson JS, Hinrichs AS, Hou M, Rosenbloom K, Clawson H, Spieth J, Hillier LW, Richards S, Weinstock GM, Wilson RK, Gibbs RA< Kent WJ, Miller W, Haussler D. 2005 .Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* 15, 1034–50. (doi:10.1101/gr.3715005)

30. Paquette SM, Leinonen K, Longabaugh WJR. 2016 .BioTapestry now provides a web application and improved drawing and layout tools. *F1000Research* 5, 39. (doi:10.12688/f1000research.7620.1)

31. Spivakov M. 2014 .Spurious transcription factor binding: Non-functional or genetically redundant? *Bioessays* 36, 798–806. (doi:10.1002/bies.201400036)

32. Wooten MW, Geetha T, Seibenhener ML, Babu JR, Diaz-Meco MT, Moscat J. 2005 .The p62 scaffold regulates nerve growth factor-induced NF-κB activation by influencing TRAF6 polyubiquitination. *J. Biol. Chem.* 280, 35625–35629. (doi:10.1074/jbc.C500237200)

33. Jain A, Lamark T, Sjøttem E, Larsen KB, Awuh JA, Øvervatn A, McMahon M, Hayes JD, Johansen T. 2010. p62/SQSTM1 is a target gene for transcription factor NRF2 and creates a positive feedback loop by inducing antioxidant response element-driven gene transcription. *J. Biol. Chem.* 285, 22576–22591. (doi:10.1074/jbc.M110.118976)

34. Ohtsuji M, Katsuoka F, Kobayashi A, Aburatani H, Hayes JD, Yamamoto M. 2008 .Nrf1 and Nrf2 play distinct roles in activation of antioxidant response element-dependent genes. *J. Biol. Chem.* 283, 33554–33562. (doi:10.1074/jbc.M804597200)

35. Stuart GW, Searle PF, Chent HY, Brinstert RL, Palmiter RD. 1984 .A 12-base-pair DNA motif that is repeated several times in metallothionein gene promoters confers metal regulation to a heterologous gene (cadmium/synthetic DNA/mouse eggs). *Biochemistry* 81, 7318–7322. (doi:10.1073/pnas.81.23.7318)

36. Dubé A, Harrisson J-F, Saint-Gelais G, Séguin C. 2011 .Hypoxia acts through multiple signaling pathways to induce metallothionein transactivation by the metal-responsive transcription factor-1 (MTF-1).

Biochem. Cell Biol. 89, 562–577. (doi:10.1139/o11-063)

37. Monolayer H, Author C, Schroeder JJ, Cousins RJ. 1990 .Interleukin 6 Regulates Metallothionein Gene Expression and Zinc Metabolism in in hepatocyte monolayer cultures. *Proc. Natl. Acad. Sci.* 87, 3137–3141.

38. Kim CH, Kim JH, Lee J, Ahn YS. In press. Zinc-induced NF-kB inhibition can be modulated by changes in the intracellular metallothionein level. *Toxicology and Applied Pharmacology* (doi:10.1016/S0041-008X(03)00167-4)

39. Farhang Ghahremani M, Goossens S, Nittner D, Bisteau X, Bartunkova S, Zwolinska A, Hulpia P, Haigh K, Haenebalke L, Drogat B, Jochemsen A, Roger PP, Marine JC, Haigh JJ. 2013. p53 promotes VEGF expression and angiogenesis in the absence of an intact p21-Rb pathway. *Cell Death Differ.* 20, 888–897. (doi:10.1038/cdd.2013.12)

40. Li L, Pan H, Wang H, Li X, Bu X, Wang Q, Gao Y. 2016 .Interplay between VEGF and Nrf2 regulates angiogenesis due to intracranial venous hypertension. *Sci. Rep.* , 1–11. (doi:10.1038/srep37338)

41. Vorrink SU, Domann FE. 2014 .Regulatory crosstalk and interference between the xenobiotic and hypoxia sensing pathways at the AhR-ARNT-HIF1α signaling node. *Chem. Biol. Interact.* 218, 82–88. (doi:10.1016/J.CBI.2014.05.001)

42. Sant KE, Hansen JM, Williams LM, Tran NL, Goldstone J V., Stegeman JJ, Hahn ME, Timme-Laragy A. 2017 .The role of Nrf1 and Nrf2 in the regulation of glutathione and redox dynamics in the developing zebrafish embryo. *Redox Biol.* 13, 207–218. (doi:10.1016/j.redox.2017.05.023)

43. Li J, Labbadia J, Morimoto RI. 2017. Rethinking HSF1 in Stress, Development, and Organismal Health. *Trends Cell Biol.*, xx,yy, 1–11. (doi:10.1016/j.tcb.2017.08.002)

44. Rousseau ME, Sant KE, Borden LR, Franks DG, Hahn ME, Timme-Laragy AR. 2015 .Regulation of Ahr signaling by Nrf2 during development: Effects of Nrf2a deficiency on PCB126 embryotoxicity in zebrafish (Danio rerio). *Aquat. Toxicol.* 167, 157–171. (doi:10.1016/j.aquatox.2015.08.002)

45. Miao W, Hu L, Scrivens PJ, Batist G. 2005 .Transcriptional regulation of

NF-E2 p45-related factor (NRF2) expression by the aryl hydrocarbon receptor-xenobiotic response element signaling pathway: Direct cross-talk between phase I and II drug-metabolizing enzymes. *J. Biol. Chem.* 280, 20340–20348. (doi:10.1074/jbc.M412081200)

46.	Yeager RL, Reisman S a, Aleksunes LM, Klaassen CD. 2009 .Introducing the 'TCDD-inducible AhR-Nrf2 gene battery'. *Toxicol. Sci.* 111, 238–46. (doi:10.1093/toxsci/kfp115)

47.	Shin S, Wakabayashi N, Misra V, Biswal S, Lee GH, Agoston ES, Yamamoto M, Kensler TW. 2007 .NRF2 Modulates Aryl Hydrocarbon Receptor Signaling: Influence on Adipogenesis. *Mol. Cell. Biol.* 27, 7188–7197. (doi:10.1128/MCB.00915-07)

48.	Korashy HM, El-Kadi AOS. 2004 .Differential effects of mercury, lead and copper on the constitutive and inducible expression of aryl hydrocarbon receptor (AHR)-regulated genes in cultured hepatoma Hepa 1c1c7 cells. *Toxicology* 201, 153–72. (doi:10.1016/j.tox.2004.04.011)

49.	Zhu QL, Guo SN, Yuan SS, Lv ZM, Zheng JL, Xia H. 2017 .Heat indicators of oxidative stress, inflammation and metal transport show dependence of cadmium pollution history in the liver of female zebrafish. *Aquat. Toxicol.* 191, 1–9. (doi:10.1016/j.aquatox.2017.07.010)

50.	Iwanaszko M, Brasier AR, Kimmel M. 2012 .The dependence of expression of NF-κB-dependent genes: statistics and evolutionary conservation of control sequences in the promoter and in the 3' UTR. *BMC Genomics* 13, 182. (doi:10.1186/1471-2164-13-182)

51.	Kruiswijk F, Labuschagne CF, Vousden KH. 2015. P53 in Survival, Death and Metabolic Health: a Lifeguard With a Licence To Kill. *Nat. Rev. Mol. Cell Biol.* 16, 393–405. (doi:10.1038/nrm4007)

52.	Buelna-Chontal M, Zazueta C. 2013. Redox activation of Nrf2 & NF-kB: A double end sword? *Cell. Signal.* 25, 2548–2557. (doi:10.1016/j.cellsig.2013.08.007)

53.	Topal A, Atamanalp M, Oruç E, Erol HS. 2017. Physiological and biochemical effects of nickel on rainbow trout (Oncorhynchus mykiss) tissues: Assessment of nuclear factor kappa B activation, oxidative

stress and histopathological changes. *Chemosphere* 166, 445–452. (doi:10.1016/j.chemosphere.2016.09.106)

54. Tiedke J, Thiel R, Burmester T. 2014. Molecular response of estuarine fish to hypoxia: a comparative study with ruffe and flounder from field and laboratory. *PLoS One* 9, e90778. (doi:10.1371/journal.pone.0090778)

55. Martínez R, Esteve-Codina A, Herrero-Nogareda L, Ortiz-Villanueva E, Barata C, Tauler R, Raldúa D, Piña B, Navarro-Martín L. 2018. Dose-dependent transcriptomic responses of zebrafish eleutheroembryos to Bisphenol A. *Environ. Pollut.* 243, 988–997. (doi:10.1016/J.ENVPOL.2018.09.043)

56. Rasti M, Arabsolghar R, Khatooni Z, Mostafavi-Pour Z. 2012. p53 Binds to Estrogen Receptor 1 Promoter in Human Breast Cancer Cells. *Pathol. Oncol. Res.* 18, 169–175. (doi:10.1007/s12253-011-9423-6)

57. Bailey ST, Shin H, Westerling T, Liu XS, Brown M. 2012. Estrogen receptor prevents p53-dependent apoptosis in breast cancer. *Proc. Natl. Acad. Sci.* 109, 18060–18065. (doi:10.1073/pnas.1018858109)

58. Zhou W, Lo S-C, Liu J-H, Hannink M, Lubahn DB. 2007. ERRbeta: a potent inhibitor of Nrf2 transcriptional activity. *Mol. Cell. Endocrinol.* 278, 52–62. (doi:10.1016/j.mce.2007.08.011)

59. Ansell PJ, Lo SC, Newton LG, Espinosa-Nicholas C, Zhang DD, Liu JH, Hannink M, Lubahn DB. 2005 .Repression of cancer protective genes by 17β-estradiol: Ligand-dependent interaction between human Nrf2 and estrogen receptor α. *Mol. Cell. Endocrinol.* 243, 27–34. (doi:10.1016/j.mce.2005.08.002)

60. He Q, Zhang C, Wang L, Zhang P, Ma D, Lv J, Liu F. Inflammatory signaling regulates hematopoietic stem and progenitor cell emergence in vertebrates. *Blood.* 125, 7, 1098-1106 (doi:10.1182/blood)

61. Juven-Gershon T, Kadonaga JT. 2010 .Regulation of gene expression via the core promoter and the basal transcriptional machinery. *Dev. Biol.* 339, 225–229. (doi:10.1016/j.ydbio.2009.08.009)

62. Eric T, Allen H, Goodman JM, Gutsell S, Russell PJ, Allen TEH, Goodman JM, Gutsell S, Russell PJ. 2014 .Defining Molecular Initiating Events in the Adverse Outcome Pathway Framework for Risk

Assessment. *Chem. Res. Toxicol.* 27, 2100–2112.
(doi:10.1021/tx500345j)

63. Meek B, Palermo CM, Bachman AN, North CM, Lewis RJ. 2014 .Mode of action human relevance (species concordance) framework: Evolution of the Bradford Hill considerations and comparative analysis of weight of evidence. *Appl. Toxicol.* 34. (doi:10.1002/jat.2984)

64. Park EY, Rho HM. 2002.The transcriptional activation of the human tetrachlorodibenzo- p -dioxin through two different regulator sites , the antioxidant responsive element and xenobiotic responsive element. *Mol. Cell. Biochem.,* 24, 47–55.

65. Almeida DV, Nornberg BFDS, Geracitano L a, Barros DM, Monserrat JM, Marins LF. 2010 .Induction of phase II enzymes and hsp70 genes by copper sulfate through the electrophile-responsive element (EpRE): insights obtained from a transgenic zebrafish model carrying an orthologous EpRE sequence of mammalian origin. *Fish Physiol. Biochem.* 36, 347–53. (doi:10.1007/s10695-008-9299-x)

66. Zhang X, Chen X, Song H, Chen H-Z, Rovin BH. Activation of the Nrf2/antioxidant response pathway increases IL-8 expression. 2005. *Journal of Immunology,* 35, 3258-3267. (doi:10.1002/eji.200535489)

67. Kobayashi EH, Suzuki T, Funayama R, Nagashima T, Hayashi M, Sekine H, Tanaka N, Moriguchi T, Motohashi H, Nakayama K, Yamamoto M. 2016 .Nrf2 suppresses macrophage inflammatory response by blocking proinflammatory cytokine transcription. *Nat. Commun.* 7, 11624. (doi:10.1038/ncomms11624)

68. Wruck CJ et al. 2011 .Nrf2 induces interleukin-6 (IL-6) expression via an antioxidant response element within the IL-6 promoter. *J. Biol. Chem.* 286, 4493–4499. (doi:10.1074/jbc.M110.162008)

69. Lacher SE, Levings DC, Freeman S, Slattery M. 2018. Identification of a functional antioxidant response element at the HIF-1A locus. *Redox Bio.* 19, 401-411 (doi:10.1016/j.redox.2018.08.014)

70. Yang SY, Ahmed S, Satheesh S V., Matthews J. 2017 .Genome-wide mapping and analysis of aryl hydrocarbon receptor (AHR)- and aryl hydrocarbon receptor repressor (AHRR)-binding sites in human breast cancer cells. *Arch. Toxicol. ,* 1–16. (doi:10.1007/s00204-017-2022-x)

71. Li W, Harper P a, Tang B-K, Okey AB. 1998. Regulation of cytochrome P450 enzymes by aryl hydrocarbon receptor in human cells. *Biochem. Pharmacol.* 56, 599–612. (doi:10.1016/S0006-2952(98)00208-1)

72. Lo R, Matthews J. 2012 .High-Resolution Genome-wide Mapping of AHR and ARNT Binding Sites by ChIP-Seq. *Toxicol. Sci.* 130(2), 349–361. (doi:10.1093/toxsci/kfs253)

73. Semenza GL, Agani F, Booth G, Forsythe J, Iyer N, Jiang BH, Leung S, Roe R, Wiener C, Yu A. 1997 .Structural and functional analysis of hypoxia-inducible factor 1. *Kidney Int.* 51, 553–5.

74. Brunnberg S, Pettersson K, Rydin E, Matthews J, Hanberg A, Pongratz I. 2003 .The basic helix-loop-helix-PAS protein ARNT functions as a potent coactivator of estrogen receptor-dependent transcription. *Proc. Natl. Acad. Sci. U. S. A.* 100, 6517–6522. (doi:10.1073/pnas.1136688100)

75. Yang J, AlTahan Alaa, Jones DT, Buffa FM, Bridges E, Interiano RB, Qu C, Vogt N, Li JL, Baban D, Ragoussis J, Nicholson R, Davidoff AM, Hariss AL. 2015 .Estrogen receptor-α directly regulates the hypoxia-inducible factor 1 pathway associated with antiestrogen response in breast cancer. *Proc. Natl. Acad. Sci.* 112, 15172–15177. (doi:10.1073/pnas.1422015112)

76. Lim W, Park Y, Cho J, Park C, Park J, Park Y-K, Park H, Lee Y. 2011 .Estrogen receptor beta inhibits transcriptional activity of hypoxia inducible factor-1 through the downregulation of arylhydrocarbon receptor nuclear translocator. *Breast Cancer Res.* 13, R32. (doi:10.1186/bcr2854)

77. Yao KS, O'Dwyer PJ. 1995 .Involvement of NF-kB in the induction of NAD(P)H: Quinone oxidoreductase (DT-diaphorase) by hypoxia, oltipraz and mitomycin C. *Biochem. Pharmacol.* 49, 275–282. (doi:10.1016/0006-2952(94)00544-V)

78. Kwak EL, Larochelle DA, Beaumont C, Torti S V, Torti FM. 1995 .Role for NF-kappa B in the regulation of ferritin H by tumor necrosis factor-alpha. *J. Biol. Chem.* 270, 15285–93. (doi:10.1074/jbc.270.25.15285)

79. Rojo AI. 2004 .Regulation of Cu/Zn-Superoxide Dismutase Expression via the Phosphatidylinositol 3 Kinase/Akt Pathway and Nuclear

Factor- B. *J. Neurosci.* 24, 7324–7334. (doi:10.1523/JNEUROSCI.2111-04.2004)

80. Ammirante M, Rosati A, Gentilella A, Festa M, Petrella A, Marzullo L, Pascale M, Belisario MA, Leone A, Turco MC. 2008 .The activity of hsp90 alpha promoter is regulated by NF-kappa B transcription factors. *Oncogene,* 27, 1175–8. (doi:10.1038/sj.onc.1210716)

81. Kunsch C, Rosen CA. 1993 .NF-kB Subunit-Specific Regulation of the Interleukin-8 Promoter. *Mol. Cell. Biol.* 13, 6137–6146.

82. Libermann TA, Baltimore D. 1990 .Activation of interleukin-6 gene expression through the NF-kappa B transcription factor. *Mol. Cell. Biol.* 10, 2327–2334. (doi:10.1128/MCB.10.5.2327.Updated)

83. Hiscott J et al. 1993 .Characterization of a functional NF-kappa B site in the human interleukin 1 beta promoter: evidence for a positive autoregulatory loop. *Mol. Cell. Biol.* 13, 6231–40. (doi:10.1128/MCB.13.10.6231)

84. Yu M et al. 2011 .Nuclear factor p65 interacts with Keap1 to repress the Nrf2-ARE pathway. *Cell. Signal.* 23, 883–892. (doi:10.1016/j.cellsig.2011.01.014)

85. Vogel CFA, Li W, Wu D, Miller JK, Sweeney C, Lazennec G, Fujisawa Y, Matsumura F. 2011 .Interaction of aryl hydrocarbon receptor and NF-κB subunit RelB in breast cancer is associated with interleukin-8 overexpression. *Arch. Biochem. Biophys.* 512, 78–86. (doi:10.1016/j.abb.2011.05.011)

86. Chen P-H, Chang H, Chang JT, Lin P. 2012 .Aryl hydrocarbon receptor in association with RelA modulates IL-6 expression in non-smoking lung cancer. *Oncogene* 31, 2555–2565. (doi:10.1038/onc.2011.438)

87. Kim DW, Gazourian L, Quadri S a, Romieu-Mourez R, Sherr DH, Sonenshein GE. 2000 .The RelA NF-kappaB subunit and the aryl hydrocarbon receptor (AhR) cooperate to transactivate the c-myc promoter in mammary cells. *Oncogene* 19, 5498–506. (doi:10.1038/sj.onc.1203945)

88. Kasner E, Hunter CA, Ph D, Kariko K, Ph D. 2010 .Metal transcription factor-1 regulation via MREs in the transcribed regions of selenoprotein H and other metal-responsive genes. *Biochim Biophys Acta* 70, 646–

656. (doi:10.1002/ana.22528.Toll-like)

89. Ovrevik J, Lag M, Lecureur V, Gilot D, Lagadic-Gossmann D, Refsnes M, Schwarza PE, Skuland T, Becher R, Holme JA. 2014. AhR and Arnt differentially regulate NF-kappaB signaling and chemokine responses in human bronchial epithelial cells. *Cell Commun Signal* 12, 48. (doi:10.1186/s12964-014-0048-8)

90. Mak P, Li J, Samanta S, Mercurio AM. 2015 .ERβ regulation of NF-kB activation in prostate cancer is mediated by HIF-1. *Oncotarget* 6, 40247–40254. (doi:10.18632/oncotarget.5377)

91. Forsythe JA, Jiang BH, Iyer N V, Agani F, Leung SW, Koos RD, Semenza GL. 1996. Activation of vascular endothelial growth factor gene transcription by hypoxia-inducible factor 1. *Mol. Cell. Biol.* 16, 4604–4613. (doi:10.1128/MCB.16.9.4604)

92. Hahn ME, McArthur AG, Karchner SI, Franks DG, Jenny MJ, Timme-Laragy AR, Stegeman JJ, Woodin BR, Cipriano MJ, Linney E. 2014 .The Transcriptional Response to Oxidative Stress during Vertebrate Development: Effects of tert-Butylhydroquinone and 2,3,7,8-Tetrachlorodibenzo-p-Dioxin. *PLoS One* 9, e113158. (doi:10.1371/journal.pone.0113158)

93. Huang W-J, Xia LM, Zhu F, Huang B, Zhou C, Zhu HF, Wang B, Chen B, Lei P, Shen GX, De-An T. 2009 .Transcriptional upregulation of HSP70-2 by HIF-1 in cancer cells in response to hypoxia. *Int. J. Cancer* 124, 298–305. (doi:10.1002/ijc.23906)

94. Samarasinghe B, Wales CTK, Taylor FR, Jacobs AT. 2014 .Heat shock factor 1 confers resistance to Hsp90 inhibitors through p62/SQSTM1 expression and promotion of autophagic flux. *Biochem. Pharmacol.* 87, 445–455. (doi:10.1016/j.bcp.2013.11.014)

95. Mendillo ML, Santagata S, Koeva M, Bell GW, Hu R, Tamimi RM, Fraenkel E, Ince TA, Whitesell L, Lindquist S. 2012 .HSF1 drives a transcriptional program distinct from heat shock to support highly malignant human cancers. *Cell* 150, 549–562. (doi:10.1016/j.cell.2012.06.031)

96. Choi J-E, Kim JH, Song NY, Suh NY, Kim DH, Kim SJ, Na HK, Nadas J, Dong Z, Cha YN, Surh YJ. 2016. 15-Deoxy-$\Delta^{12,14}$-prostaglandin J$_2$

stabilizes hypoxia inducible factor-1α through induction of heme oxygenase-1 and direct modification ofprolyl-4-hydroxylase 2. *Free Radic. Res.* 50, 1140–1152. (doi:10.1080/10715762.2016.1219352)

97. Wu J, Williams D, Walter GA, Thompson WE, Sidell N. 2014. Estrogen increases Nrf2 activity through activation of the PI3K pathway in MCF-7 breast cancer cells. *Exp. Cell Res.* 328, 351–360. (doi:10.1016/j.yexcr.2014.08.030)

98. Tebay LE, Robertson H, Durant ST, Vitale SR, Penning TM, Dinkova-Kostova AT, Hayes JD. 2015 .Mechanisms of activation of the transcription factor Nrf2 by redox stressors, nutrient cues, and energy status and the pathways through which it attenuates degenerative disease. *Free Radic. Biol. Med.* 88, 108–146. (doi:10.1016/j.freeradbiomed.2015.06.021)

99. Lim JW, Kim H, Kim KH. 2001 .Nuclear factor-kappaB regulates cyclooxygenase-2 expression and cell proliferation in human gastric cancer cells. *Lab. Invest.* 81, 349–60.

100. Montano MM, Deng H, Liu M, Sun X, Singal R. 2004. Transcriptional regulation by the estrogen receptor of antioxidative stress enzymes and its functional implications. *Oncogene* 23, 2442–53. (doi:10.1038/sj.onc.1207358)

101. Bruning U, Fitzpatrick SF, Frank T, Birtwistle M, Taylor CT, Cheong A. 2012. NFkB and HIF display synergistic behaviour during hypoxic inflammation. *Cell. Mol. Life Sci.* 69, 1319–1329. (doi:10.1007/s00018-011-0876-2)

102. Chen W, Sun Z, Wang XJ, Jiang T, Huang Z, Fang D, Zhang DD. 2009 .Direct Interaction between Nrf2 and p21Cip1/WAF1 Upregulates the Nrf2-Mediated Antioxidant Response. *Mol. Cell* 34, 663–673. (doi:10.1016/j.molcel.2009.04.029)

103. Xing D, Oparil S, Yu H, Gong K, Feng W, Black J, Chen YF, Nozell S. 2012. Estrogen modulates NFkB signaling by enhancing IkBa levels and blocking p65 binding at the promoters of inflammatory genes via estrogen receptor-b. *PLoS One* 7, 1–10. (doi:10.1371/journal.pone.0036890)

104. Fischer M. 2017 .Census and evaluation of p53 target genes.

Oncogene 36, 3943–3956. (doi:10.1038/onc.2016.502)

105. Pal S, Datta K, Mukhopadhyay D, Endothelial V, Factor G, Vegf VPF. 2001 .Central Role of p53 on Regulation of Vascular Permeability Factor Vascular Endothelial Growth Factor (VPF / VEGF) Expression in Mammary Carcinoma. *Cancer Res.* 61, 6952–6957.

106. Arizti P, Fang L, Park I, Yin Y, Solomon E, Ouchi T, Aaronson SA, Lee SW. 2000 .Tumor suppressor p53 is required to modulate BRCA1 expression. *Mol. Cell. Biol.* 20, 7450–7459.

107. Hafner A, Stewart-Ornstein J, Purvis JE, Forrester WC, Bulyk ML, Lahav G. 2017. p53 pulses lead to distinct patterns of gene expression albeit similar DNA-binding dynamics. *Nat. Struct. Mol. Biol.* , 1–14. (doi:10.1038/nsmb.3452)

108. Sano M, Minamino T, Toko H, Miyauchi H, Orimo M, Qin Y, Akazawa H, Tateno K, Kayama Y, Harada M, Schimizu I, Ashara T, Hamada H, Tomita S, Molentin JD, Zou Y, Komuro I. 2007. p53-induced inhibition of Hif-1 causes cardiac dysfunction during pressure overload. *Nature* 446, 444–448. (doi:10.1038/nature05602)

109. Meiller A, Alvarez S, Drané P, Lallemand C, Blanchard B, Tovey M, May E. 2007. p53-dependent stimulation of redox-related genes in the lymphoid organs of γ-irradiated mice - Identification of Haeme-oxygenase 1 as a direct p53 target gene. *Nucleic Acids Res.* 35, 6924–6934. (doi:10.1093/nar/gkm824)

110. Robles AI, Bemmels NA, Foraker AB, Harris CC. 2001 .APAF-1 Is a Transcriptional Target of p53 in DNA Damage-induced Apoptosis Advances in Brief APAF-1 Is a Transcriptional Target of p53 in DNA Damage-induced Apoptosis. *Cancer Res.* 61, 6660–6664.

111. Catz SD, Johnson JL. 2001. Transcriptional regulation of bcl-2 by nuclear factor kB and its signifcance in prostate cancer. *Oncogene,* 20, 7342-7351.

112. Mitsuishi Y, Taguchi K, Kawatani Y, Shibata T, Nukiwa T, Aburatani H, Yamamoto M, Motohashi H. 2012 .Nrf2 Redirects Glucose and Glutamine into Anabolic Pathways in Metabolic Reprogramming. *Cancer Cell*, 22, 66–79. (doi:10.1016/j.ccr.2012.05.016)

113. Catz SD, Johnson JL. 2001 .Transcriptional regulation of bcl-2 by

nuclear factor κB and its significance in prostate cancer. *Oncogene* 20, 7342–7351. (doi:10.1038/sj.onc.1204926)

114. Gabai VL, Meng L, Kim G, Mills TA, Benjamin IJ, Sherman MY. 2012 .Heat Shock Transcription Factor Hsf1 Is Involved in Tumor Progression via Regulation of Hypoxia-Inducible Factor 1 and RNA-Binding Protein HuR. *Mol. Cell. Biol.* 32, 929–940. (doi:10.1128/MCB.05921-11)

115. Hu Y, Dietrich D, Xu W, Patel A, Thuss JAJ, Wang JJ, Yin W, Qiao K, Houk KN, Vedras JC, Tang Y. 2015 .HSF1 regulation of b-catenin in mammary cancer cells through the control of HUR/ELAVL1 expression. *Oncogene,* 10, 552–554. (doi:10.1038/nchembio.1527.A)

116. Kothari S, Cizeau J, McMillan-Ward E, Israels SJ, Bailes M, Ens K, Kirshenbaum LA, Gibson SB. 2003 .BNIP3 plays a role in hypoxic cell death in human epithelial cells that is inhibited by growth factors EGF and IGF. *Oncogene* 22, 4734–4744. (doi:10.1038/sj.onc.1206666)

117. Kaidi A, Qualtrough D, Williams AC, Paraskeva C. 2006. Direct Transcriptional Up-regulation of Cyclooxygenase-2 by Hypoxia-Inducible Factor (HIF) -1 Promotes Colorectal Tumor Cell Survival and Enhances HIF-1 Transcriptional Activity during Hypoxia. *Cancer Res.,* 66, 13, 6683–6692. (doi:10.1158/0008-5472.CAN-06-0425)

118. Trisciuoglio D, Gabellini C, Desideri M, Ziparo E, Zupi G, del Bufalo D. 2010. Bcl-2 regulates HIF-1a protein stabilization in hypoxic melanoma cells via the molecular chaperone HSP90. *PLoS One*, 5, 7, e11772 (doi:10.1371/journal.pone.0011772)

119. Nakano K, Vousden KH, 2001 .PUMA, a novel proapoptotic gene, is induced by p53. *Mol. Cell* 7, 683–94. (doi:10.1016/S1097-2765(01)00214-3)

120. Sachdeva M, Zhu S, Wu F, Wu H, Walia V, Kumar S, Elble R, Watabe K, Mo Y-Y. 2009 .p53 represses c-Myc through induction of the tumor suppressor miR-145. *Proc. Natl. Acad. Sci. U. S. A.* 106, 3207–3212. (doi:10.1073/pnas.0808042106)

121. Sax JK, Fei P, Murphy ME, Bernhard E, Korsmeyer SJ, El-Deiry WS. 2002. BID regulation by p53 contributes to chemosensitivity. *Nat. Cell Biol.* 4, 842–849. (doi:10.1038/ncb866)

122. Abdel-Razzak Z, Garlatti M, Aggerbeck M, Barouki R. 2004

.Determination of interleukin-4-responsive region in the human cytochrome P450 2E1 gene promoter. *Biochem. Pharmacol.* 68, 1371–1381. (doi:10.1016/J.BCP.2004.06.003)

123. Itoh K, Mochizuki M, Ishii Y, IshiiT, Shibata T, Kawamoto Y, Kelly V, Sekizawa K, Uchida K, Yamamoto M. 2004. Transcription factor Nrf2 regulates inflammation by mediating the effect of 15-deoxy-Delta(12,14)-prostaglandin j(2). *Mol. Cell. Biol.* 24, 36–45. (doi:10.1128/MCB.24.1.36-45.2004)

124. Hoesel B, Schmid JA. 2013 .The complexity of NF-κB signaling in inflammation and cancer. *Mol. Cancer* 12, 86. (doi:10.1186/1476-4598-12-86)

125. Luecke S, Wincent E, Backlund M, Rannug U, Rannug A. 2010. Cytochrome P450 1A1 gene regulation by UVB involves crosstalk between the aryl hydrocarbon receptor and nuclear factor κB. *Chem. Biol. Interact.* 184, 466–473. (doi:10.1016/J.CBI.2010.01.038)

126. Zamorano J, Mora AL, Boothby M, Keegan AD. 2001 .NF-kappa B activation plays an important role in the IL-4-induced protection from apoptosis. *Int. Immunol.* 13, 1479–87.

127. Liu M-J, Bao S, Galvez-Peralta M, Pyle C, Rudawsky AC, Pavlovicz RE, Killilea DW, Li C, Nebert DW, Wewers MD, Knoell DJ. 2013. ZIP8 regulates host defense through zinc-mediated inhibition of NF-κB. *Cell Rep.* 3, 386–400. (doi:10.1016/j.celrep.2013.01.009)

128. Kelleher ZT, Matsumoto A, Stamler JS, Marshall HE. 2007 .NOS2 Regulation of NF-κB by *S* -Nitrosylation of p65. *J. Biol. Chem.* 282, 30667–30672. (doi:10.1074/jbc.M705929200)

129. Sowter HM, Ratcliffe PJ, Watson P, Greenberg AH, Harris AL. 2001. HIF-1-dependent regulation of hypoxic induction of the cell death factors BNIP3 and NIX in human tumors. *Cancer Res.* 61, 6669–6673.

130. Huang LE. 2008. Carrot and stick: HIF-α engages c-Myc in hypoxic adaptation. *Cell Death Differ.* 15, 672–677. (doi:10.1038/sj.cdd.4402302)

131. De Bruin A, Cornelissen PWA, Kirchmaier BC, Mokry M, Lich R, Nirmala E, Liang KH, Vegh AMD, Scholman KT, Groot Koerkamp MJ, Holstege FC, Cuppen E, Schulte-Merker S, Bakker WJ. 2015.Genome-

wide analysis reveals NRP1 as a direct HIF1α-E2F7 target in the regulation of motorneuron guidance in vivo. *Nucleic Acids Res.* 44, 3549–3566. (doi:10.1093/nar/gkv1471)

132. Zheng X, Krakowiak J, Patel N, Beyzavi A, Ezike J, Khalil AS, Pincus D. 2016 .Dynamic control of Hsf1 during heat shock by a chaperone switch and phosphorylation. *Elife* 5. e18638 (doi:10.7554/eLife.18638)

133. Kong F, Wang H, Guo J, Peng M, Ji H, Yang H, Liu B, Wang J, Zhang X, Li S, 2016 .Hsp70 suppresses apoptosis of BRL cells by regulating the expression of Bcl-2, cytochrome C, and caspase 8/3. *Vitr. Cell. Dev. Biol. - Anim.* 52, 568–575. (doi:10.1007/s11626-016-0005-5)

134. Bellezza I, Tucci A, Galli F, Grottelli S, Mierla AL, Pilolli F, Minelli A. 2012 .Inhibition of NF-κB nuclear translocation via HO-1 activation underlies α-tocopheryl succinate toxicity. *J. Nutr. Biochem.* 23, 1583–1591. (doi:10.1016/J.JNUTBIO.2011.10.012)

Chapter 3

A critical analysis of *in silico* methods used for identifying transcription factor binding sites across fish species.

3.1 Abstract

Transcription factor binding sites (TFBS) exhibit a high level of evolutionary conservation across distantly related species and *in silico* predictive methods have been widely used to inform on the identification of target genes for specific transcription factors (TFs). Currently, a number of databases exist that contain positional weight matrixes (PWMs), a form of DNA binding motif largely derived from mammalian cell lines. PWMs are used across vertebrate groups to predict downstream target genes for TFs of interest. Considering that multiple TFs can regulate the expression of a single gene, identifying functional binding sites across distantly related species can establish the conservation of regulatory interactions such as those at the basis of gene regulatory networks (GRNs). However, it is also necessary to establish the efficiency of binding-site identification tools in distantly related vertebrate species, such as fish, where there is the potential that sequence divergence could affect search efficiency. This analysis assessed the level of divergence in validated binding sites of the adaptive stress response transcription factors nuclear-factor erythroid 2-related factor 2 (Nrf2), aryl hydrocarbon receptor (AhR), metal transcription factor 1 (MTF1) and hypoxia inducible factor-1α (HIF-1α) in fish species compared to mammalian-derived consensus sequences. Divergent sequences with a base change from the consensus were identified in Nrf2 and HIF-1α target genes suggesting that traditional motif-discovery tools would not identify functional binding sites. Using random forest models, the ability of mammalian Nrf2 and HIF-1α motifs derived from JASPAR and HOCOMOCO databases to predict downstream targets across selected teleost fish species was investigated. Whilst mammalian motifs for Nrf2 showed little difference in the ability to predict target genes across fish and mammals, for HIF-1α, random forest models were less efficient at predicting targets in positive and negative groups. The conservation of binding sites in target genes regulated by multiple adaptive stress response TFs was then investigated for heme-oxygenase 1 and heat shock protein 70 using both mammalian and fish-derived motifs. This research showed that whilst mammalian PWMs provided putative predictions of binding sites across fish species, fish-specific motif files were also able to identify putative TFBS that differed from mammalian search results.

3.2 Introduction:

Aquatic environments have experienced the largest decline in biodiversity of any ecosystem[1] and there is a clear need to generate methods that predict the potential effects of toxic pollution and environmental change. The activation of adaptive stress response transcription factors (TFs) is a fundamental outcome to toxicological insult, targeted by a wide range of pollutants of concern. These TFs include nuclear factor erthyriod-like 2 (Nrf2), metal transcription factor 1 (MTF1), the aryl hydrocarbon receptor (AhR), heat shock factor 1 (HSF1), tumour protein p53 (p53), nuclear factor kappa B (NFkB) and hypoxia inducible factor 1 alpha (HIF-1α) which have all been shown to be evolutionary conserved across aerobic organisms [2]. Each factor associates with specific DNA binding regions, termed transcription factor binding sites (TFBS) - short degenerate sequences that exhibit a high level of evolutionary conservation [3–6]- in order to control the regulation of target genes: Nrf2 binds to the Electrophile response element (EpRE), HIF-1α with hypoxia responsive elements (HREs) [7], AhR with dioxin response elements (DREs) [8] and MTF1 with metal response elements (MREs) (Figure 1.6) [9]. Although historically associated with distinct response outcome processes, the identification of cross talk and shared downstream targets between TFs in mammalian cell lines suggests that similar stress-response processes are activated under a range of inducers, such as antioxidant defence processes (Chapter 2). This underpins the adaptive stress response as a gene regulatory network (GRN) in mammals and has the potential to provide a novel approach in predictive toxicology by establishing the regulatory responses in vertebrates at high risk of pollutant exposures such as fish.

The identification of TFBS in non-coding regions provides putative predictions of downstream target genes and can aid in determining the conservation of regulatory networks at the basis of adverse outcome pathways (AOPs)[10]. Currently, *in silico* searches for TFBS are largely based on sequence motifs in the form of positional weight matrices (PWMs), derived from multiple sequence alignments (MSAs) of experimentally validated sites [11,12]. Databases of vertebrate PWMs have become widely available and include Transfac [13],

JASPAR [14], HOCOMOCO[15] , the Transcription Factor encyclopedia (TFe) [16] and PAZAR [17], largely comprising of PWMs from systematic evolution of ligands by exponential enrichment (SELEX) experiments in mammalian cell lines.

However, searches for PWMs produce a large number of false positive hits as the probability of a short DNA sequence occurring at random in the genome is high[18]. PWMs are therefore used as a means of establishing target genes in combination with experimental evidence. Identifying binding sequences in returned peaks in whole genome sequencing methods such as Chromatin Immunoprecipitation sequencing (ChIP-Seq) and Assay for Transposase-Accessible Chromatin using sequencing (ATAC-seq) is reliant on PWM search systems [12]. Gene-specific binding site identification methods are also reliant on PWM searches; regulatory regions are searched for putative binding sites which are later validated experimentally using methods such as reporter constructs and TF knockouts [21–23]. Despite the chance of false positive hits, binding site predictions can be used to predict regulatory interactions from gene-expression datasets in the absence of direct TF-DNA binding evidence, a method widely used to identify GRNs [24]. Reducing the number of false-positive hits from PWM searches is therefore essential and functional binding sites have been associated with enrichment, position, distance from the transcription start site (TSS), orientation and evolutionary conservation[25–27]. Importantly, the similarity of the returned hits to the motif file used is widely associated with a greater probability that the identified sequence will be functional [14].

From an aquatic toxicology setting, as the majority of PWMs are derived from mammals and there is, to current knowledge, only one ChIP-Seq study for adaptive-stress response TFs at the whole genome level (for HIF-1α as discussed below), the level of conservation of binding sites across distantly related vertebrates (mammals to teleost fish) is largely unknown. The lack of fish-specific PWMs is not necessarily seen as problematic in predicting regulatory regions because of the ability of TF proteins derived from fish to bind to mammalian TFBS and vice-versa as documented in *Danio rerio*

(zebrafish) for Nrf2 [28]. However, in the context of Nrf2, which has been shown to have a conserved function in regulating antioxidant genes across vertebrates [29], differences in the underlying binding sequence have been identified in validated sites in comparison to mammalian counterparts. Nrf2 binds to the electrophile response element (EpRE) to initiate transcription of downstream target genes, including glutathione-s-transferase P (*gstp1*) [30]. Whilst few Nrf2 binding sites have been identified in fish species, a functional EpRE region in the *gstp1* promoter in zebrafish differed from the consensus sequence with a base changed to a T [30]. This change in binding sequence prevents its identification using both conventional consensus sequences searches and with mammalian-derived EpRE PWMs, which are searched with high stringency settings. Similarly, a whole-genome ChIP-Seq study on HIF-1α in zebrafish using *de novo* motif discovery identified a binding sequence that showed some variation from the mammalian HIF-1α matrix [7]. The impact of species-specific TFBS matrices was shown for p53, where a PWM formed from validated binding sites in zebrafish was able to identify novel target genes that failed to be recovered using the mammalian matrix alone [21].

Divergences in binding sequences between genes and across species can lead to variations in transcriptional responses. For example, specific binding sequences for NFkB have been associated with gene-specific response trajectories [31]. In addition, single nucleotide polymorphisms (SNPs) in TFBS are widely associated with an increased susceptibility to disease and cancer [32]. In fish species, divergences in TFBS identified across two species of cichilid fish were hypothesised to contribute to observed differences in opsin gene expression between species, suggesting the binding sequence was under an evolutionary pressure [33]. Differences in tissue-specific expression pattern have been reported in zebrafish transgenics where human and zebrafish-derived reporter constructs consisting of conserved non-coding elements (CNEs) have been compared [34]; here, divergences in transcription factor binding sites were shown to be correlated to the differences in expression pattern [34]. Biochemically, changes in sequence composition can alter the binding strength of the transcription factor-DNA interactions resulting in a change in the transcriptional output [35].

Changes in binding sequence from mammals to fish is problematic as gene-expression datasets in fish toxicology studies are regularly searched for binding sites to predict the regulatory mechanisms behind response processes[36]. *In silico*, binding sites are predicted by searching non-coding DNA sequences for matches to the regulatory motifs of interest, returning a score (as the q-value, p-value or z-score depending on the statistical method), representing the similarity of the identified sequence to the motif file used. Transcription factor sequence divergence (TFSD) across fish species may cause PWMs derived from mammals to miss-identify regulatory sequences through *in silico* approaches, excluding functional sites from being validated and potentially limiting the identification of novel target genes. The development of transgenic models (in aquatic toxicology zebrafish and *Oryzias latipes* (medaka) are the most widely used fish model organisms[37]) is also reliant on the selection of sensitive binding sequences for transcription factors which are expected to be targeted by pollutants of concern, and this is supported by *in silico*-based sequence identification [38].

This research investigated whether TFBS derived from mammals were sufficiently conserved for adaptive stress response TFs to accurately predict downstream target genes in the adaptive stress-response pathway in fish. A literature search highlighted TFBD in some but not all validated sites across fish species. Alignments of validated binding sequences in fish species were used to generate PWMs for fish-specific TFs for Nrf2, HIF-1α, MTF1 and the AhR. Teleost fish species were selected to cover both a broad evolutionary range (Figure S3.1) and as representatives of widely used species in ecotoxicology and included the zebrafish, stickleback (*Gasterosteus aculeatus)*, tilapia (*Orechromis niloticus*), medaka and fugu *(Takifugu rubripes*) (Figure S3.1)[39]. Where TFBD was observed for Nrf2:EpRE and HIF-1α:HRE, the efficiency of mammalian motifs to identify target genes from a background gene set was conducted using random forest classification models to distinguish between known and unknown target genes. A random forest model of the efficiency of the fish-specific Nrf2:EpRE PWM searches was also conducted but showed a poor predictive performance using this motif. Finally, the conservation of shared interactions was identified in heme oxygenase 1

(*hmox1*) and heat shock protein 70 (*hsp70*), which are targeted by multiple adaptive stress response factors in mammals. Binding site analysis of promoter regions using mammalian motifs and the fish-specific motifs generated in this study was conducted and this showed variations in the ability to identify sites using motifs from different species. The results of this study show that whilst mammalian motifs on the whole work well for identifying putative binding sites for adaptive-stress response factors in fish species, novel TFBS targets can also be identified using fish-specific motifs.

3.3 Methods

3.3.1 Identification of experimentally validated sites across fish species:

A literature analysis identified experimentally validated binding sites across fish species for Nrf2, HIF-1α, MTF1, HSF1 and the AhR. Experimental validation was determined through either gene-reporter assays conducted *in vitro* or *in vivo* or through ChIP-Seq assays where appropriate. All sequences were compared to the consensus sequence for the factors of interest and against mammalian PWMs from the JASPAR database [44] (Figure S3.2) .

3.3.2 Promoter sequence retrieval:

The efficiency of PWMs derived from mammals for Nrf2 and HIF-1α to predict downstream targets in fish was conducted following the identification of divergent binding sites in the literature analysis. Twenty genes that are validated downstream targets for each factor were used as positive datasets and are shown in the Table S3.1 and Table S3.2, respectively. 1:1 orthologs were obtained against the human gene ids for the selected genes using a tree-based method in Ensembl Biomart [40]. For all genes, sequences were obtained from the Ensembl Biomart site with 3.5 kb and 5 kb of the upstream flank-coding region extracted for each species respectively[41].

Where paralog genes existed, all sequences were used in subsequent analysis. The genome assembly files for human (GRCh38.p10), mouse (GRCm38.p5), zebrafish (GRCzV9), medaka (HdrR), tilapia (Orenil1.0) and fugu (FUGU 4.0) were used in the analysis.

3.3.3 Background gene-set generation:

A list of eighty human gene names were generated using the Molbiotools software [42] and used as a background gene-set of unknown downstream targets (Table S3.3). These genes are not known to be regulated by HIF-1α or Nrf2 and all genes used in the analysis are annotated across all species. For each id, the orthologs were collected as described above for each species and the flank-coding region was extracted for 3.5 kb and 5 kb of each gene using Biomart.

3.3.4 Identification of binding sites using the MEME software:

Binding sites were identified using the command line version of the Find Individual Motif Occurrences (FIMO) program in the MEME suite 4.10.2. [43]. FIMO provides the q-value and p-value score. The p-value represents the probability match to the PWM and the q-value is the p-value adjusted against the background genomic file generated from the whole genome GC% content to give a measure of the number of significant results that are incorrect. In this case, a q-value = 1 indicates that all matches that are significant are caused by chance whereas a value = 0.01-0.05 is widely regarded as a reliable threshold score for identifying positive binding sites[43].

The results from the FIMO analysis are highly sensitive to background GC% content. For each species, the FIMO search was compared to a genomic background frequency file created from 3.5 kb and 5 kb of the flank-coding region of all genes extracted using the Ensembl Biomart [41] search tool for each species respectively. The GC% content was calculated using the *fasta-get-Markov* command in the MEME suite to generate a 0-order Markov Model file as shown in Table S3.4).

Transcription factor motifs obtained from JASPAR (2016) [44] and HOCOMOCO v10 [15] were converted into the MEME file format using the *jaspar2meme* command. Motif images were created using the *cseq* program in MEME [43]. Using FIMO, a matrix search was conducted on positive (known target genes) and background (unknown target genes) sequences by searching both DNA sequence strands using the species-specific genomic background frequencies.

3.3.5 Variations in q-value

For each species following the FIMO analysis, the distribution of top-scoring q-values for every gene in the known downstream target and unknown downstream target gene groups were compared. Differences in q-values between known and unknown target gene groups not meeting the assumptions of normality were calculated using the Mann-Whitney U test in R studio (version 3.4.2). Data normally distributed was analysed using a t-test.

3.3.6 Random forest classification models

Classification models were built to determine the efficiency of mammalian-derived motif files to categorize regulated genes against the background gene-sets across fish species using the information returned from the FIMO analysis. The top-scoring q-value, p-value, enrichment, matrix score and position were used as variables. Preliminary analysis identified random forest models as the best classifier using the *caret* package [41] in R studio (version 3.4.2) as shown in Table S3.5.

Random forest models comprise of a series of decision trees which are constructed on information from variables in training data to classify unknown data points into their respective groups (e.g. known downstream target/unknown downstream target groups). Trees are trained on known subsets of observations before being used to predict unknown observations. Each decision tree votes on the classification category with the combined votes averaged to give an overall prediction of the classification group for the sample. The predictions of the model can therefore be improved by increasing the number of trees until improvements to classification performance become stabilized.

In this research, positive (known target genes) and background (unknown target genes) datasets across mouse and humans were combined to create a mammalian training dataset. A fish-specific training dataset was also created by combining the data from the FIMO analysis for stickleback, zebrafish, medaka, tilapia and fugu. Training datasets consisted of a minimum of 20% positive and 80% background data in the case of the mammalian dataset. Two random forest models for each motif (for 3.5 kb and 5 kb search results) where constructed and trained on the mammalian dataset and fish-specific dataset respectively. To determine the model efficiency, models were used firstly to classify the training dataset (e.g. mammalian dataset for the mammalian-trained model) before being used to predict the classification of the unknown dataset (e.g. fish). Successful random forest models where then used to classify species-specific datasets.

Model tuning parameters and optimal model selection was conducted using the inbuilt training features of the *caret* package in R studio. To avoid over-fitting outcomes to the training data, models were generated using leave-out out cross-validation (LOOCV). LOOCV builds models on all data in the training set, leaving one sample out in each case, with its classification predicted based on the model generated from all other data. This process was repeated so all data were used for both training and prediction purposes. Training data predictions were pre-processed using the center and scale features and the number of trees was set to 1000. The final model in each case was selected based on the accuracy score. The accuracy score and number of decision trees for each model is shown in the supplementary information (Table S3.6 – S3.7).
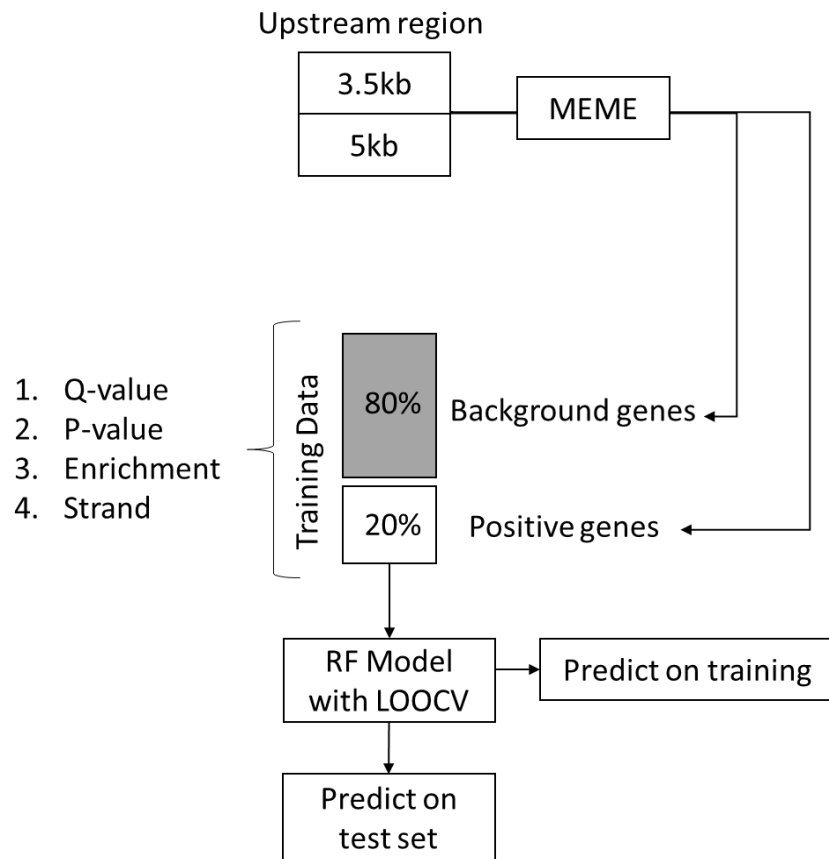
Figure 3.1. Schematic of classification models: Random forest models were trained on either mammalian or fish-specific data containing the q-value, p-value, enrichment score and strand for each gene-set using leave on out cross-validation (LOOCV). Model performance was assessed by predicting the classification of the training data (e.g. mammalian model predicting mammalian data) before predicting classification of the test data (e.g. mammalian model predicting fish data). 'Background genes' represent unknown downstream targets randomly generated from the Molbiotools software [42] whereas 'Positive genes' represent known targets identified from the literature.

3.3.7 Receiver Operator Characteristics (ROC) curves:

The efficiency of each model to predict the classification categories of fish-derived and mammalian-derived sequences can be represented in a confusion matrix as shown in Figure 3.2. This represents the performance of the model by showing if the data recalled is either correctly classified (e.g. as a true positive or true negative) or incorrectly classified (e.g. as a false positive or false negative).

Predicted class

Actual class

TP	FN
FP	TN

Figure 3.2. Confusion matrix of classification scores where T = True Positive, FP= False positive, FN= False negative, TN = True negative.

Receiver operator characteristics (ROC) curves provide a visual representation of a model's classification performance by plotting the sensitivity value against 1-specificty values across all data points. The sensitivity of the model is a measure of the number of correct classifications (true positive hits) against the number of total positive hits and is calculated as:

$$\frac{TP}{TP + TN}$$

TP= True positive, TN = True negative.

The specificity of the model predicts the likelihood that the results will be a true negative hit and is calculated as:

$$\frac{TN}{FN + FP}$$

TN = True negative, FN= False negative, FP = False positive.

The area under the curve (AUC) for each ROC plot can be used as an indicator of model performance. A model that correctly predicts the classification of data points will have an AUC = 1.0 whereas a model that randomly classifies data points will have an AUC = 0.5. Therefore, AUC values in the range of 0.5- 0.7 = poor, 0.7- 0.8 = good and 0.8 – 1.0 = excellent.

For each species, ROC curves were plotted using the *pROC* package [41] in R studio (version 3.4.2). The standard deviation for AUC values was calculated taking the square root of the variance value, derived from the *var(roc.plot)* command in *pROC*.

3.3.8 Conserved binding motifs in shared downstream targets:
Heme oxygenase 1 (*hmox1*) and heat shock protein 70 (*hsp70*) have both been identified as being regulated by multiple stress-responsive transcription factors with HMOX1 responding to NFkB [45], p53 [46] and Nrf2 [45] and Hsp70 to Nrf2 [47], P53 [48], HIF-1α [49] and HSF1[50]. For each gene, promoters were obtained from Ensembl (as described in section 3.3.2.) with 3.5 kb of the promoter region was searched using the FIMO command line in the MEME suite [42]. Alignments of the fish-specific sequences derived from the literature search were used to create position-factor matrixes and these were converted to meme motifs using the *jaspar2meme* command in the meme suite. For p53, the fish-specific motif as described by Mandriani B.*et al.* 2016 [21] was used to search promoter regions. The mammalian motifs for NFkB (MA0105.4, MA0778.1), p53 (MA0106.2), HSF1 (MA0486.1), Nrf2 (MA0150.2), HIF-1α (MA1106.1) and AhR (MA0006.1) from the JASPAR database were also used and are shown in the supplementary information (Figure S3.2). Putative sites were plotted based on position from transcription start site (TSS) and sites were aligned using the T-coffee multiple sequence alignment (MSA) tool [51].

3.4 Results

3.4.1 Validated binding sites across teleost fish:

The literature search identified few occurrences of experimentally validated sequences across fish species for the EpRE (Figure 3.3). EpRE elements have been validated for the *Oncorhynchus kisutch* (Coho salmon) and zebrafish. TFSD was identified in the *gstp1* gene in zebrafish, validated by a GFP-promoter assay using diethlymalate (DEM) as an inducer. EpRE elements identified in Coho Salmon matched the mammalian consensus sequence. Alignments of validated sequences produced a PWM that strongly matched the consensus with the exception of the T in position 8, which is not present in mammalian motifs (as shown in Figure 3.5).

For HIF-1α, ChIP-Seq analysis in zebrafish identified a binding region that diverged from the consensus sequence in the first two base pairs where R (A/G) was substituted for Y (A/G/C). All other sequences across species matched the consensus except for in the lactate dehydrogenase B gene in *Fundulus heteroclitus* (Atlantic killifish) where the second position in the sequence did not match the consensus and was changed to a T (Figure 3.4). The aligned motif file was formed of a short sequence that matched the core GTG motif.

Validated binding sites for the AhR:DRE showed no variation in sequence hit to the consensus (Figure 3.5) with sites validated across the zebrafish, medaka, rainbow trout and Atlantic killifish. Aligned sequence therefore matched the consensus strongly, which is similar to the mammalian AhR:DRE motif (Figure S3.5) as shown in the JASPAR database.

There were a number of validated binding sites for MTF1 across fish species, each of which matched the consensus motif file (Figure 3.6). Multiple MRE were associated with each gene and validated in the promoter region. The alignment of MTF1 sequences produced a matrix file with little variation from the consensus. This differs from the mammalian motif, which comprises of a

longer sequence (14 bp) though little variation around the core consensus (Figure S3.2).

A.

TABLE 1.   Nrf2:EpRE binding sites.
Consensus = 5'TGACnnnGC3'

Species	Sequence	Pos. (bp)	Method	Inducer	Ref.
Zebrafish					
gstp	TGACTCA**T**C	-45	GFP-promoter	DEM	[4]
Coho Salmon					
hmox1	TGACAGGGC	-134	*In vitro* Luc Assay (ZEM2S)	CMV-driven cDNAs of zebrafish paralogs Nrf2a or -2b	[52]
prdx1	TGACTTTGC	-141			
	TGACTGCGC	-74			
gclc	TGACTCAGC	-194			
gstp	TGACTTTGC	-71			

B.

Figure 3.3. Nrf2:EpRE binding sites across teleost fish compared to the consensus sequence TGACnnnGC. A). Validated binding sites were identified in glutathione-s-transferase pi (gstp), heme-oxygenaase 1 (homx1), perodoxin 1 (*prdx1*), glutamate-cysteine-ligase (gclc) in zebrafish and coho salmon). Divergences from the consensus sequence are shown in red. The zebrafish and Coho Salmon where the only teleost species identified which had experimentally validated EpRE regions. Position from TSS shown in base pairs (bps). B). Weblogo of validated Nrf2:EpRE binding sites in fish species.

A.

TABLE 2. HIF-1α:HRE binding sites. Consensus = 5' RCGTG 3'					
Species	**Sequence**	**Pos. (bp)**	**Method**	**Inducer**	**Ref.**
Zebrafish Whole Genome	**VY**GTGH	N/A	ChIP-Seq	Vhl mutant	[7]
Igfbp-1	ACGTG	-1086	EMSA assay	Competition	[53]
	GCGTG	-1066			
Fugu					
epo	ACGTG	-2643	*In vitro* Luc Assay	Hypoxia	[54]
Atlantic Killifish					
epo	A**T**GTG	-195	*In vitro* Luc Assay	Hypoxia	[55]

B.

Figure 3.4. Validated HIF-1α:HRE binding sites across teleost fish species compared to the consensus sequence RCGTG. A). Validated binding sites were identified in the whole genome and for insulin-like growth factor binding protein 1 (igfbp1) in zebrafish and erythroprotien (epo) in fugu and the atlantic killifish. Binding sites are shown by the IUPAC alphabet where V = A/C/G, R =A/G, H= A/C/T and Y = C/T . Variation in binding sites were identified in the zebrafish and atlantic killifish and are highlighted in red. B). Weblogo of validated HIF-1α:HRE binding sites in fish species.

A.

Species	Sequence	Pos. (bp)	Method	Inducer	Ref.
TABLE 3.		AhR:DRE binding sites.			
		Consensus = 5' TNGCGTG 3'			
Zebrafish					
Cyp1a1	TCGCGTG	-2256	*In vitro* Luc Assay (Hepa1c1c7)	TCDD	[56]
	TCGCGTG	-2550			
	TTGCGTG	-2592			
Atlantic Killifish					
Cyp1a	TTGCGTG	-197	EMSA (mAhR + kfARNT)	TCDD	[57]
	TCGCGTG	-197			
Rainbow Trout					
Cyp1a3	TGCGTGC	-451	*In vitro* Luc Assay (Hepa1c1c7)	TCDD	[57]
	TGCGTGC	-1357			
	TGCGTGC	-1771			
	TGCGTGT	-1775			

B.
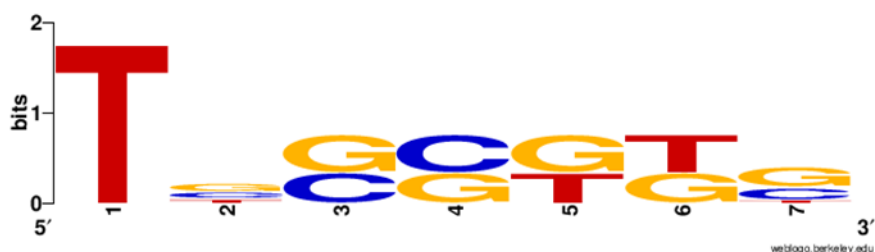
Figure 3.5. Validated AhR:DRE across fish species compared to the consensus sequence TNGCGTG. A). Validated binding sites were identified in cytochrome 1a1 (cyp1a1) in zebrafish, cytochrome 1a in Atlantic killifish and cytrochrome 1a3 in rainbow trout. Validated binding sites for the AhR matched the consensus sequence for all species. For luciferase reporter assays, constructs where used invitro in human cell lines using the fish-regulatory sequence. Sequences are also shown matching the complimentary consensus (5' CACGCNA 3'). B). Weblogo of validated AhR:DRE binding sites in fish species.

MTF1:MRE Binding sites
Consensus = 5' TGCRCNC 3'

Species	Sequence	Pos. (bp)	Method	Inducer(s)	Ref.
Zebrafish					
mt2	TGCACTC	-100	In vitro Luc Assay	Cd, Cu^{2+}, Hg, Zn	[58]
	TGCGCGC				
	TGCACTC				
	TGCACCC				
	TGCACAC	-740			
mt	TGCGCGC		In vitro Luc Assay	Cd, Cu^{2+}, Zn, H$_2$O$_2$	[58]
	TGCACTC	-			
	TGCACCC				
Rainbow Trout					
mta	TGCACAC		In vitro Luc Assay (RTG)	Ag^{2+} and Zn.	[59]
	TGCACAC	-118			
	TGCGCAC	-576			
	TGCACAC	-760			
mtb			Luc Assay (HepG2)		[60]
	TGCACCC	-62			
	TGCACAC	-89			
Pike		-599	In vitro Luc Assay (HepG2)	Zn, Cu^{2+}, Cd	[61]
mt	TGCACAC				
	TGCACAC				
	TGCACAC	-890			
	TGCGCGC				
Stone Loach					
mt	TGCACAC	-40		Zn, Cu^{2+}, Cd	[61]
		-140	In vitro Luc Assay (HepG2)		
	TGCACCC				
	TGCACTC	-140			
	TGCACTC	-500			
	TGCACCC				
	GCGTGCA				
	GTGTGCA	-720			

B.

Figure 3.6. Validated binding sites across teleost fish species compared to the consensus MRE, TGCRCNC. Validated binding sites were identified in metallothionein (*mt*) genes in clusters. Validated binding sites for the MRE matched the consensus sequence for all species and genes had multiple MRE sites. All sequences matched the consensus.

3.4.2 Differences in q-value across species for Nrf2:EpRE models:

For all random forest models for the Nrf2:EpRE motifs, the q-value provided the best predictor of classification. In all cases, q-values of known downstream target genes where significantly different from unknown targets as shown in Figure 3.7. Known targets were lower scoring than unknown targets and the level of significance was greatest for the mouse, human, zebrafish and stickleback. The difference in q-value was lowest for the medaka which showed a reduced significance between gene-sets. Q-values for returned hits for human and mouse where on average lower than across fish species. Only mammalian true positive hits fell within the 0.05 cut-off range across 3.5 kb promoter region (Figure 3.7a). Differences between the returned q-values between 3.5 kb and 5 kb searched sequences identified a reduced variability in the returned hits across all fugu and stickleback motifs. For zebrafish, q-values decreased for the MA0150.1 in 5 kb (Figure 3.7b) compared to 3.5 kb (Figure 3.7a).
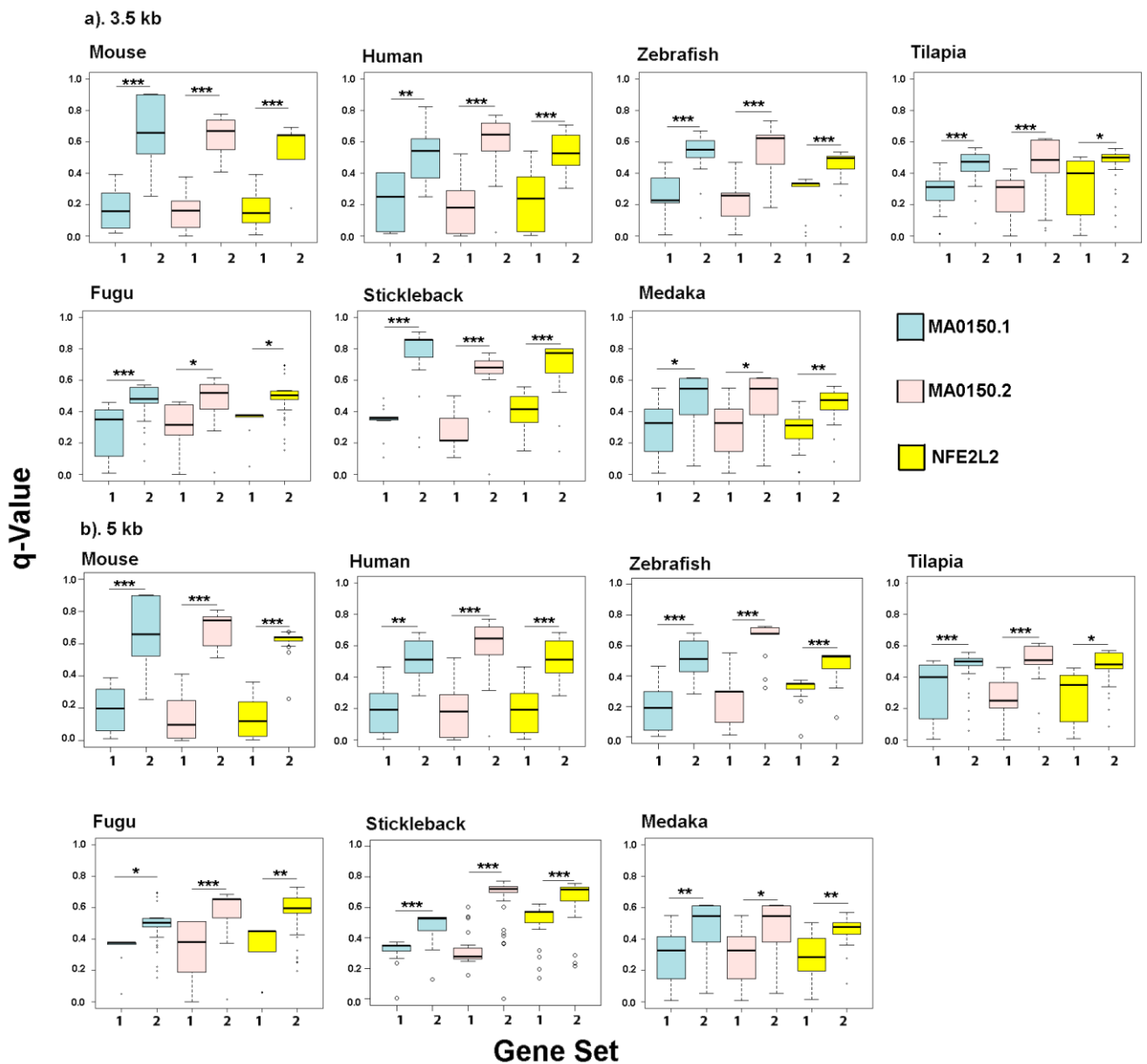
Figure 3.7. Distribution of q-value scores for returned Nrf2:EpRE matrices across species. Distribution of top-scoring q-values for each gene at a) 3.5 kb and b) 5 kb from TSS for the mouse, human, zebrafish, tilapia, fugu, stickleback and medaka following searches with NFE2L2, MA0150.1 and MA0150.2 PWM files in FIMO. Number 1 indicates results from the positive gene list (known gene targets) whereas 2 indicate the background targets (unknown gene targets). Colours indicate matrix file used as shown. Significance as calculated with either the Mann-Whitney U test or t-test where appropriate and indicated as 0.05*, 0.0001** and 0.00001***.

193

3.4.3 Random forest classification models for Nrf2:EpRE binding motifs:
Random forest models were first created on combined data consisting of top scoring q-values across all genes for mammalian and fish datasets respectively. These models were used to predict the performance of first, the ability of the model to predict the classification of the data used to create the model (e.g. for the mammalian model to predict the classification of the mammalian dataset) and secondly, the ability of the model to predict the classification of the remaining dataset (e.g. for the mammalian model to predict the classification of the fish dataset). In all cases, the random forest models had high accuracy values and high kappa values following LOOCV during model optimization with the q-value the greatest predictor of classification group (Table S3.8).

This research showed that overall, the models built on the mammalian datasets had a better classification performance across all motifs in comparison to those built using the fish dataset. For mammalian models, the classification score was high with an AUC score between 0.873 and 0.917 for mammalian predictions. The classification performance of the mammalian model predictions of the fish dataset where considerably lower with AUC scores between 0.701 and 0.768, identifying a reduced predictive performance across fish species (Figure 3.8a).

For models built on the combined fish-specific data, models where better able to predict the classification of mammalian datasets compared to the fish-specific datasets. In this case, the AUC values were between 0.813 - 0.843 for mammalian predictions, considerably lower than the predictions produced from the mammalian-trained models (Figure 3.8a). Fish-specific models also had a decreased performance when predicting the classification of the fish datasets with an AUC score between 0.709 and 0.768 across motifs (Figure 3.8b). In both cases, the NFE2L2 motif from HOCOMOCO was the least successful at predicting known-genes against unknown gene groups (Figure 3.8b).

An analysis of the species-specific classification performance using the mammalian model was conducted, identifying differences in the classification performance between species as shown in Figure 3.8c. In all cases, the classification performance was better for the species-specific predictions compared to the predictions of the combined dataset. The zebrafish dataset had the highest classification performance across all motifs with an AUC of 0.985 to 0.99, highlighting that the mammalian model was able to successfully classify zebrafish datasets into the known downstream and unknown downstream target groups.

The NFE2L2 motif had the lowest predictive performance across all species with the lowest AUC score of 0.796 for the tilapia. In all other cases, classification scores ranged from 0.92 to 0.98 across species and motifs, indicating that the models were reasonably successful at identifying gene groups. An exception to this was the medaka, which had an AUC score of 0.86 for MA0150.1 and the lowest AUC score of 0.92 for the MA0150.2 motif (Figure 3.6d). For the 5 kb search parameters, there was an overall decrease in the AUC score for all motifs across species (Figure 3.8d). The NFE2L2 motif had the lowest classification performance and the AUC score was lowest in the fugu at 0.62, which is close to a result for a random dataset. The zebrafish had high AUC scores across all motifs and models where successfully able to classify the dataset into known and unknown target gene groups. Overall, the MA0150.1 provided the most robust classification across the species-specific analysis, with AUC scores ranging from 0.92 to 1 (Figure 3.8b).

MA0150.1   MA0150.2   NFE2L2

a). Mammalian trained models

b). Fish trained models

c).

d).

Sensitivity

1 - Specificity

196
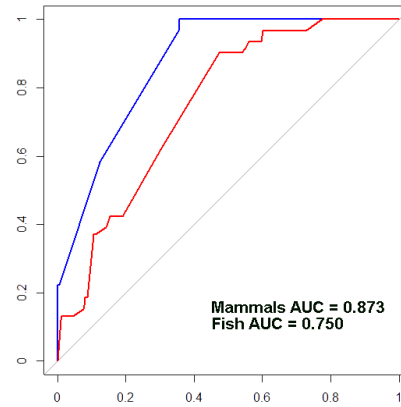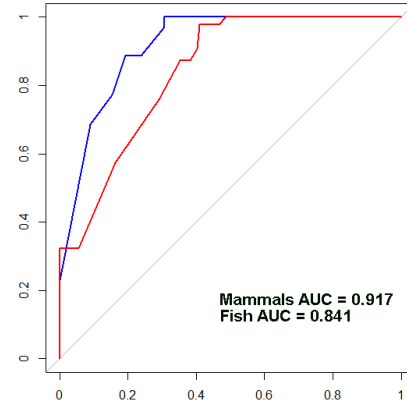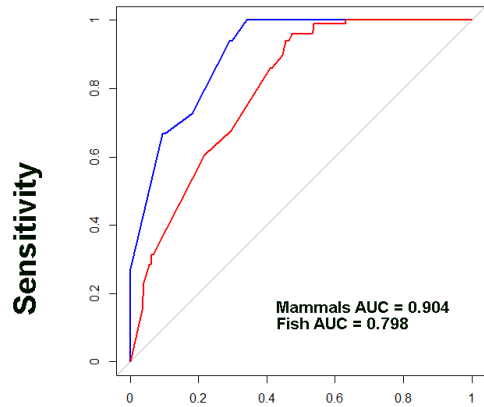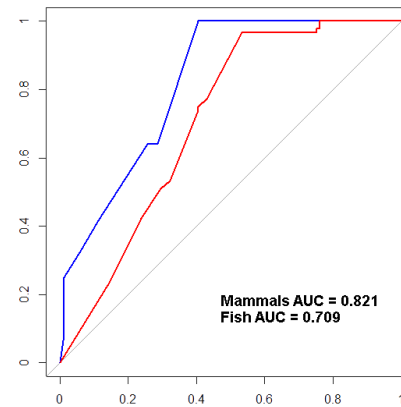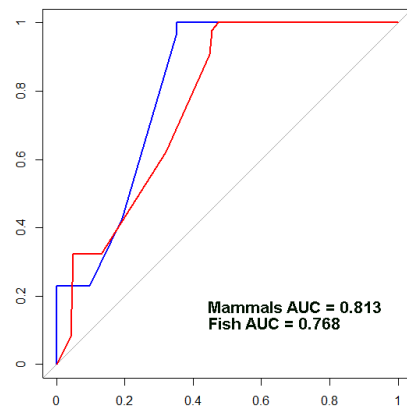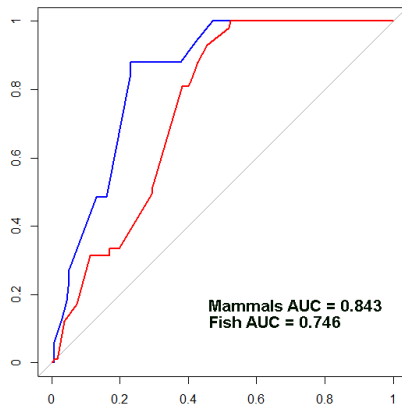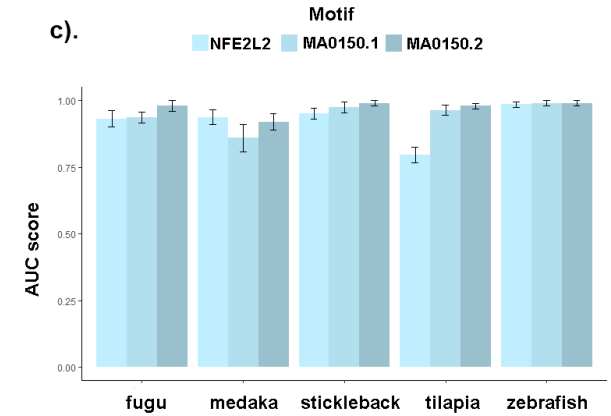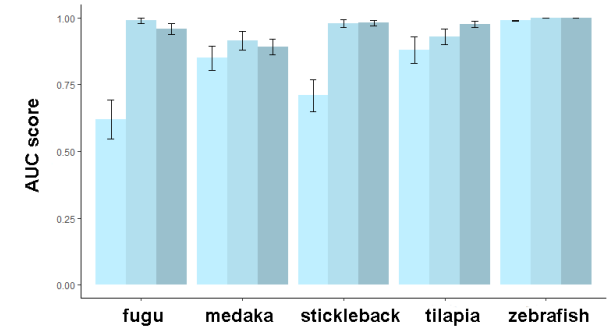
Figure 3.8. Random forest model classification scores for Nrf2:EpRE datasets following FIMO motif analysis using PWM motifs from the Jaspar database (MA0150.1, MA0150.2) and HOCOMOCO database (NFE2L2). a). ROC curves of Nrf2:EpRE RF models trained using the combined mammalian data at 3.5 kb. Models were used to classify datasets into known targets or unknown target gene groups from searches at 3.5 kb. b). ROC curves of Nrf2:EpRE RF models trained using the combined mammalian data. Models were used to classify datasets into known targets or unknown target gene groups. c). Bar-plot indicating the AUC values of the predictions using mammalian-trained models to classify species-specific gene sets at 3.5 kb. Error bars indicate the standard deviation. d). Bar-plot indicating the AUC values of the predictions using mammalian-trained models to classify species-specific gene sets at 5 kb. Error bars indicate the standard deviation.

3.4.4 Differences in q-value across HIF-1α:HRE models

Q-values for the returned highest scoring hits for HIF-1α showed instances where both the occurrence of hits in known targets and unknown targets was no different from random (Figure 3.9). This was the case for the MA01106.1 motif in humans and was also the case in the tilapia for MA01106.1 and MA0259.1 motifs. For mouse, the q-values across the 3.5 kb search showed a significant difference between known and unknown target gene groups across all motifs. In zebrafish, there was a significant difference in q-value between known and unknown gene targets for the matrix MA1106.1 and MA0259.1 but in the case of the latter, unknown targets had lower scoring q-values than known for the 3.5 kb search.

In fugu and stickleback there was a significant difference in between all scores with HIF1_si showing the lowest q-value on average. Across fish species, the greatest difference in q-value was between HIF1_si scores across fugu, medaka and tilapia showing the clearest difference between positive and negative groups. There was little difference in returned highest q-value scores between the 3.5 kb and 5 kb promoter regions across both groups.
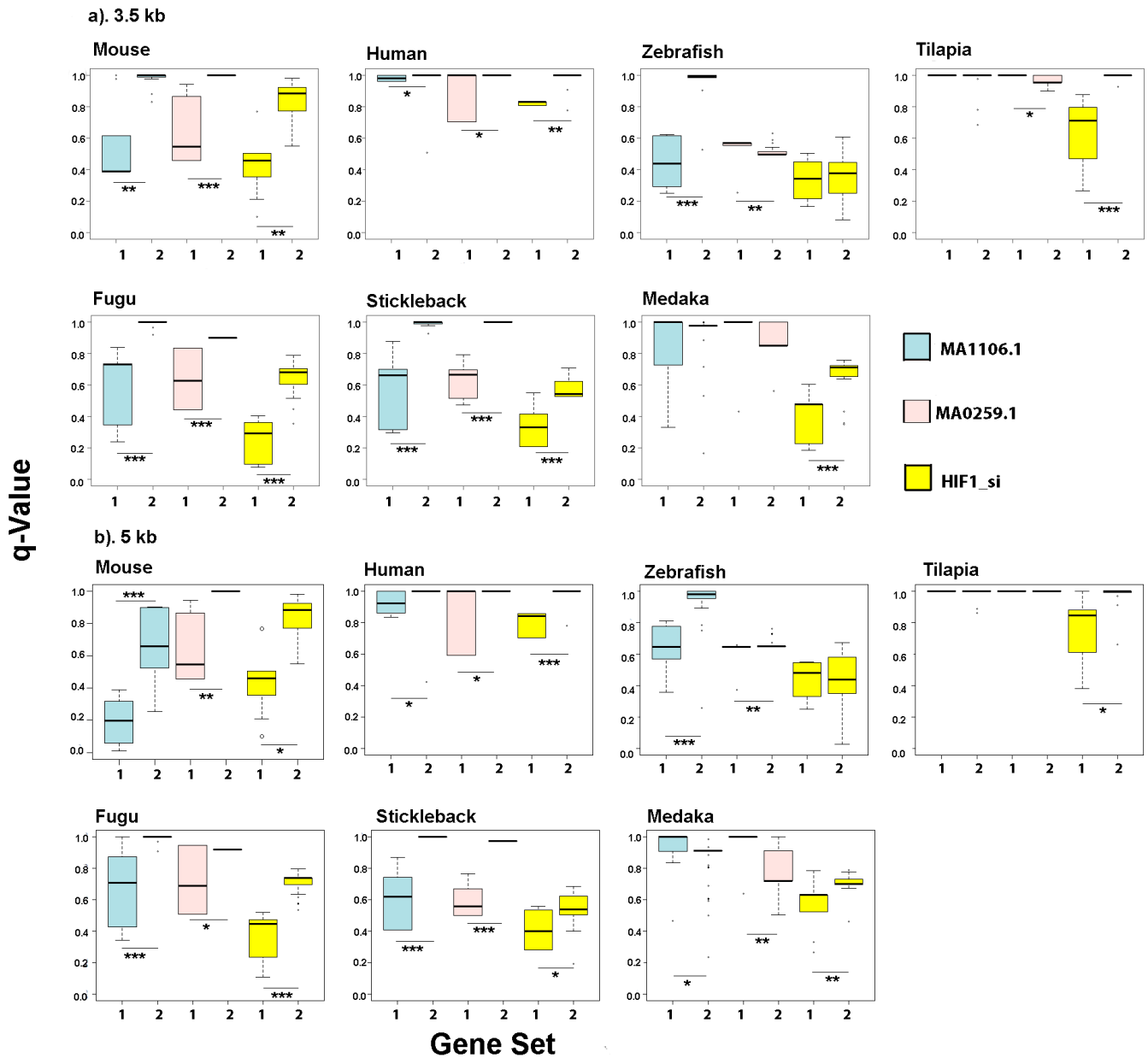
Figure 3.9. Distribution of q-value scores for returned HIF1:HRE matrixes from JASPAR (MA1106.1, MA0259.1) and HOCOMOCO (HIF1_si) across species. Q-value scores from A) 3.5 kb and B) 5 kb from TSS for the mouse, human, zebrafish, tilapia, fugu, stickleback and medaka following searches with MA1106.1, MA0259.1 and HIF1_si PWM files in FIMO. Number 1 indicates results from the positive gene list (known gene targets) whereas 2 indicates the q-value of background targets (unknown gene targets). Colours indicate the matrix file used as shown. Significance as calculated with either the Mann-Whitney U test or t-test where appropriate and indicated as 0.05*, 0.0001** and 0.00001***.

Models built on both the combined mammalian dataset and fish-specific dataset at 3.5 kb had a low classification performance across all motifs for HIF-1α (Figure 3.10). The q-value was the greatest indicator of classification group across models. In all instances, the AUC score was below 0.8 and in the case of HIF-1α_si, the AUC score of 0.59 to 0.69 for the mammalian-trained model shows a random classification performance. For the MA01106.1 motif, the AUC scores for the fish-specific model were greater than the mammalian model, between 0.71 and 0.78 for fish and mammals respectively compared to AUC scores 0.45 and 0.43 for the mammalian trained model. The motif MA0259.1 provided the best classification performance of any mammalian trained model at predicting the mammalian classification performance.

All models returned low kappa and accuracy values during model tuning in comparison to the Nrf2:EpRE models as shown in Table S3.6. In addition, as ROC curves identified low predictive performance across all mammalian trained models, species-specific analysis of predictive performance was not conducted in this case.
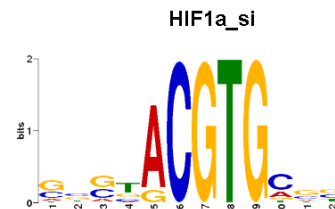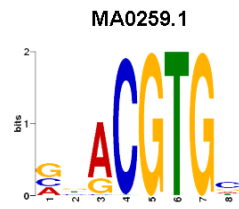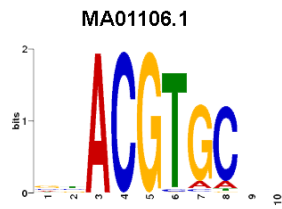
MA01106.1

MA0259.1

HIF1a_si

a.   **Mammalian trained models**

Mammals AUC = 0.45
Fish AUC =0.432

Mammals AUC = 0.708
Fish AUC = 0.598

Mammals AUC = 0.593
Fish AUC = 0.691

b.   **Fish trained models**

Mammals AUC = 0.778
Fish AUC = 0.718

Mammals AUC = 0.628
Fish AUC = 0.728

Mammals AUC = 0.613
Fish AUC = 0.581

Sensitivity

1 - Specificity

201
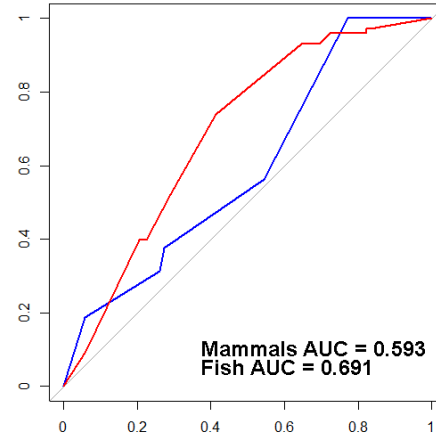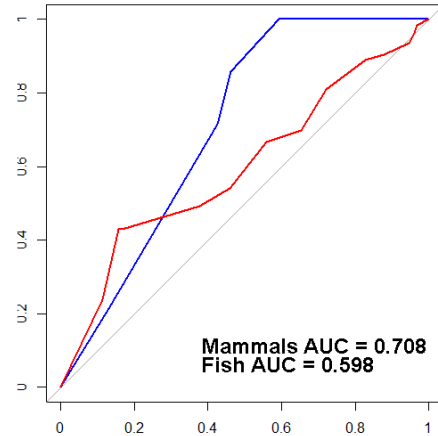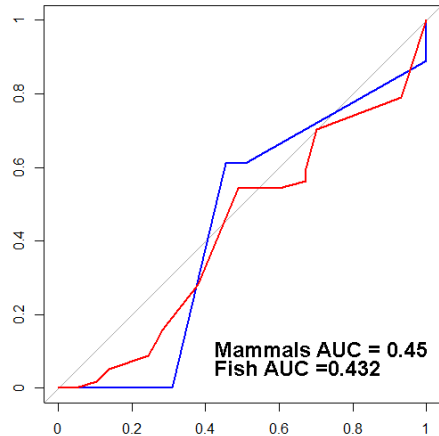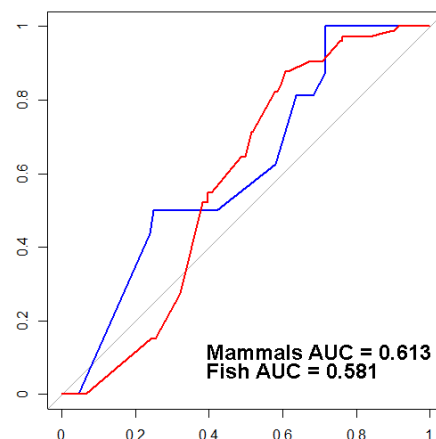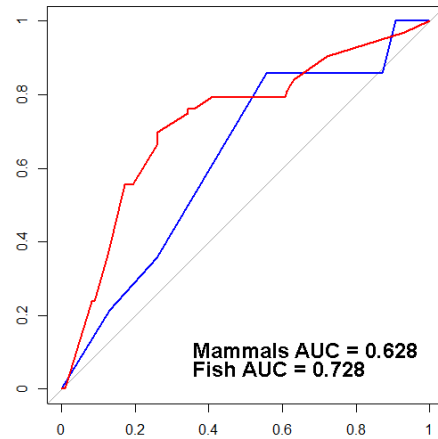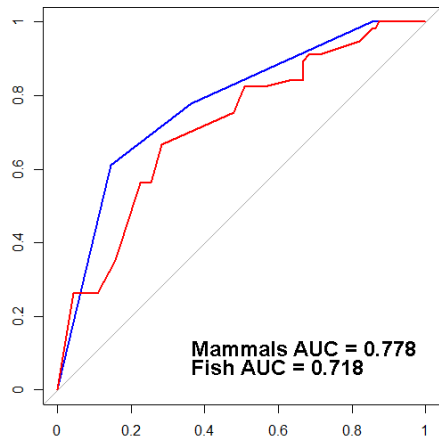
Figure 3.10. Random forest model classification scores for HIF-1α:HRE datasets following FIMO motif analysis. a). ROC curves of HIF-1α:HRE RF models trained using the combined mammalian data. Models were used to classify mammalian and fish-specific datasets into known targets or unknown target gene groups with the AUC score as shown as a measure of classification accuracy. Motifs used left to right as MA01106.1, Ma0259.1 and HIF1_si. b). ROC curves for HIF-1α:HRE RF models trained using the combined fish-specific dataset. Models were used to classify mammalian and fish-specific datasets into known targets or unknown target gene groups with the AUC score as shown as a measure of classification accuracy. Motifs used left to right as MA01106.1, MA0259.1 and HIF1_si.

3.4.5 Hmox1 and hsp70 promoter binding site identification

For hmox1, promoter regions were searched for Nrf2:EpRE, NFkB:kB and p53 binding sites across each species. Putative binding sites were identified for every transcription factor across species (Figure 3.11A). The identified binding sites are aligned in the supplementary information and shown in Figure S3.3.

There was little association with the functional binding sites and position across mammals to fish for the EpRE with TFBS identified near the TSS in mammals but were more distally located in fish (Figure 3.11A). Alignments of identified binding sites using the mammalian and fish-specific motifs showed that the majority matched the core consensus (TGACnnnGC) but there were some hits across both mammalian and fish species for the zebrafish-specific sequence (TGACnnnTC). The fish-specific motif identified hits in all fish species except stickleback where a sequence was identified using the mammalian motif (Figure S3.3). More hits were identified in humans, tilapia and fugu using the mammalian motif in comparison to the fish-specific motif. The same binding sites were identified for mammalian and fish searches in zebrafish and human regulatory regions.

For p53 motifs in hmox1, more hits were identified using the fish-specific motif in comparison to the mammalian motif (Figure 3.11a). Using the latter, high scoring hits were identified in the human and tilapia (Figure S3.3). Putative binding sequences were identified in the medaka and fugu using the fish-specific motif but not the mammalian motif and no sequences were identified in zebrafish (Figure 3.11A). Multiple NFkB sites were identified in all species except humans with the lowest scoring sequences identified in the fugu (Figure S3.3).

For hsp70, the promoter region was searched for Nrf2:EpRE, P53:PDB, HSF1:HSE, HIF1:HRE and AhR:DRE binding sites. EpRE binding sites were identified across human, mouse, zebrafish, tilapia, medaka and fugu but were not identified in the stickleback using the mammalian motif (Figure 3.10). However, sites were identified in the stickleback using the fish-specific motif and this matched the consensus TGACnnnGC sequence (Figure S3.4). The

medaka had the greatest number of putative binding sites and included hits that had the lowest q-values for the mammalian motif but only 2 hits using the fish-specific motif (Figure S3.4). The same binding sites were identified in medaka and tilapia using the mammalian and fish specific motifs (Figure 3.11B)

In *hsp70*, p53 binding sites were located after -3 kb from the TSS in all species except in humans and stickleback (Figure 3.11B). Searches for p53 binding sites using the mammalian motif identified sequences across human, mouse, tilapia and stickleback. Alignments of binding regions showed the strongest correlation between the stickleback and human sequence across fish species (Figure S3.4). The zebrafish p53 motif identified a greater number of p53 binding sites with hits in medaka and fugu. However, identified sequences in the human and mouse showed the greatest similarity to the motif file with q-values of 0.054 and 0.081 respectively (Figure S3.4).

Despite validation in mammals, few HIF-1α binding sites were identified using the mammalian motif with single hits shown in the mouse, medaka and fugu. In contrast, AhR:DRE sites were not identified using the mammalian motif but hits were found using the fish specific motif in the zebrafish, medaka and tilapia (Figure 3.11B).

hsp70 is traditionally associated with being regulated by HSF1 and multiple HSF1 binding sites were identified across all species investigated. All species had HSF1 binding sites between -3.5 and -5kb of the promoter regions and multiple sequences were identified in human, medaka, tilapia, zebrafish and stickleback. Alignments of binding sequences showed the most variation from the mammalian hits existed in the tilapia and these identified sequences subsequently had the lowest q-value score (Figure S3.5).
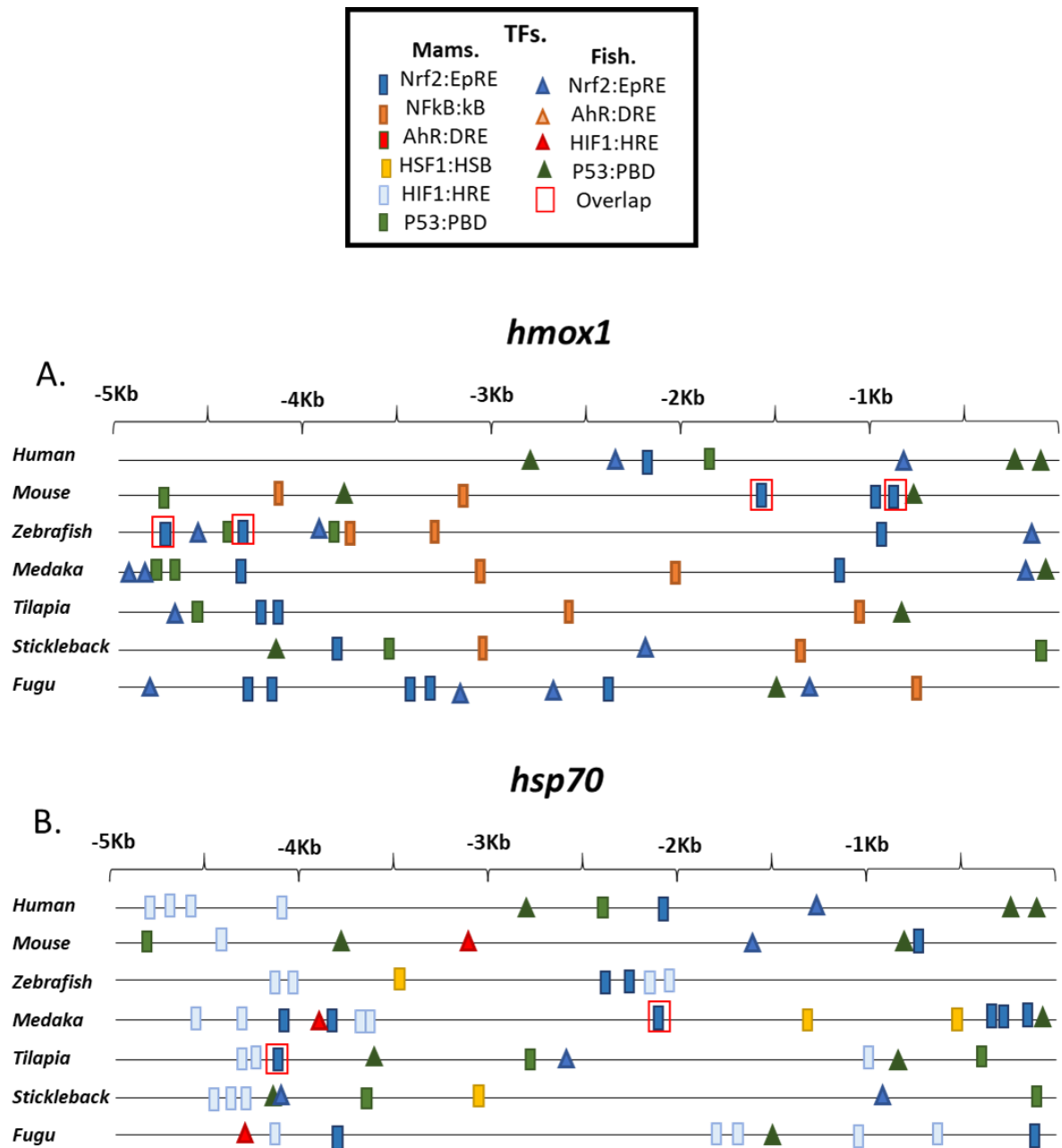
Figure 3.11. Promoter analysis for a). hmox1 and b). hsp70: Schematic representation of binding sites identified in the hmox1 and hsp70 promoter regions following 5 kb searches upstream of the transcription start site (TSS). Hmox1 was searched for binding sites for P53, Nrf2 and NFkB. Hsp70 was searched for sites for HIF-1α, P53, AhR, HSF1 and p53 in accordance with validated sites in mammalian studies. Promoter regions were searched using PWMs from JASPAR database. Overlap indicates sites that were shared between mammalian and fish-specific searches.

3.5 Discussion

3.5.1 Validated binding sites across fish species:

The literature search identified that validated binding sites across fish species matched the consensus sequence in the majority of cases. Sequences searches for AhR:DRE and MTF1:MRE showed no variation from the consensus sequence suggesting a strong evolutionary conservation of these sites across vertebrates. In addition, for MTF1, sites were shown to be necessary in clusters, a phenomenon conserved in mammals[63]. All identified sequences for these TFs were responsive to inducers under luciferase reporter assays [56-61], providing strong evidence that sequences matching the mammalian consensus were functional sites.

Validated binding sequences differed from the consensus in Nrf2:EpRE and HIF-1α:HRE TFBS in zebrafish as well as in the *epo* gene in the Atlantic Killifish [55]. However, across other fish species, functional sites were identified matching the consensus mammalian sequence which suggests the motifs are widely conserved. However, as the evolutionary distance and diversity across teleost fish species is great, sharing an estimated ~350 million years of evolutionary history [39], and as TFBS evolve faster than TFs [64], there is a strong likelihood of species-specific divergences in TFBS composition. As whole genome studies for TF binding have yet to be widely performed across fish-species for adaptive stress response factors, the levels of species divergence in binding sites are not yet known. Identified variations in binding sequence were most prevalent in the zebrafish in this research but this may be biased as the zebrafish is the most widely used model fish species and there is therefore far more information available for this species in comparison to overs.

In addition, as the identified studies for MTF1:MRE and AhR:DRE were searched using the consensus sequence, this could prevent any variations in binding motif from being identified [52]. PWMs alignments of the functional binding sites also showed little informational content in comparison to mammalian motifs with little variation from the consensus in validated sites. Information such as flanking regions to the consensus and the frequency of variations in base pairs can be identified from large amounts of information (e.g. 10+ validated sites). Changes in frequency of base composition is important

considering that transcriptional output can be altered by the strength of TF-DNA interactions and therefore has evolutionary consequences[65].

3.5.2 Threshold specificity altered across distantly related species:
Given that there were no differences between validated sites and mammalian consensus for MTF1:MRE and AHR:DRE, further analysis of binding motifs focused on the Nrf2:EpRE and HIF-1α:HRE PWMs only. The results from the FIMO analysis showed that the threshold q-value, where an arbitrary threshold score between 0.01-0.05 is recommended [43], was too stringent to identify most known downstream target genes in fish species.

Only scores for human and mouse hits for Nrf2:EpRE showed values of known downstream targets within this range, though this was not significantly different from the unknown target gene set. For fish species, there was an overall difference in the returned q-value between known and unknown downstream target genes across Nrf2:EpRE hits, with known targets on average having a lower q-value than their mammalian counterparts.
Furthermore, no known downstream targets for HIF-1α:HRE genes had binding sites within 3.5 kb and 5 kb of the promoter region that were within the 0.01-0.05 q-value threshold with the exception of the mouse MA1106.1 results at 5 kb search, identifying that additional parameters are needed to identify true candidate sequences in the case of HIF-1α:HRE.

However, it should be noted that the q-value only determines how well an identified sequence matches the matrix file; it is a measure of confidence but not necessarily of function. The difference in q-value between known and unknown targets reflects a difference in the overall sequence composition of identified binding sites and this suggests that there could be TFBD in distantly related species to mammals such as stickleback, where there was a reduced difference between q-value scores in known and unknown targets. However, incomplete genome annotations for the stickleback, tilapia and fugu may have caused a bias in the results; the q-value used in FIMO as a measure of accuracy is highly sensitive to the background GC% content and due to incomplete genomic information for these species, it is likely that not all gene promoters were compiled to make the background GC% files in this case.

3.5.3 Variation in classification performance across species:

Random forest models were built to assess whether there was both a variation between motifs at predicting known downstream targets and a variation in classification performance across mammalian and fish species. For all random forest models, the q-value was the best predictor of whether a hit belonged to a downstream target or a gene in the background gene set. Parameters including enrichment, strand and p-value had a less significant effect on model performance, with predictions of true and false targets largely defined by the q-value.

For the Nrf2:EpRE datasets, all models constructed on the results from the 3.5 kb FIMO analysis where able to successfully predict the classification of mammalian datasets with a high accuracy rate as shown by the AUC value. Likewise, all models were able to classify fish-specific gene sets but with a lower accuracy than for predictions using the mammalian dataset. Variation in the predictive scores and therefore the identified sequences is suggested across the known and unknown downstream target groups in fish, reflected by both a lower AUC value in the mammalian-trained models and a reduced classification score on fish-trained models. This research showed clear differences in the predictive performance of motif sequences, with MA0150.2 having the highest AUC value across both mammals and fish when compared to the mammalian-trained model. The NFE2L2 motif, from the HOCOMOCO database, had the lowest performance despite being generated from multiple matrix files.

Predicting the classification for unknown and known targets of the species-specific datasets produced AUC values higher than for the combined fish-dataset alone. However, the fact that fewer data points are used for this prediction could have caused some level of over-fitting of the test dataset to the training data but this is considered to be largely avoidable in random forest models, particularly where cross-validation techniques are used. In addition, as the q-value was the greatest predictor of performance across all models, variations in the q-value between species could have caused an overall reduced performance of the combined dataset.

This analysis identified some variation in the species-specific predictive performance, notably for the medaka and stickleback at 3.5 kb search parameters and for fugu, tilapia and stickleback at 5 kb. Despite the variations in binding site composition identified in the zebrafish, Nrf2:EpRE motifs successfully classified the zebrafish dataset into known and unknown target gene groups. The zebrafish is the most evolutionary distinct of the teleost investigated (Figure S3.1), having the most distant common ancestor with the other teleost fish species studied, which may partly explain the differences in performance across the modelling results [66].

In the case of HIF-1α, the absence of a difference between q-value across known downstream targets and unknown target gene groups affected the ability to build successful classification models for both mammalian and fish datasets. Random forest models were less able to characterise genes into known and unknown gene target groups with less significant differences in q-value returned in input sequences. This difference in classification performance may in part be due to the shorter sequence length compared the to EpRE motifs which could cause a higher frequency of negative hits within the genomic background. This research was based on the assumption that the binding sites for each factor would be located with 3.5 kb and 5 kb of the transcription start site (TSS) and whilst HIF-1α binding sites have been identified within this range [7], it is possible that in some gene targets, the regulatory binding sites are more distally located.

Additional parameters used within the model, such as protein conservation, could further aid to the efficiency of the predictive platform. The study was also limited to the number of genes selected for the analysis and increasing the amount of information used within the model would significantly improve the accuracy of the results. In addition, this search only focused on the 3.5 kb and 5 kb from the TSS, excluding distal and intragenic response elements that have roles in gene expression. Consequently, there is a chance that binding sites in known targets could be missed although most sites are within the same position as mammalian regions.

3.5.4 Putative binding sites were identified in shared downstream targets in the adaptive stress response pathway:

Based on the results of this research, binding sites were identified in the target genes *hmox1* and *hsp70*, which are regulated by multiple adaptive stress response factors in mammals. This analysis identified that sites were conserved across species using both mammalian and fish motifs.

Hmox1 is a well-documented target of Nrf2 and had a large number of hits for EpRE regions across fish-species in this study using both mammalian and fish-specific binding motifs. This fish-specific EpRE motif increased the number of identified sequences across fish species in combination with the mammalian motif but there were some cases of overlap between hits in the zebrafish and mouse. As a good indicator of functional sites as shown in this study, being the main predictor of functional genes against background gene sets, the EpRE from the JASPAR database, MA0150.2, was used to search regulatory regions. For these hits, the q-value was low (<0.05) in identified hits across mammals and fish species using the mammalian motif. This shows that sites could be identified in traditional promoter analysis techniques if less stringent threshold scores, necessary to reduce false positive rates, are used. However, the fish EpRE motif did identify binding sequences in regions that were not found using the mammalian motif suggesting some sites could be missed based on mammalian motifs alone. This was also the case for fish specific p53 motif, which identified different sites in promoter regions and little overlap with mammalian motif search results.

Hsp70 is targeted by multiple TFs. The AhR and HIF-1α have been demonstrated to form a binding complex [67] and variations were observed in the number of HIF-1α and AhR TFBS across each species. HIF-1α and AhR are thought to form a complex and the TFBS for these factors in mammals is similar (ending in GTG). Whilst there were few hits for AhR or HIF-1α sites using both mammalian and fish-specific matrices, there is a possibility that the dimer formed from AhR-HIF-1α interactions binds to a variation of this binding region. Hits for AhR and HIF-1α were few and low scoring (based on the q-value) and

this also suggests that binding regions could be outside of the 5 kb search parameters used in this analysis.

Though the extent of conservation of NFkB sites was not researched in this study, kB sites were identified across species for hmox1 using mammalian motifs. Clusters of kB sites are associated with functional gene regulation with the number of occupied sites regulating the levels of gene expression [35]. NFkB sites were not identified in distinct clusters in this research across species. As a transgenic zebrafish model for NFkB showed predicted responses to inflammation caused by exposure to microbial colonization, suggesting a level of conservation with the mammalian motif [68].

There are few validated sites for HSF1 across fish species, which has a short consensus sequence (nGAAn) but expected target genes have been shown to have putative binding regions in gene promoters [50,69]. Using the mammalian motif from the Jaspar database, HSE were enriched in the *hsp70* gene promoters across species, often located distally between -4 and -5 kb. As hsp70 is a well-known target of HSF1 in mammals and as heat-shock responses and HSF1 is conserved in fish species [50], this is a good indicator that atleast some of these sites are functional.

The results from the hmox1 and hsp70 promoter analysis support the use of mammalian motifs, but with less stringent threshold scores. It should be noted that fish-specific motifs identified different numbers of TFBS in fish species compared to the mammalian motifs. However, as these are formed from only a limited number of binding sequences (<10 sites from the literature search) and were largely composed of the core consensus sequences, there is little informational content in the fish-specific PWM in comparison to the mammalian motifs. This includes limited information on flank coding regions as well as variations between bases which leads to altered levels in the q-value scores, as a measure of sequence comparison to the consensus. In addition, enriched motifs were only identified for HSE in hsp70 and for EpRE in Hmox1, which are the traditional regulators of these genes. Enrichment, largely associated with functional binding sites, did not improve the predictive performance of random forest models despite being a good indicator of functional sites. It is possible

that enrichment of TFBS is dependent on the TF in question as well as the length of the motif file and consensus sequence being searched.

3.5.5 Further research:

This research identified that whilst there are variations in TFBS across fish species, mammalian derived binding matrixes can provide good predictions of putative TFBS in distantly related species. However, there is still a need to validate binding regions across fish species to identify the levels and frequency of divergence across the genome. This would allow the ability to build more accurate identification models based on stronger species-specific empirical information and would indicate if there is a selective pressure on binding sequence composition between species. The identification of robust TFBS is challenging in non-model organisms, particularly as ChIP-Seq/ATAC-seq are reliant on mammalian antibodies that do not always function effectively across vertebrates[70]. However, the results from this analysis, particularly for Nrf2 and HIF-1α, demonstrate the advantages of using fish-specific motifs in the context of identifying target genes across greater evolutionary ranges where divergence and informational content can change the likelihood of identifying target genes.

The regulation of genes through distal enhancer elements where non-coding regions contain clusters of binding motifs associated with multiple factors[71] could increase the predictive performance of TFBS identification by combining motif searches for genes that are known to be shared targets for multiple factors. In addition, predictive models that consider the DNA-shape and TF binding have been shown to produce more accurate predictions of functional TFBS [72]. Models that incorporate data on chromatin composition along with PWM searches have also improved predictions in mammals[73]. Where the data is available, these methods should be combined with the current motif prediction techniques using PWM in a fish-specific context.

In addition, although MTF1 had multiple binding sequences, the performance of this motif may be more obvious in comparison to mammalian PWMs. However, as few known target genes have been efficiently annotated across species to conduct this analysis, this was not conducted in this research.

3.6 Conclusion

The initial analysis of binding regions showed some differences in fish-specific sequences. Using random forest models to assess binding site conservation showed that mammalian derived PWMs could classify downstream target genes across fish species but with a lower efincency than mammalian gene sets for Nrf2:EpRE. However, for HIF-1α, data provided from motif hits was insufficient to classify gene sets into known and unknown target gene groups. The research showed that whilst threshold parameters are a widely used in *in silico* identification methods, less-stringent values were necessary to identify putative regulatory regions in genes which are highly likely downstream targets of the TFs analyzed. Fish-specific binding motifs identified novel putative binding sites in comparison to mammalian models. This research highlights that different information can be obtained using fish-speciifc and mammalian motifs, though in general, binding sites show strong conservation with mammalian consensus sequences. More research is necessary to validated putative hits to establish the significance of these findings.

Chapter 3: A critical analysis of in silico methods used for identifying transcription factor binding sites across fish species

3.7 Supplementary Information

This supplementary information contains:

Figure S3.1: Evolutionary Tree of relatedness between fish and mammalian species.

Figure S3.2: PWM for adaptive stress response factors.

Table S3.1: Gene ids for validated Nrf2:EpRE targets.

Table S3.2: Gene ids for known target genes of HIF-1α:HRE

Table S3.3: Background ensemble gene ids comprising of unknown targets of Nrf2:EpRE and HIF-1α:HRE

Table S3.4: 0-order Markov Model of whole genome promoter sequences.

Table S3.5: Model selection results for 5kb mammalian datasets.

Table S3.6: Random forest model parameter for Nrf2:EpRE models

Table S3.7: Random forest model parameters for HIF-1α:HRE based models:

Table S3.8: Variable importance in Nrf2:EpRE random forest models.

Table S3.9: Variable importance in HIF-1α:HRE random forest models.

Figure S3.3: Alignments of identified binding sites in the hmox1 promoter

Figure S3.4: Alignments of identified binding sites in the hsp70 promoter for the EpRE and p53.

Figure S3.5: Alignments of identified binding sites in the hsp70 promoter for the HSE.

Figure S3.1: Evolutionary trees of species relatedness and transcription factor conservation. A). Evolutionary tree of species relatedness as adapted from Volff. (2005) [39].

A.

Homo sapiens
Mus musculus
Danio rerio
Oryzias latipes
Gasterosteus aculeatus
Takifugu rubripes
Orechromis niloticus
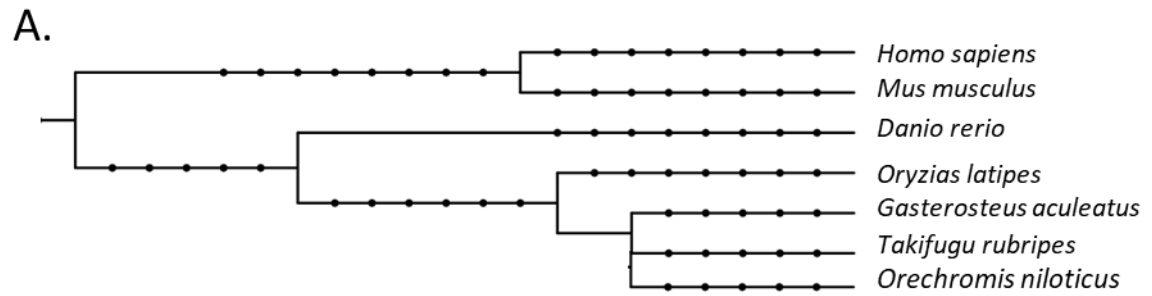
Figure S3.2: PWM for adaptive stress response factors. PWM from the JASPAR database for AhR:DRE, NFkB:kB, HSF1:HSE, p53 and MTF1. AhR, NFkB, HSF1 and p53 motifs were used to search promoter regions in hmox1 and hsp70.
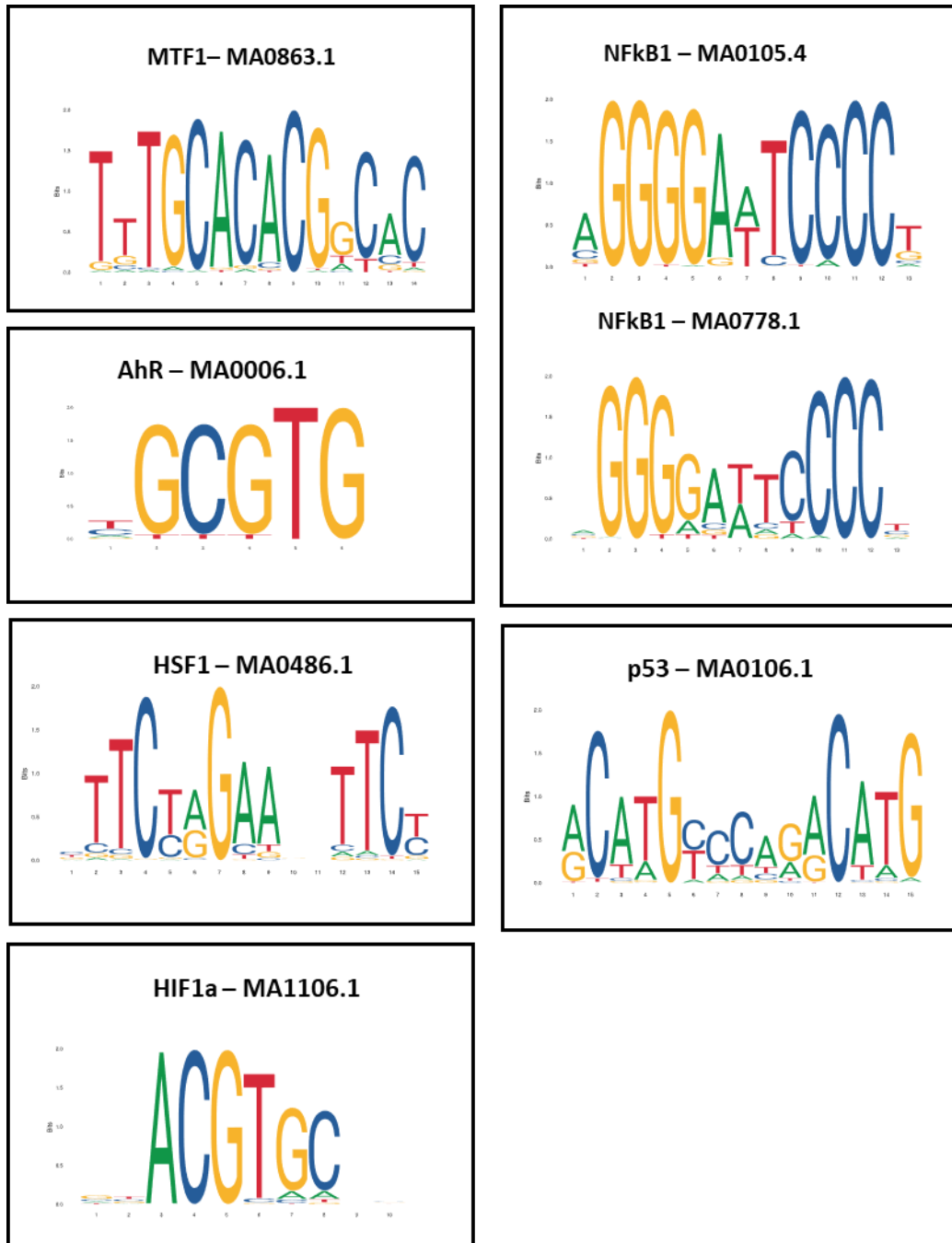
Table S3.1: Gene ids for validated Nrf2:EpRE targets: Downstream gene targets for Nrf2 were identified from the literature. In each case, ensemble gene ids were selected for humans and the gene orthologs were derived from the Ensembl database. Paralogs for all genes were used in the subsequent analysis.

Gene Name	Human	Mouse	Medaka	Stickleback	Tilapia	Zebrafish	Fugu
UGDH	ENSG00000109814	ENSMUSG00000029201	ENSORLG00000005642	ENSGACG00000016178	ENSONIG00000008996	ENSDARG00000019838	ENSTRUG00000009509
NFE2L2	ENSG00000116044	ENSMUSG00000015839	ENSORLG00000017568	ENSGACG00000004470	ENSONIG00000008604	ENSDARG00000042824	ENSTRUG00000017486
PRDX1	ENSG00000117450	ENSMUSG00000028691	ENSORLG00000010533	ENSGACG00000016172	ENSONIG00000018611	ENSDARG00000058734	ENSTRUG00000012465
SQSTM1	ENSG00000161011	ENSMUSG00000015837	ENSORLG00000020572	ENSGACG00000016271	ENSONIG00000000103	ENSDARG00000075014	ENSTRUG00000017345
HMOX1	ENSG00000100292	ENSMUSG00000005413	ENSORLG00000010457	ENSGACG00000006771	ENSONIG00000018076	ENSDARG00000027529	ENSTRUG00000004710
		ENSMUSG00000005413		ENSGACG00000019469	ENSONIG00000008232	ENSDARG00000042533 (gstm1)	
						ENSDARG00000088116 (gstm3)	
GSTM1	ENSG00000134184	ENSMUSG00000058135	ENSORLG00000005927	ENSGACG00000007655	ENSONIG00000002477	ENSDARG00000029473	ENSTRUG00000006620
			ENSORLG00000005961	ENSGACG00000007674			ENSTRUG00000001751
SRXN1	ENSG00000271303	ENSMUSG00000032802	ENSORLG00000015074	ENSGACG00000011801	ENSONIG00000016757	ENSDARG00000079160	ENSTRUG00000007666
GCLC	ENSG00000001084	ENSMUSG00000032350	ENSORLG00000009800	ENSGACG00000006384	ENSONIG00000009700	ENSDARG00000013095	ENSTRUG00000015144
ABCC2	ENSG00000023839	ENSMUSG00000025194	ENSORLG00000008249	ENSGACG00000007419	ENSONIG00000018322	ENSDARG00000014031	ENSTRUG00000009329
					ENSONIG00000009455		
GSTA1	ENSG00000243955	ENSMUSG00000057933	ENSORLG00000009674	ENSGACG00000006489	ENSONIG00000009685	ENSDARG00000039832	ENSTRUG00000014918
		ENSMUSG00000074183			ENSONIG00000009676	ENSDARG00000090228	
		ENSMUSG00000074179					
		ENSMUSG00000111709					
AHR	ENSG00000106546	ENSMUSG00000019256	ENSORLG00000000135	ENSGACG00000008466	ENSONIG00000012216	ENSDARG00000020046	ENSTRUG00000018462
		ENSMUSG00000019256	ENSORLG00000018196	ENSGACG00000015617	ENSONIG00000010232		ENSTRUG00000009303
							ENSTRUG00000000070
KEAP1	ENSG00000079999	ENSMUSG00000003308	ENSORLG00000003823	ENSGACG00000019479	ENSONIG00000018066	ENSDARG00000074634	ENSTRUG00000003855
GCLM	ENSG00000023909	ENSMUSG00000028124	ENSORLG00000017101	ENSGACG00000011182	ENSONIG00000002354	ENSDARG00000018953	ENSTRUG00000000742
SOD1	ENSG00000142168	ENSMUSG00000022982	ENSORLG00000004389	ENSGACG00000020581	ENSONIG00000004690	ENSDARG00000043848	ENSTRUG00000008179
FTH1	ENSG00000167996	ENSMUSG00000024661	ENSORLG00000006156	ENSGACG00000015484	ENSONIG00000002989	ENSDARG00000015551	ENSTRUG00000017266
			ENSORLG00000005872		ENSONIG00000015578	ENSDARG00000007975	
NQO1	ENSG00000181019	ENSMUSG00000003849	ENSORLG00000017876	ENSGACG00000003764	ENSONIG00000015892		ENSTRUG00000010839
		ENSMUSG00000003849	ENSORLG00000007464	ENSGACG00000003770	ENSONIG00000003376	ENSDARG00000010250	ENSTRUG00000011870
				ENSGACG00000003783			
				ENSGACG00000003776			

217

			ENSGACG00 000003788			
			ENSGACG00 000003762			
			ENSGACG00 000003754			
			ENSGACG00 000003769			
			ENSGACG00 000003752			
GAPDH	ENSG0000 0111640	ENSMUSG00 000098456	ENSORLG00 000012224	ENSGACG00 000010219	ENSONIG000 00012916	ENSDARG000000 43457	ENSTRUG00 000003708
		ENSMUSG00 000081607					
CAT	ENSG0000 0121691	ENSMUSG00 000027187	ENSORLG00 000001746	ENSGACG00 000003491	ENSONIG000 00012865	ENSDARG000001 04702	ENSTRUG00 000016160
GPX1	ENSG0000 0233276	ENSMUSG00 000063856	ENSORLG00 000000823	ENSGACG00 000010455	ENSONIG000 00019002	ENSDARG000000 18146	ENSTRUG00 000008247
			ENSORLG00 000010182	ENSGACG00 000000882		ENSDARG000000 06207	ENSTRUG00 000004998
MGST3	ENSG0000 0143198	ENSMUSG00 000026688		ENSGACG00 000016252	ENSONIG000 00018637	ENSDARG000001 02744	ENSTRUG00 000000749

Table S3.2: Gene ids for known target genes of HIF-1α:HRE: Downstream gene targets for HIF-1α were identified from the literature. In each case, ensemble gene ids were selected for humans and the gene orthologs were derived from the Ensembl database. Paralogs for all genes were used in the subsequent analysis.

Gene Name	Human	Mouse	Zebrafish	Fugu	Tilapia	stickleback	Medaka
EPO	ENSG00000130427	ENSMUSG00000029711	ENSDARG00000055163	ENSTRUG00000003617	ENSONIG00000019900	ENSGACG00000019930	ENSORLG00000003612
	ENSG00000112715	ENSMUSG00000023951	ENSDARG00000034700	ENSTRUG00000008580		ENSGACG00000014012	
			ENSDARG00000103542	ENSTRUG00000015230		ENSGACG00000013374	ENSORLG00000016887
EGF	ENSG00000138798	ENSMUSG00000028017	ENSDARG00000052739	ENSTRUG00000015626	ENSONIG00000001112	ENSGACG00000016495	ENSORLG00000020598
CHD2	ENSG00000173575	ENSMUSG00000020826	ENSDARG00000060687	ENSTRUG00000008842	ENSONIG00000015136	ENSGACG00000013246	ENSORLG00000008403
SNAPC1	ENSG00000023608	ENSMUSG00000021113	ENSDARG00000029569	ENSTRUG00000010116	ENSONIG00000001477	ENSGACG00000010844	ENSORLG00000018054
			ENSDARG00000003827		ENSONIG00000001765	ENSGACG00000016750	ENSONIG00000001765
FOS	ENSG00000170345	ENSMUSG00000021250	ENSDARG00000040135	ENSTRUG00000010431	ENSONIG00000015264	ENSGACG00000007617	ENSORLG00000017504
			ENSDARG00000031683	ENSTRUG00000018314	ENSONIG00000020185	ENSGACG00000010481	ENSORLG00000017504
SLC2A3	ENSG00000059804	ENSMUSG00000003153	ENSDARG00000013295	ENSTRUG00000007710	ENSONIG00000012832	ENSGACG00000001994	ENSORLG00000011838
				ENSTRUG00000005156	ENSONIG00000007021	ENSGACG00000010483	ENSORLG00000006093
NDRG1	ENSG00000104419		ENSDARG00000032849	ENSTRUG00000010966	ENSORLG00000004785	ENSGACG00000005968	ENSONIG00000006850
			ENSDARG00000010420	ENSTRUG00000010413	ENSORLG00000003558	ENSGACG00000006984	ENSONIG00000002093
PKM2	ENSG00000067225	ENSMUSG00000032294	ENSDARG00000099730	ENSTRUG00000004651	ENSORLG00000013279	ENSGACG00000016600	ENSONIG00000002725
CGA	ENSG00000135346	ENSMUSG00000028298	ENSDARG00000040479	ENSTRUG00000005818		ENSGACG00000009153	ENSORLG00000014810
GPI	ENSG00000105220	ENSMUSG00000036427	ENSDARG00000012987	ENSTRUG00000015872	ENSONIG00000018246	ENSGACG00000007744	ENSORLG00000000249
			ENSDARG00000103826	ENSTRUG00000001879	ENSONIG00000002874	ENSGACG00000014922	ENSORLG00000014082
				ENSTRUG00000001618	ENSONIG00000007356		
BCL2L	ENSG00000171791	ENSMUSG00000003190	ENSDARG00000008434	ENSTRUG00000005363	ENSONIG00000009766	ENSGACG00000013368	ENSORLG00000016791
					ENSONIG00000017372		
ZFAT1	ENSG00000066827	ENSMUSG00000022335		ENSTRUG00000017967			ENSONIG00000001807
				ENSTRUG00000010724	ENSONIG00000007011	ENSGACG00000006955	ENSORLG00000004805
				ENSTRUG00000012678	ENSONIG00000008419	ENSGACG00000006393	ENSORLG00000004943
					ENSONIG00000002725	ENSGACG00000016600	
NARF	ENSG00000141562	ENSMUSG00000000056	ENSDARG00000024184	ENSTRUG00000008423	ENSONIG00000019672	ENSGACG00000005244	ENSORLG00000019873
GPR37L1	ENSG00000170075	ENSMUSG00000026424	ENSDARG00000006079	ENSTRUG00000013954	ENSONIG00000014671	ENSGACG00000000278	ENSORLG00000002758
				ENSTRUG00000009206	ENSONIG00000000494	ENSGACG00000004094	ENSORLG00000018276

SEMA4B	ENSG0000 0185033	ENSMUSG000 00030539	ENSDARG000 00074414	ENSTRUG000 00002328	ENSONIG000 00002592	ENSGACG000 00010807	ENSORLG000 00012629
			ENSDARG000 00062352	ENSTRUG000 00006523	ENSONIG000 00015171	ENSGACG000 00016484	ENSORLG000 00008247
PDK1	ENSG0000 0152256	ENSMUSG000 00006494	ENSDARG000 00013128	ENSTRUG000 00017692	ENSONIG000 00008703	ENSGACG000 00004843	ENSORLG000 00017416
SCL2A	ENSG0000 0117394	ENSMUSG000 00028645	ENSDARG000 00007412	ENSTRUG000 00002051		ENSGACG000 00012574	ENSORLG000 00019077
			ENSDARG000 00001437	ENSTRUG000 00002157		ENSGACG000 00004850	ENSORLG000 00007473
				ENSTRUG000 00001412			
MMP2	ENSG0000 0087245	ENSMUSG000 00031740	ENSDARG000 00017676	ENSTRUG000 00010165	ENSONIG000 00002793	ENSGACG000 00016670	ENSORLG000 00013688
BNIP3	ENSG0000 0176171	ENSMUSG000 00078566		ENSTRUG000 00006491		ENSGACG000 00009527	ENSORLG000 00005914
		ENSMUSG000 00078566	ENSDARG000 00022832	ENSTRUG000 00006491	ENSONIG000 00000438	ENSGACG000 00009527	
			ENSDARG000 00099961	ENSTRUG000 00006491	ENSONIG000 00000438	ENSGACG000 00009527	
PIGF1	ENSG0000 0151665	ENSMUSG000 00024145	ENSDARG000 00032780	ENSTRUG000 00004839	ENSONIG000 00015903	ENSGACG000 00002773	ENSORLG000 00018733

Table S3.3: Background ensemble gene ids comprising of unknown targets of Nrf2:EpRE and HIF-1α:HRE. A list of 80 human gene names were randomly generated using molbiotools (http://www.molbiotools.com). Ensembl gene ids were then obtained for the genes in human, mouse, zebrafish, tilapia, stickleback, medaka and fugu suing the Ensembl Biomart database (as shown). Gene paralogs were excluded from this analysis so the final gene list comprises of 80 genes for each species.

Training Set	Human	Stickleback	Zebrafish	Mouse	Medaka	Fugu	Tilapia
CRB1	ENSG00000134376	ENSGACG00000009835	ENSDARG00000100506	ENSMUSG00000063681	ENSORLG00000009129	ENSTRUG00000009186	ENSONIG00000010727
FOXJ3	ENSG00000198815	ENSGACG00000007398	ENSDARG00000075774	ENSMUSG00000032998	ENSORLG00000004623	ENSTRUG00000014404	ENSONIG00000002167
SUN1	ENSG00000164828	ENSGACG00000011433	ENSDARG00000055350	ENSMUSG00000036817	ENSORLG00000008524	ENSTRUG00000003138	ENSONIG00000019101
MTUS1	ENSG00000129422	ENSGACG00000018595	ENSDARG00000071562	ENSMUSG00000045636	ENSORLG00000018633	ENSTRUG00000009546	ENSONIG00000006729
KIAA1217	ENSG00000120549	ENSGACG00000003838	ENSDARG00000059333	ENSMUSG00000036617	ENSORLG00000010340	ENSTRUG00000017841	ENSONIG00000009695
PTPN11	ENSG00000179295	ENSGACG00000015575	ENSDARG00000020334	ENSMUSG00000043733	ENSORLG00000000470	ENSTRUG00000018529	ENSONIG00000015302
SPRYD4	ENSG00000176422	ENSGACG00000003092	ENSDARG00000023309	ENSMUSG00000051346	ENSORLG00000012455	ENSTRUG00000002839	ENSONIG00000017030
GRXCR1	ENSG00000215203	ENSGACG00000017551	ENSDARG00000069865	ENSMUSG00000068082	ENSORLG00000000654	ENSTRUG00000014150	ENSONIG00000002391
PSG3	ENSG00000221826	ENSGACG00000009412	ENSDARG00000105324	ENSMUSG00000074272	ENSORLG00000013389	ENSTRUG00000003595	ENSONIG00000013122
RPL29	ENSG00000162244	ENSGACG00000007041	ENSDARG00000077717	ENSMUSG00000048758	ENSORLG00000004752	ENSTRUG00000007770	ENSONIG00000014377
SS18	ENSG00000141380	ENSGACG00000013236	ENSDARG00000002970	ENSMUSG00000037013	ENSORLG00000016977	ENSTRUG00000008991	ENSONIG00000009737
TRMT10B	ENSG00000165275	ENSGACG00000019076	ENSDARG00000060176	ENSMUSG00000035601	ENSORLG00000007862	ENSTRUG00000018087	ENSONIG00000012990
CCAR1	ENSG00000060339	ENSGACG00000007905	ENSDARG00000074759	ENSMUSG00000020074	ENSORLG00000007436	ENSTRUG00000013280	ENSONIG00000018419
AVP	ENSG00000101200	ENSGACG00000006569	ENSDARG00000042845	ENSMUSG00000037727	ENSORLG00000003907	ENSTRUG00000015631	ENSONIG00000015218
RALGDS	ENSG00000160271	ENSGACG00000014771	ENSDARG00000042409	ENSMUSG00000026821	ENSORLG00000020651	ENSTRUG00000007032	ENSONIG00000013950
RBMY1E	ENSG00000242389	ENSGACG00000017237	ENSDARG00000014244	ENSMUSG00000096520	ENSORLG00000001692	ENSTRUG00000006031	ENSONIG00000002700
PITX1	ENSG00000069011	ENSGACG00000016483	ENSDARG00000042785	ENSMUSG00000021506	ENSORLG00000000858	ENSTRUG00000009140	ENSONIG00000017304
ARSB	ENSG00000113273	ENSGACG00000015775	ENSDARG00000108788	ENSMUSG00000042082	ENSORLG00000002398	ENSTRUG00000003412	ENSONIG00000015486
FGFR2	ENSG00000066468	ENSGACG00000003443	ENSDARG00000058115	ENSMUSG00000030849	ENSORLG00000013277	ENSTRUG00000017610	ENSONIG00000010192
DIEXF	ENSG00000117597	ENSGACG00000012761	ENSDARG00000017696	ENSMUSG00000016181	ENSORLG00000019107	ENSTRUG00000011153	ENSONIG00000019525
NSMCE4A	ENSG00000107672	ENSGACG00000005042	ENSDARG00000024311	ENSMUSG00000040331	ENSORLG00000011593	ENSTRUG00000015029	ENSONIG00000009943
GOLGA6L6	ENSG00000277322	ENSGACG00000016644	ENSDARG00000063197	ENSMUSG00000002546	ENSORLG00000006984	ENSTRUG00000018651	ENSONIG00000002830
SHE	ENSG00000169291	ENSGACG00000004491	ENSDARG00000087956	ENSMUSG00000046280	ENSORLG00000009685	ENSTRUG00000005642	ENSONIG00000006285
CDKN3	ENSG00000100526	ENSGACG00000011791	ENSDARG00000039130	ENSMUSG00000037628	ENSORLG00000009158	ENSTRUG00000010247	ENSONIG00000019575
FTCD	ENSG00000160282	ENSGACG00000015411	ENSDARG00000007421	ENSMUSG00000001155	ENSORLG00000018636	ENSTRUG00000000589	ENSONIG00000009387
LRWD1	ENSG00000161036	ENSGACG00000020765	ENSDARG00000035147	ENSMUSG00000029703	ENSORLG00000014559	ENSTRUG00000012493	ENSONIG00000009466

Gene	Human	Gorilla	Zebrafish	Mouse	Medaka	Fugu	Tetraodon
SLC25A29	ENSG00000197119	ENSGACG00000012925	ENSDARG00000057352	ENSMUSG00000021265	ENSORLG00000017564	ENSTRUG00000009698	ENSONIG00000000788
KLF7	ENSG00000118263	ENSGACG00000002124	ENSDARG00000073857	ENSMUSG00000025959	ENSORLG00000010342	ENSTRUG00000008149	ENSONIG00000013381
ADRB1	ENSG00000043591	ENSGACG00000006578	ENSDARG00000007490	ENSMUSG00000035283	ENSORLG00000004803	ENSTRUG00000012396	ENSONIG00000020665
STRBP	ENSG00000165209	ENSGACG00000017377	ENSDARG00000021455	ENSMUSG00000026915	ENSORLG00000010162	ENSTRUG00000008384	ENSONIG00000013919
SLC35B1	ENSG00000121073	ENSGACG00000009684	ENSDARG00000038213	ENSMUSG00000020873	ENSORLG00000002514	ENSTRUG00000011550	ENSONIG00000000758
PPAT	ENSG00000128059	ENSGACG00000014340	ENSDARG00000004517	ENSMUSG00000029246	ENSORLG00000001965	ENSTRUG00000015737	ENSONIG00000002917
HOXB1	ENSG00000120094	ENSGACG00000003939	ENSDARG00000054033	ENSMUSG00000018973	ENSORLG00000012369	ENSTRUG00000010036	ENSONIG00000018209
CD109	ENSG00000156535	ENSGACG00000006787	ENSDARG00000060609	ENSMUSG00000046186	ENSORLG00000017286	ENSTRUG00000010621	ENSONIG00000008083
HIBCH	ENSG00000198130	ENSGACG00000002609	ENSDARG00000054867	ENSMUSG00000041426	ENSORLG00000015037	ENSTRUG00000012966	ENSONIG00000007161
SPTLC2	ENSG00000100596	ENSGACG00000005393	ENSDARG00000018976	ENSMUSG00000021036	ENSORLG00000010589	ENSTRUG00000018026	ENSONIG00000001084
CYP2A7	ENSG00000198077	ENSGACG00000012840	ENSDARG00000101423	ENSMUSG00000074254	ENSORLG00000001225	ENSTRUG00000015833	ENSONIG00000005899
HSF4	ENSG00000102878	ENSGACG00000013994	ENSDARG00000013251	ENSMUSG00000033249	ENSORLG00000013768	ENSTRUG00000000721	ENSONIG00000010204
PARP10	ENSG00000178685	ENSGACG00000008798	ENSDARG00000087145	ENSMUSG00000063268	ENSORLG00000013501	ENSTRUG00000013287	ENSONIG00000007669
HINFP	ENSG00000172273	ENSGACG00000020411	ENSDARG00000004851	ENSMUSG00000032119	ENSORLG00000007490	ENSTRUG00000016621	ENSONIG00000005379
WDR19	ENSG00000157796	ENSGACG00000016145	ENSDARG00000037406	ENSMUSG00000037890	ENSORLG00000005532	ENSTRUG00000008814	ENSONIG00000008982
PDLIM1	ENSG00000107438	ENSGACG00000003243	ENSDARG00000019845	ENSMUSG00000055044	ENSORLG00000014111	ENSTRUG00000017854	ENSONIG00000010331
APBB1	ENSG00000166313	ENSGACG00000011369	ENSDARG00000076560	ENSMUSG00000037032	ENSORLG00000004141	ENSTRUG00000006654	ENSONIG00000016866
DUSP10	ENSG00000143507	ENSGACG00000010062	ENSDARG00000052465	ENSMUSG00000039384	ENSORLG00000012052	ENSTRUG00000006678	ENSONIG00000005979
ATAD3A	ENSG00000197785	ENSGACG00000005036	ENSDARG00000086848	ENSMUSG0000236	ENSORLG00000018296	ENSTRUG00000010114	ENSONIG00000018915
LRRFIP1	ENSG00000124831	ENSGACG00000001620	ENSDARG00000030012	ENSMUSG00000026305	ENSORLG00000008211	ENSTRUG00000008570	ENSONIG00000016777
EVL	ENSG00000196405	ENSGACG00000008613	ENSDARG00000099720	ENSMUSG00000021262	ENSORLG00000015624	ENSTRUG00000010357	ENSONIG00000020106
MYBBP1A	ENSG00000132382	ENSGACG00000020754	ENSDARG00000078214	ENSMUSG00000040463	ENSORLG00000014339	ENSTRUG00000008378	ENSONIG00000017370
IGF2	ENSG00000167244	ENSGACG00000011125	ENSDARG00000033307	ENSMUSG00000048583	ENSORLG00000018930	ENSTRUG00000009557	ENSONIG00000014499
IGF2	ENSG00000167244	ENSGACG00000011125	ENSDARG00000018643	ENSMUSG00000048583	ENSORLG00000018930	ENSTRUG00000009557	ENSONIG00000014499
UQCRH	ENSG00000173660	ENSGACG00000015611	ENSDARG00000059128	ENSMUSG00000063882	ENSORLG00000008554	ENSTRUG00000007187	ENSONIG00000010536
HCN4	ENSG00000138622	ENSGACG00000016590	ENSDARG00000061685	ENSMUSG00000032338	ENSORLG00000013180	ENSTRUG00000005413	ENSONIG00000008418
TPCN1	ENSG00000186815	ENSGACG00000009537	ENSDARG00000062362	ENSMUSG00000032741	ENSORLG00000008165	ENSTRUG00000016229	ENSONIG00000013599
SNORD118	ENSG00000200463	ENSGACG00000022675	ENSDARG00000082850	ENSMUSG00000064899	ENSORLG00000021460	ENSTRUG00000019317	ENSONIG00000021946
AGL	ENSG00000162688	ENSGACG00000014614	ENSDARG00000106630	ENSMUSG00000033400	ENSORLG00000005432	ENSTRUG00000012741	ENSONIG00000001656
MRPL32	ENSG00000106591	ENSGACG00000005754	ENSDARG00000060489	ENSMUSG00000015672	ENSORLG00000006185	ENSTRUG00000014307	ENSONIG00000007276
MGAT4A	ENSG00000071073	ENSGACG00000002932	ENSDARG00000063330	ENSMUSG00000026110	ENSORLG00000012322	ENSTRUG00000014086	ENSONIG00000014241
SLCO1C1	ENSG00000139155	ENSGACG00000008749	ENSDARG00000016749	ENSMUSG00000030235	ENSORLG00000008279	ENSTRUG00000016998	ENSONIG00000008514
WBP1	ENSG00000239779	ENSGACG00000000318	ENSDARG00000092260	ENSMUSG00000030035	ENSORLG00000006422	ENSTRUG00000002699	ENSONIG00000017549
FBXO4	ENSG00000151876	ENSGACG00000006857	ENSDARG00000074170	ENSMUSG00000022184	ENSORLG00000004065	ENSTRUG00000015025	ENSONIG00000015180
GLRB	ENSG00000109738	ENSGACG00000016556	ENSDARG00000052782	ENSMUSG00000028020	ENSORLG00000014832	ENSTRUG00000015901	ENSONIG00000001184
BCL2L13	ENSG00000099968	ENSGACG00000013693	ENSDARG00000062370	ENSMUSG00000009112	ENSORLG00000015896	ENSTRUG00000003370	ENSONIG00000010327

EPHB1	ENSG0000 0154928	ENSGACG000 00013951	ENSDARG000 00076757	ENSMUSG00 000032537	ENSORLG000 00011003	ENSTRUG000 00018413	ENSONIG000 00010010
TRIL	ENSG0000 0255690	ENSGACG000 00007203	ENSDARG000 00100791	ENSMUSG00 000043496	ENSORLG000 00005131	ENSTRUG000 00018191	ENSONIG000 00021346
DUOX1	ENSG0000 0137857	ENSGACG000 00006163	ENSDARG000 00062632	ENSMUSG00 000033268	ENSORLG000 00005268	ENSTRUG000 00009066	ENSONIG000 00008359
CRBN	ENSG0000 0113851	ENSGACG000 00000750	ENSDARG000 00054250	ENSMUSG00 000005362	ENSORLG000 00001509	ENSTRUG000 00011346	ENSONIG000 00019050
DGCR6	ENSG0000 0183628	ENSGACG000 00011927	ENSDARG000 00005500	ENSMUSG00 000003531	ENSORLG000 00013085	ENSTRUG000 00006563	ENSONIG000 00010855
NOM1	ENSG0000 0146909	ENSGACG000 00003920	ENSDARG000 00060027	ENSMUSG00 000001569	ENSORLG000 00010541	ENSTRUG000 00010239	ENSONIG000 00005958
GDE1	ENSG0000 0006007	ENSGACG000 00018741	ENSDARG000 00055108	ENSMUSG00 000033917	ENSORLG000 00015102	ENSTRUG000 00004050	ENSONIG000 00005169
NRDC	ENSG0000 0078618	ENSGACG000 00003577	ENSDARG000 00019596	ENSMUSG00 000053510	ENSORLG000 00003474	ENSTRUG000 00015775	ENSONIG000 00004642
ECI1	ENSG0000 0167969	ENSGACG000 00005907	ENSDARG000 00018002	ENSMUSG00 000024132	ENSORLG000 00016704	ENSTRUG000 00014501	ENSONIG000 00006534
NPR3	ENSG0000 0113389	ENSGACG000 00003906	ENSDARG000 00035253	ENSMUSG00 000022206	ENSORLG000 00017232	ENSTRUG000 00013730	ENSONIG000 00013952
PLIN1	ENSG0000 0166819	ENSGACG000 00017870	ENSDARG000 00054048	ENSMUSG00 000030546	ENSORLG000 00006290	ENSTRUG000 00001733	ENSONIG000 00014533
ABCB6	ENSG0000 0115657	ENSGACG000 00008399	ENSDARG000 00074254	ENSMUSG00 000026198	ENSORLG000 00005333	ENSTRUG000 00009019	ENSONIG000 00012196
NHEJ1	ENSG0000 0187736	ENSGACG000 00015560	ENSDARG000 00058893	ENSMUSG00 000026162	ENSORLG000 00001663	ENSTRUG000 00012232	ENSONIG000 00012157
CYP2U1	ENSG0000 0155016	ENSGACG000 00017462	ENSDARG000 00026548	ENSMUSG00 000027983	ENSORLG000 00006861	ENSTRUG000 00008392	ENSONIG000 00017505
CALML4	ENSG0000 0129007	ENSGACG000 00016957	ENSDARG000 00075800	ENSMUSG00 000032246	ENSORLG000 00001729	ENSTRUG000 00016707	ENSONIG000 00005694
DYRK2	ENSG0000 0127334	ENSGACG000 00000712	ENSDARG000 00094646	ENSMUSG00 000028630	ENSORLG000 00017348	ENSTRUG000 00006772	ENSONIG000 00018022
COL9A2	ENSG0000 0049089	ENSGACG000 00007343	ENSDARG000 00024492	ENSMUSG00 000028626	ENSORLG000 00005372	ENSTRUG000 00015261	ENSONIG000 00004658
DDC	ENSG0000 0132437	ENSGACG000 00006397	ENSDARG000 00016494	ENSMUSG00 000020182	ENSORLG000 00005340	ENSTRUG000 00009415	ENSONIG000 00007162

Table S3.4: 0-order Markov Model of whole genome promoter sequences.

0-order Markov Model for the whole genome for the human, mouse, zebrafish, tilapia, stickleback, medaka and fugu. 3.5 kb and 5 kb of the upstream flank coding region was extracted for every gene in the genome for each species using the ensemble biomart tool.

Species	Human	Mouse	Zebrafish	Tilipia	Stickleback	Medaka	Fugu
A	2.624	2.718	0.3222	2.999	2.802	3.029	0.2785
C	2.376	2.282	1.778	2.001	2.198	1.971	0.2215
G	2.376	2.282	1.778	2.001	2.198	1.971	0.2215
T	2.624	2.718	3.222	2.999	2.802	3.029	0.2785

Species	Human	Mouse	Zebrafish	Tilipia	Stickleback	Medaka	Fugu
A	2.656	2.737	3.213	2.996	2.792	3.02	0.2796
C	2.344	2.263	1.787	2.004	2.208	1.98	0.2231
G	2.344	2.263	1.787	2.004	2.208	1.98	0.2231
T	2.656	2.737	3.213	2.996	2.792	3.02	0.2796

Table S3.5: Model selection results for 5kb mammalian datasets. Model selection results for 5kb datasets for HIF-1α and Nrf2 matrix files. Learning vector quantization (LVQ), stochastic gradient boosting (GBM), support vector machine (SVM) and Random Forest (RF) models were built on the mammalian dataset using the q-value, p-value, enrichment and position as parameters to predict positive and negative gene sets. Random forest models performed best for all matrix files and were selected for subsequent analysis.

	hif1_si		MA0259.1		MA1106.1	
	Accuracy	Kappa	Accuracy	Kappa	Accuracy	Kappa
LVQ	0.824	0	0.774	0	0.877	0.746
GBM	0.904	0.63	0.76	0.02	0.888	0.715
SVM	0.88	0.42	0.845	0.378	0.818	0.396
RF	0.982	0.93	0.887	0.58	0.865	0.643
	NFE2L2		MA0150.1		MA0150.2	
	Accuracy	Kappa	Accuracy	Kappa	Accuracy	Kappa
LVQ	0.81	0	0.816	0	0.804	0
GBM	0.984	0.942	0.984	0.942	0.985	0.949
SVM	0.96	0.859	0.96	0.859	0.985	0.949
RF	0.99	0.962	0.99	0.962	0.981	0.94

Table S3.6: Random forest model parameter for Nrf2:EpRE models.

Random Forest model training parameters showing the final model accuracy and kappa score following leave-one-out cross validation using the mammalian dataset (combined human and mouse top-scoring q-value result for each gene in the positive and background datasets).

Mammalian	kB	mtry	Accuracy	Kappa
Ma0150.1	3.5	5	0.98	0.97
	5	3	0.99	0.96
MA0150.2	3.5	2	0.98	0.95
	5	3	0.99	0.95
HOCOMOCO	3.5	5	0.98	0.93
	5	5	1.0	0.98
Fish				
Ma0150.1	3.5	3	0.98	0.95
	5	2	0.97	0.92
MA0150.2	3.5	4	0.99	0.97
	5	3	0.98	0.95
HOCOMOCO	3.5	5	0.93	0.85
	5	5	0.97	0.91

Table S3.7: Random forest model results for HIF-1α:HRE based models:
Random Forest model training parameters showing the final model accuracy
and kappa score following leave-one-out cross validation using the mammalian
dataset (combined human and mouse top-scoring q-value result for each gene
in the positive and background datasets).

Mammalian	kb	mtry	Accuracy	Kappa
Hifa_si	3.5	5	0.99	0.96
MA0259.1	3.5	5	0.96	0.86
MA1106.1	3.5	3	0.88	0.63
Fish				
Hifa_si	3.5	2	0.92	0.76
MA0259.1	3.5	2	0.94	0.80
MA1106.1	3.5	2	0.94	0.80

Table S3.8: Variable importance in Nrf2:EpRE random forest models. Plots of the variable importance for each random forest model in a) mammals and b) fish for 3.5kb and 5kb respectively using the q.value, p-value, z-score and strand as input variables.

	NFE2L2		MA0150.1		MA0150.2	
3.5 kb						
	Fish	**Mams.**	**Fish**	**Mams**	**Fish**	**Mams.**
q.value	100	100	100	100	100	100
p.value	57.61	21.99	66.47	27.97	60.66	17.18
freq	1.23	1.43	3.56	1.2	5.03	4.747
strand	0	0	0	0	0	0
5 kb						
q.value	100	100	100	100	100	100
p.value	61.98	9.47	57.49	74.43	37.4	16.26
freq	3.09	0.06	2.61	2.18	3.7	0.812
strand	0	0	0	0	0	0

Table S3.9: Variable importance in HIF-1α:HRE random forest models.

Plots of the variable importance for each random forest model in a) mammals and b) fish for 3.5kb and 5kb respectively using the q.value, p-value, z-score and strand as input variables.

	3.5 kb					
	HIF1_si		MA01106.1		MA0259.1	
	Fish	**Mams.**	**Fish**	**Mams**	**Fish**	**Mams.**
q.value	100	100	100	100	100	100
p.value	32.62	25.62	35.161	72.4	35.24	28.45
freq	24.22	3.034	4.2	12.86	37.84	5.8
strand	0	0	0	0	0	0
	5 kb					
q.value	100	100	100	100	100	100
p.value	42.94	34.52	48.08	90.44	43.77	21.3
freq	3.56	6.83	3.18	1.65	3.5	2.47
strand	0	0	0	0	0	0

Figure S3.3 Alignments of identified TFBS in the hmox1 promoter.

Alignments and q-values of identified binding sequences using T-Coffee for A). the mammalian and fish EpRE motifs and B). the mammalian and fish p53 motifs in the hmox1 promoter.

A.

	EpRE	Q-value
Human_1	ACTGGTGACTCAGCA	0.0365
Human_2	AAACATGACGCAGCA	0.0365
Human_3	CGGAATGACTCGGCG	0.0365
Human_4	GAGGCTGAATCAGCA	0.139
Mouse_1	GAGGGTGACTCAGCA	0.00204
Mouse_2	AACCATGACACAGCA	0.00451
Mouse_3	GACCGTGAGCGAGCA	0.282
Zebrafish_1	AAACATGACACAGCA	0.00862
Zebrafish_2	CGGAATGAACCAGCA	0.0569
Zebrafish_3	GAGTGTGATTCAGCT	0.0569
Zebrafish_4	AAACATGACTGGGCG	0.663
Zebrafish_5	TCTCATGACAAAGCA	0.663
Medaka_1	TGGAATGAGGCAGCA	0.101
Medaka_2	AAACATGACAGGGCA	0.174
Medaka_3	TAACATGACAGGGCA	0.174
Medaka_4	CTCACTGATTCAGCG	0.174
Tilapia_1	AAGCATGACTCAGCA	0.221
Tilapia_2	CCATTTGACTCAGCA	0.0389
Tilapia_3	AAACATGACGAGGCA	0.259
Stickleback_1	CACCGTGACACATCA	0.323
Fugu_1	AATCATGAGTCAGCA	0.00641
Fugu_2	AAGCGTGAGCCAGCA	0.031
Fugu_3	AAGCATGACCCTGCA	0.031
Fugu_4	AAACATGACACAACA	0.103
Fugu_5	CAGTGTGACTGATCA	0.11

	EpRE_fish	Q-value
Human_1	TGACTCAGC	0.276
Human_2	TGACTTATC	0.434
Mouse_1	TGACACAGC	0.172
Mouse_2	TGACTTGGC	0.172
Mouse_3	TGACTCAGC	0.131
Zebrafish_1	TGACTGGGC	0.0601
Zebrafish_2	TGACACAGC	0.120
Zebrafish_3	TGACTGCGC	0.0601
Zebrafish_4	TGACTGATC	0.139
Zebrafish_5	TGACAGGGC	0.225
Medaka_1	TGACAGGGC	0.225
Medaka_2	TGACTCATC	0.226
Medaka_3	TGACTCAGC	0.0516
Tilapia_1	TGACTCAGC	0.0516
Tilapia_2	TGACTGGGC	0.232
Fugu_1	TGACACAGC	0.232
Fugu_2	TGACTGATC	0.232
Fugu_3	TGACAGGGC	0.232
Fugu_4		

B.

	p53	Q-value
Human_1	GCAAGCCCAGACCGG	0.284
Human_2	GCATGCTGCAACTTG	0.284
Human_3	ACATGATCAGACTTC	0.441
Human_4	ACATGATCAGACTTC	0.441
Mouse_1	GCTAGTTCAGGCTGG	0.81
Zebrafish_1	GTGTGTCTAAACATG	0.461
Zebrafish_2	GCTAGTCTTTGCAAG	0.461
Medaka_1	ACCTGCCCTGTCATG	0.0968
Medaka_2	ACATGACAGGGCAGG	0.285
Stickleback_1	ACATGTCAAGTCTTG	0.113
Stickleback_2	GCAAGTCCTGGCTGG	0.17
Stickleback_3	TCAAGACTTGACATG	0.113
Tilapia_1	GTCAGCTCAAACATG	0.713

	p53_fish	Q-value
Human_1	AGACAAGCTTACAGGGACCT	1
Mouse_1	GGGCTAGTTCAGGCTGGTTC	0.039
Mouse_2	AAGCAAGCTTTAAATATTAA	0.284
Zebrafish_1	GAGCAGGTCCAGGACACACA	0.16
Zebrafish_2	AGGCAAGTTTGATAATTAAA	0.267
Zebrafish_3	AAACAGATCTAGAATGATTC	0.273
Medaka_1	GGGCAGGTTTGAAAGAGTCA	0.176
Stickleback_1	AAACTGGTCCAAAGTGAATA	0.713
Tilapia_1	GGACTGGTGTAGAAACTGTC	0.00913
Fugu_1	GGACTGGTCTGGGGGAAATC	0.0719
Fugu_2	AAACAGGTCTTGTGTATCCT	

230

Figure S3.4 Alignments of EpRE and p53 identified TFBS in the hsp70 promoter. Alignments and q-values of identified binding sequences using T-Coffee for A). the mammalian and fish EpRE motifs and B). the mammalian and fish p53 motifs in the hsp70 promoter.

A.

EpRE — Q-value

	Sequence	Q-value
Human_1	AT----TACTCAGCC	0.633
Mouse_1	CC---TGTCTCAGCA	0.195
Zebrafish_1	TTCTGTGACCCGGCA	0.312
Zebrafish_2	CACTGTGAATCTGCC	0.409
Medaka_1	GACGGTGACTCACCA	0.061
Medaka_2	TGCTGTGACTCAGGA	0.061
Medaka_3	CTGAGTGACTAAGCT	0.135
Medaka_4	AGACCTGACTTGGCA	0.135
Medaka_5	CCACCTGACTGAGCT	0.143
Medaka_6	AAGAATGAGTCACCT	0.143
Fugu_1	AAAGCTGAGACAGCA	0.424
Fugu_2	GCTGGTGAGTCAGCG	0.424
Tilapia_1	CAGAGTGAGACAGCA	0.0114

EpRE_fish — Q-value

	Sequence	Q-value
Human_1	TGACTGCGC	0.171
Human_2	TGACTTTGC	0.171
Medaka_1	TGACTGAGC	0.0996
Medaka_2	TGACTTGGC	0.145
Stickleback_1	TGACTGCGC	0.264
Stickleback_2	TGACAGCGC	0.36
Tillapia_1	TGACTTGGC	0.202
Tilapia_2	TGACAGAGC	0.202

B.

p53 — Q-value

	Sequence	Q-value
Human_1	ACATGTTACAACATG	0.0584
Mouse_1	ACAAGCGAAGACAAG	0.184
Stickleback_1	-ACATGTCAAGTCTTG	0.113
Stickleback_2	-GCAAGTCCTGGCTGG	0.113
Stickleback_3	-TCAAGACTTGACATG	0.17
Tilapia_1	ACCTGACAAGACATG	0.099
Tilapia_2	ACAAGCAGCAACTTG	0.219
Tilapia_3	ACATGTCTTGTCAGG	0.259

p53_fish — Q-value

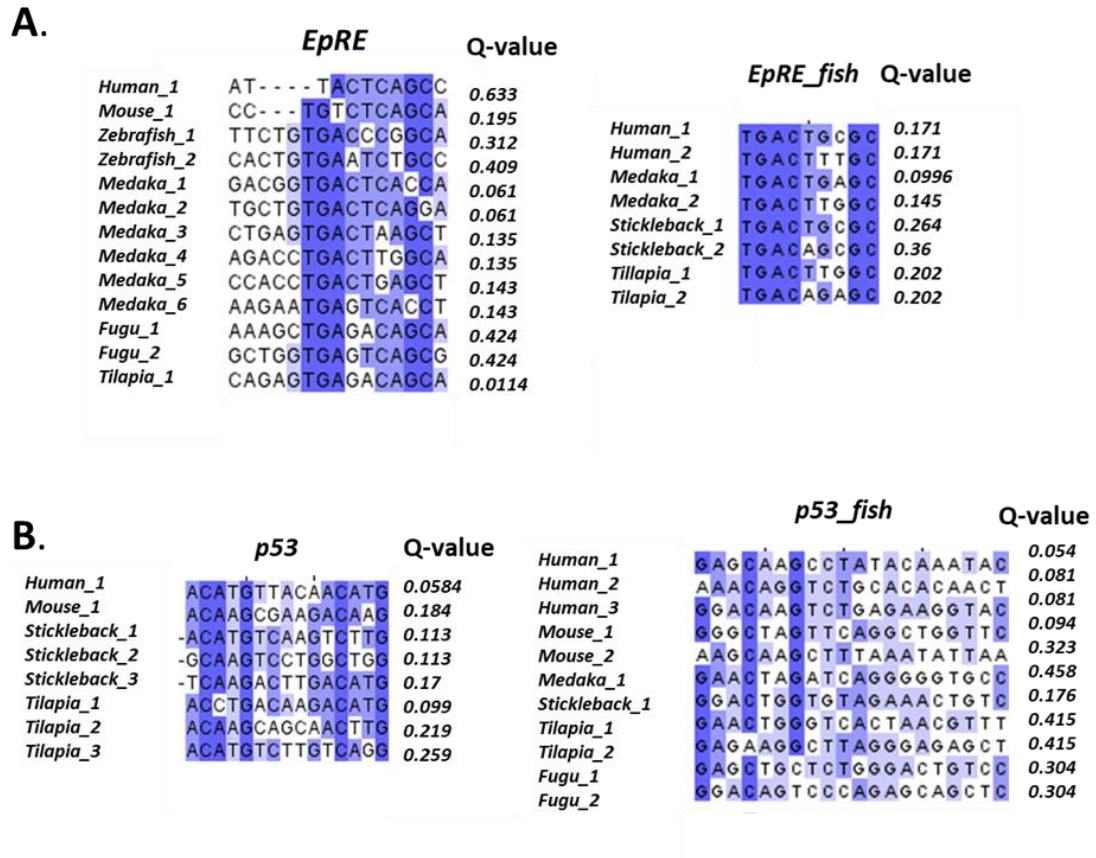|	Sequence	Q-value
Human_1	GAGCAAGCCTATACAAATAC	0.054
Human_2	AAACAGGTCTGCACACAACT	0.081
Human_3	GGACAAGTCTGAGAAGGTAC	0.081
Mouse_1	GGGCTAGTTCAGGCTGGTTC	0.094
Mouse_2	AAGCAAGCTTTAAATATTAA	0.323
Medaka_1	GAACTAGATCAGGGGGTGCC	0.458
Stickleback_1	GGACTGGTGTAGAAACTGTC	0.176
Tilapia_1	GAACTGGGTCACTAACGTTT	0.415
Tilapia_2	GAGAAGGCTTAGGGAGAGCT	0.415
Fugu_1	GAGCTGCTCTGGGACTGTCC	0.304
Fugu_2	GGACAGTCCCAGAGCAGCTC	0.304

231

Figure S3.5 Alignments of identified TFBS in the hsp70 promoter.

Alignments and q-values of identified binding sequences using T-Coffee for A) HSE search in hsp70, C) fish-sepcific HRE search in hsp70 E) AhR search in hsp70.

3.8 References:

1. OECD. 2017. Diffuse Pollution, Degraded Waters. Emerging Policy Solutions. *OECD Policy Highlights* (doi:10.1787/9789264269064-en)

2. Simmons SO, Fan CY, Ramabhadran R. 2009. Cellular stress response pathway system as a sentinel ensemble in toxicological screening. *Toxicol. Sci.* 111, 202–225. (doi:10.1093/toxsci/kfp140)

3. Günther V, Lindert U, Schaffner W. 2012. The taste of heavy metals: Gene regulation by MTF-1. *Biochim. Biophys. Acta - Mol. Cell Res.* 1823, 1416–1425. (doi:10.1016/j.bbamcr.2012.01.005)

4. Suzuki T, Takagi Y, Osanai H, Li L, Takeuchi M, Katoh Y, Kobayashi M. 2005 .Pi class glutathione S-transferase genes are regulated by Nrf2 through an evolutionarily conserved regulatory element in zebrafish. *Biochem. J.* 73, 65–73.

5. Timme-Laragy AR, Karchner SI, Franks DG, Jenny MJ, Harbeitner RC, Goldstone J V, McArthur AG, Hahn ME. 2012. Nrf2b, novel zebrafish paralog of oxidant-responsive transcription factor NF-E2-related factor 2 (NRF2). *J. Biol. Chem.* 287, 4609–27. (doi:10.1074/jbc.M111.260125)

6. Kobayashi M, Itoh K, Suzuki T, Osanai H, Nishikawa K, Katoh Y, Takagi Y, Yamamoto M. 2002. Identification of the interactive interface and phylogenic conservation of the Nrf2-Keap1 system. *Genes Cells* 7, 807–20.

7. Greenald D, Jeyakani J, Pelster B, Sealy I, Mathavan S, van Eeden FJ. 2015. Genome-wide mapping of Hif-1α binding sites in zebrafish. *BMC Genomics* 16, 923. (doi:10.1186/s12864-015-2169-x)

8. Lo R, Matthews J. 2012. High-Resolution Genome-wide Mapping of AHR and ARNT Binding Sites by ChIP-Seq. *Toxicol. Sci.* 130(2), 349–361. (doi:10.1093/toxsci/kfs253)

9. Kasner E, Hunter CA, Ph D, Kariko K. 2010. Metal transcription factor-1 regulation via MREs in the transcribed regions of selenoprotein H and other metal-responsive genes. *Biochim Biophys Acta* 70, 646–656. (doi:10.1002/ana.22528.Toll-like)

10. Wittwehr C, Whelan M. 2017. How Adverse Outcome Pathways Can Aid the Development and Use of Computational Prediction Models for Regulatory Toxicology. *Toxicol. Sci.* 155, 326–336. (doi:10.1093/toxsci/kfw207)

11.   Nioi P, McMahon M, Itoh K, Yamamoto M, Hayes JD. 2003. Identification of a novel Nrf2-regulated antioxidant response element (ARE) in the mouse NAD(P)H:quinone oxidoreductase 1 gene: reassessment of the ARE consensus sequence. *Biochem. J.* 374, 337–48. (doi:10.1042/BJ20030754)

12.   Wang XJ, Hayes JD, Wolf CR. 2006. Generation of a stable antioxidant response element-driven reporter gene cell line and its use to show redox-dependent activation of nrf2 by cancer chemotherapeutic agents. *Cancer Res.* 66, 10983–94. (doi:10.1158/0008-5472.CAN-06-2298)

13.   Matys V, Fricke E, Geffers R, Grossling E, Haubrock M, Hehl R, Hornischer K, Keras D, Kel A.E, Kel-Margoulis O.V, Kloos D.U, Land S, Lewicki-Potapov B, Micheal H, Munch R, Reuter I, Rotert S, Saxel H, Scheer M, Thiele S, Wingender E. 2003 .TRANSFAC®: Transcriptional regulation, from patterns to profiles. *Nucleic Acids Res.* 31, 374–378. (doi:10.1093/nar/gkg108)

14.   Mathelier A, Zhao X, Zhang A.W, Parcy F, Worsley-Hunt R, Arenillas D, Buchman S, Chen C Y, Chou A, Lenasescu H, Lim J, Shyr C, Tan G, Zhou M, Lenhard B, Sandelin A, Wasserman WW. 2014. JASPAR 2014: An extensively expanded and updated open-access database of transcription factor binding profiles. *Nucleic Acids Res.* 42, 142–147. (doi:10.1093/nar/gkt997)

15.   Kulakovskiy I V., Medvedeva YA, Schaefer U, Kasianov AS, Vorontsov IE, Bajic VB, Makeev VJ. 2013 .HOCOMOCO: a comprehensive collection of human transcription factor binding sites models. *Nucleic Acids Res.* 41, D195–D202. (doi:10.1093/nar/gks1089)

16.   Yusuf D, Butland SL, Swanson MI, Bolotin E, Ticoll A, Cheung WA, Yu X, Zhang C, Dickman CTD, Fulton DL, Lim JS, Schnabl JM, Ramos OHP, Vasseur-cognet M, Leeuw CND, Simpson EM, Ryffel GU, Lam EW, Kist R, Wilson MSC, Marco-ferreres R, Brosens JJ, Beccari LL, Nanan KK, Wegner M, Hou J, Mullen RD, Colvin SC, Noy PJ. 2012. The Transcription Factor Encyclopedia. *Genome Biol.* 13, R24. (doi:10.1186/gb-2012-13-3-r24)

17.   Portales-Casamar E, Kirov S, Lim J, Lithwick S, Swanson MI, Ticoll A, Snoddy J, Wasserman WW. 2007. PAZAR: a framework for collection and dissemination of cis-regulatory sequence annotation. *Genome Biol.* 8,

R207. (doi:10.1186/gb-2007-8-10-r207)

18. Pique-Regi R, Degner JF, Pai AA, Gaffney DJ, Gilad Y, Pritchard JK. 2011. Accurate inference of transcription factor binding from DNA sequence and chromatin accessibility data. *Genome Res.* 21, 447–55. (doi:10.1101/gr.112623.110)

19. Chorley BN, Campbell MR, Wang X, Karaca M, Sambandan D, Bangura F, Xue P, Pi J, Kleeberger SR, Bell DA. 2012. Identification of novel NRF2-regulated genes by ChIP-Seq: influence on retinoid X receptor alpha. *Nucleic Acids Res.* 40, 7416–29. (doi:10.1093/nar/gks409)

20. Dhakshinamoorthy S, Jain AK, Bloom D a, Jaiswal AK. 2005. Bach1 competes with Nrf2 leading to negative regulation of the antioxidant response element (ARE)-mediated NAD(P)H:quinone oxidoreductase 1 gene expression and induction in response to antioxidants. *J. Biol. Chem.* 280, 16891–900. (doi:10.1074/jbc.M500166200)

21. Mandriani B, Castellana S, Rinaldi C, Manzoni M, Venuto S, Rodriguez-Aznar E, Galceran J, Nieto MA, Borsani G, Monti E, Mazza T, Merla G, Micale L. 2016. Identification of p53-target genes in Danio rerio. *Scientific Reports,* 6, 1–13. (doi:10.1038/srep32474)

22. Miao W, Hu L, Scrivens PJ, Batist G. 2005. Transcriptional regulation of NF-E2 p45-related factor (NRF2) expression by the aryl hydrocarbon receptor-xenobiotic response element signaling pathway: Direct cross-talk between phase I and II drug-metabolizing enzymes. *J. Biol. Chem.* 280, 20340–20348. (doi:10.1074/jbc.M412081200)

23. van Uden P, Kenneth NS, Rocha S. 2008. Regulation of hypoxia-inducible factor-1α by NF-κB. *Biochem. J.* 412, 477–484. (doi:10.1042/BJ20080476)

24. Piechota M, Korostynski M, Przewlocki R. 2010. Identification of cis-regulatory elements in the mammalian genome: the cREMaG database. *PLoS One* 5, e12465. (doi:10.1371/journal.pone.0012465)

25. Broos S, Hulpiau P, Galle J, Hooghe B, Van Roy F, De Bleser P. 2011. ConTra v2: a tool to identify transcription factor binding sites across species, update 2011. *Nucleic Acids Res.* 39, W74-8. (doi:10.1093/nar/gkr355)

26. Kwon AT, Arenillas DJ, Worsley Hunt R, Wasserman WW. 2012. oPOSSUM-3: advanced analysis of regulatory motif over-representation

across genes or ChIP-Seq datasets. *G3 (Bethesda).* 2, 987–1002. (doi:10.1534/g3.112.003202)

27. Won KJ, Ren B, Wang W. 2010. Genome-wide prediction of transcription factor binding sites using an integrated model. *Genome Biol.* 11, R7. (doi:10.1186/gb-2010-11-1-r7)

28. Carvan MJ, Sonntag DM, Cmar CB, Cook RS, Curran MA, Miller GL. 2001. Oxidative stress in zebrafish cells: potential utility of transgenic zebrafish as a deployable sentinel for site hazard ranking. *Sci. Total Environ.* 274, 183–196.

29. Mukaigasa K, Nguyen LTP, Li L, Nakajima H, Yamamoto M, Kobayashi M. 2012. Genetic Evidence of an Evolutionarily Conserved Role for Nrf2 in the Protection against Oxidative Stress. *Mol. Cell. Biol.* 32, 4455–4461. (doi:10.1128/MCB.00481-12)

30. Mukaigasa K, Nguyen LTP, Li L, Nakajima H, Yamamoto M, Kobayashi M. 2012. Genetic evidence of an evolutionarily conserved role for Nrf2 in the protection against oxidative stress. *Mol. Cell. Biol.* 32, 4455–61. (doi:10.1128/MCB.00481-12)

31. Iwanaszko M, Brasier AR, Kimmel M. 2012. The dependence of expression of NF-κB-dependent genes: statistics and evolutionary conservation of control sequences in the promoter and in the 3' UTR. *BMC Genomics* 13, 182. (doi:10.1186/1471-2164-13-182)

32. Siggers T, Chang AB, Teixeria A, Wong D, Williams KJ, Ahmed B, Ragousis J, Udalova IA, Smale ST, Martha L. 2012. Principles of dimer-specific gene regulation revealed by a comprehensive characterization of NF- κ B family DNA binding. *Nat. Immunol.* 13, 95–102. (doi:10.1038/ni.2151.Principles)

33. O'Quin KE, Smith D, Naseer Z, Schulte J, Engel SD, Loh Y-HE, Streelman JT, Boore JL, Carleton KL. 2011. Divergence in cis-regulatory sequences surrounding the opsin gene arrays of African cichlid fishes. *BMC Evol. Biol.* 11, 120. (doi:10.1186/1471-2148-11-120)

34. Ritter DI, Li Q, Kostka D, Pollard KS, Guo S, Chuang JH. 2010. The importance of Being Cis: Evolution of Orthologous Fish and Mammalian enhancer activity. *Mol. Biol. Evol.* 27, 2322–2332. (doi:10.1093/molbev/msq128)

35. Giorgetti L, Siggers T, Tiana G, Caprara G, Notarbartolo S, Corona T,

Pasparakis M, Milani P, Bulyk ML, Natoli G. 2010. Noncooperative Interactions between Transcription Factors and Clustered DNA Binding Sites Enable Graded Transcriptional Responses to Environmental Inputs. *Mol. Cell* 37, 418–428. (doi:10.1016/j.molcel.2010.01.016)

36. Hahn ME, McArthur AG, Karchner SI, Karchner SI, Franks DG, Jenny MJ, Timme Laragy AR, Stegemen JJ, Woodin BR, Cipriano MJ, Linney E. 2014. The Transcriptional Response to Oxidative Stress during Vertebrate Development: Effects of tert-Butylhydroquinone and 2,3,7,8-Tetrachlorodibenzo-p-Dioxin. *PLoS One,* 9, 11, e113158. (doi:10.1371/journal.pone.0113158)

37. Lee O, Green JM, Tyler CR. 2015. Transgenic fish systems and their application in ecotoxicology. *Crit. Rev. Toxicol.* 45, 124–141. (doi:10.3109/10408444.2014.965805)

38. Carvan MJ, Solis W a, Gedamu L, Nebert DW. 2000. Activation of transcription factors in zebrafish cell cultures by environmental pollutants. *Arch. Biochem. Biophys.* 376, 320–7. (doi:10.1006/abbi.2000.1727)

39. Volff JN. 2005. Genome evolution and biodiversity in teleost fish. *Heredity (Edinb).* 94, 280–294. (doi:10.1038/sj.hdy.6800635)

40. Ensembl - Ensembl Protein Trees, https://www.ensembl.org/info/genome/compara/homolo.

41. Kuhn M. 2008. Building Predictive Models in R Using the caret Package. *JSS J. Stat. Softw.* 28.

42. Molbiotools, *http://www.molbiotools.com.*

43. Grant CE, Bailey TL, Noble WS. 2011. FIMO: scanning for occurrences of a given motif. *Bioinforma. Appl. NOTE* 27, 1017–101810. (doi:10.1093/bioinformatics/btr064)

44. Mathelier A, Fornes O, Arenillas DJ, Chen C, Denay G, Lee J, Shi W, Shyr C, Tan G, Worsley-Hunt R, Zhang AW, Parcy F, Lenhard B, Sandelin A, Wasserman WW. 2016. JASPAR 2016: a major expansion and update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.* 44, D110–D115. (doi:10.1093/nar/gkv1176)

45. Pronk TE, van der Veen JW, Vandebriel RJ, van Loveren H, de Vink EP, Pennings JLA. 2014. Comparison of the molecular topologies of stress-activated transcription factors HSF1, AP-1, NRF2, and NF-κB in their induction kinetics of HMOX1. *BioSystems* 124, 75–85.

(doi:10.1016/j.biosystems.2014.09.005)

46. Alam J, Cook JL. 2007. How many transcription factors does it take to turn on the heme oxygenase-1 gene? *Am. J. Respir. Cell Mol. Biol.* 36, 166–174. (doi:10.1165/rcmb.2006-0340TR)

47. Almeida DV, Nornberg BFDS, Geracitano L a, Barros DM, Monserrat JM, Marins LF. 2010. Induction of phase II enzymes and hsp70 genes by copper sulfate through the electrophile-responsive element (EpRE): insights obtained from a transgenic zebrafish model carrying an orthologous EpRE sequence of mammalian origin. *Fish Physiol. Biochem.* 36, 347–53. (doi:10.1007/s10695-008-9299-x)

48. Fischer M. 2017. Census and evaluation of p53 target genes. *Oncogene* 36, 3943–3956. (doi:10.1038/onc.2016.502)

49. Huang W-J, Xia LM, Zhu F, Huang B, Zhou C, Zhu HF, Wang B, Chen B, Lei P, Shen GX, De-AnTian. 2009. Transcriptional upregulation of HSP70-2 by HIF-1 in cancer cells in response to hypoxia. *Int. J. Cancer* 124, 298–305. (doi:10.1002/ijc.23906)

50. Evans TG, Belak Z, Ovsenek N, Krone PH. 2007. Heat shock factor 1 is required for constitutive Hsp70 expression and normal lens development in embryonic zebrafish. *Comp. Biochem. Physiol. A. Mol. Integr. Physiol.* 146, 131–140. (doi:10.1016/j.cbpa.2006.09.023)

51. Notredame C, Higgins DG, Heringa J. 2000. T-coffee: A novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.* 302, 205–217. (doi:10.1006/jmbi.2000.4042)

52. Ramsden R, Gallagher EP. 2016. Dual NRF2 paralogs in Coho salmon and their antioxidant response element targets. *Redox Biol.* 9, 114–123. (doi:10.1016/j.redox.2016.07.001)

53. Kajimura S, Aida K, Duan C. 2005. Insulin-like growth factor-binding protein 1 (IGFBP-1) mediates hypoxia-induced embryonic growth and developmental retardation. *PNAS* 102, 4, 1204-1245 (doi:10.1073)

54. Kulkarni RP, Tohari S, Ho A, Brenner S, Venkatesh B. 2010. Characterization of a hypoxia-response element in the Epo locus of the pufferfish, Takifugu rubripes. *Mar. Genomics* 3, 63–70. (doi:10.1016/j.margen.2010.05.001)

55. Rees BB, Figueroa YG, Wiese TE, Beckman BS, Schulte PM. 2009. A novel hypoxia-response element in the lactate dehydrogenase-B gene of

the killifish Fundulus heteroclitus. *Comp. Biochem. Physiol. - A Mol. Integr. Physiol.* 154, 70–77. (doi:10.1016/j.cbpa.2009.05.001)

56. ZeRuth G, Pollenz RS. 2007. Functional analysis of cis-regulatory regions within the dioxin-inducible CYP1A promoter/enhancer region from zebrafish (Danio rerio). *Chem. Biol. Interact.* 170, 100–113. (doi:10.1016/j.cbi.2007.07.003)

57. Powell WH, Morrison HG, Weil EJ, Karchner SI, Sogin ML, Stegeman JJ, Hahn ME. 2004 .Cloning and analysis of the CYP1A promoter from the atlantic killifish (Fundulus heteroclitus). *Mar. Environ. Res.* 58, 119–24. (doi:10.1016/j.marenvres.2004.03.005)

58. Yan CHM, Chan KM. 2004. Cloning of zebrafish metallothionein gene and characterization of its gene promoter region in HepG2 cell line. *Biochim. Biophys. Acta* 1679, 47–58. (doi:10.1016/j.bbaexp.2004.04.004)

59. Olsson' E, Kling' P, Erkell' LJ, Kille3 P. 1995. Structural and functional analysis of the rainbow trout (Oncorhyncus mykiss) metallothionein-A gene. *Eur. J. Biochem.* 230, 344-349.

60. Samson SL, Gedamu L. 1995. Metal-responsive elements of the rainbow trout metallothionein-B gene function for basal and metal-induced activity. *J. Biol. Chem.* 270, 6864–71. (doi:10.1074/JBC.270.12.6864)

61. Olsson PE, Kille P. 1997. Functional comparison of the metal-regulated transcriptional control regions of metallothionein genes from cadmium-sensitive and tolerant fish species. *Biochim. Biophys. Acta* 1350, 325–34.

62. Scudiero R, Carginale V, Capasso C, Riggio M, Filosa S, Parisi E. 2001. Structural and functional analysis of metal regulatory elements in the promoter region of genes encoding metallothionein isoforms in the Antarctic fish Chionodraco hamatus (icefish). *Gene* 274, 199–208.

63. Searle PF, Stuart GW, Palmiter RD. 1985. Building a metal-responsive promoter with synthetic regulatory elements. *Mol. Cell. Biol.* 5, 1480–9.

64. Tuğrul M, Paixão T, Barton NH, Tkačik G, Fay JC. 2015. Dynamics of Transcription Factor Binding Site Evolution. *PLOS Genet,* 11 (11):e1005639 (doi:10.1371/journal.pgen.1005639)

65. Spivakov M. 2014. Spurious transcription factor binding: Non-functional or genetically redundant? *Bioessays* 36, 798–806. (doi:10.1002/bies.201400036)

66. Postlethwait J, Amores A, Cresko W, Singer A, Yan Y. 2004. Subfunction

partitioning , the teleost radiation and the annotation of the human genome. *TRENDS Genet.* 20, 10 (doi:10.1016/j.tig.2004.08.001)

67. Vorrink SU, Domann FE. 2014. Regulatory crosstalk and interference between the xenobiotic and hypoxia sensing pathways at the AhR-ARNT-HIF1α signaling node. *Chem. Biol. Interact.* 218, 82–88. (doi:10.1016/J.CBI.2014.05.001)

68. Oyarbide U, Iturria I, Rainieri S, Pardo MA. 2015. Use of gnotobiotic zebrafish to study Vibrio anguillarum pathogenicity. *Zebrafish* 12, 71–80. (doi:10.1089/zeb.2014.0972)

69. Lin GG, Scott JG. 2012. An essential role for heat shock transcription factor binding protien 1 (hsbp1) during early embryonic development. 100, 130–134. (doi:10.1016/j.pestbp.2011.02.012.Investigations)

70. Wardle FC, Tan H. 2015. A ChIP on the shoulder? Chromatin immunoprecipitation and validation strategies for ChIP antibodies. *F1000Research* 235, 1–10. (doi:10.12688/f1000research.6719.1)

71. Barolo S. 2012. Shadow enhancers: Frequently asked quations about distriuted cis-regulatory information and enhancer redundancy. *Bioessays* 32, 135–141. (doi:10.1016/j.jacc.2007.01.076.White)

72. Yang L, Zhou T, Dror I, Mathelier A, Wasserman WW, Gordân R, Rohs R. 2014 .TFBSshape: a motif database for DNA shape features of transcription factor binding sites. *Nucleic Acids Res.* 42, D148–D155. (doi:10.1093/nar/gkt1087)

73. Seok H-S, Kim J. 2013. Computational prediction of transcription factor binding sites based on an integrative approach incorporating genomic and epigenomic features. *Genes Genomics* 36, 25–30. (doi:10.1007/s13258-013-0136-y)

Chapter 4

Identifying the conservation of adaptive-stress responses in vertebrates based on experimental evidence from *in vitro* and *in vivo* studies.

4.1 Abstract

Despite growing evidence for shared downstream target genes and regulatory interactions between transcription factors, environmental pollutants with prooxidant effects are widely considered to act through distinct molecular pathways. The adaptive stress-response transcription factor (TF) nuclear factor (erythroid-derived)- like 2 (Nrf2) is considered a central mediator of antioxidant defence processes but its interactions with other regulatory factors remains largely unexplored. A boolean model of regulatory connectivity between adaptive stress response factors in mammals suggested that responses to inducers can be conserved and hypothesized that activation of Nrf2 results in coherent responses with other TF pathways. These include hypoxia inducible factor 1 –alpha (HIF1a), heat shock factor 1 (HSF1), aryl hydrocarbon receptor (AhR) and metal transcription factor 1 (MTF1) which are predicted to lead to canalised response processes. This research aimed to validate this hypothesis through a comparative analysis on gene expression signatures across fish species exposed to pollutants. To explore connectivity between pathways in greater depth, the zebrafish was used as a model to identify potential interactions between Nrf2 and the transcription factors metal transcription factor 1 (*mtf1*), hypoxia inducible factor 1 (*hif1a*) and nuclear factor kappa b (*nfkb1*) at 2 and 4 dpf and throughout early development. Exposures to the inducer tert-butyl hydroquinone (tBHQ) identified that the antioxidant genes sequestosome-1 (*sqstm1)* and glutathione-s-transferase pi (*gstp1*) had a greater level of inducible expression at 4 dpf compared to 2 dpf. Further analysis identified that both *mtf1* and *nfkb1* were upregulated following 12 hr exposures to tBHQ in comparison to 6 hrs, suggesting interactions between regulatory cascades following Nrf2 induction. In addition, acridine orange staining identified increases in the number of pre-apoptotic cells in the gill, jaw and hindbrain in a concentration dependant manner after 6 hrs tBHQ exposures at 4 dpf, suggesting these are likely target tissues of tBHQ. This research indicated some similarities in gene-expression responses across inducers in fish specific exposures, providing some support for the modelling outcomes presented in chapter 2. tBHQ exposures altered levels of *nfkb1* and *mtf1,* suggesting some level of connectivity with Nrf2 activation. Further research is necessary to establish if regulatory connectivity between these pathways is Nrf2 dependent.

4.2 Introduction

The nuclear factor (erythroid-derived 2)-like *2* (Nrf2) – Kelch-like ECH associated protein 1 (Keap1) pathway is evolutionary conserved across vertebrate groups and widely regarded as being a "master regulator" of the oxidative stress response (OSR), initiating the transcription of antioxidant defence genes and phase II detoxification enzymes [1],[2]. As such, the Nrf2-Keap1 pathway is of particular interest from both medical and environmental perspectives where its induction by a range of electrophilic compounds and by cellular redox imbalance has been identified as preventing the progression of a number of neurological disorders[3] and diseases such as cancer [4]. Under basal conditions, Nrf2 is held in the cytoplasm by the inhibitory protein Keap1 and is degraded by the Cul3 ubiquitination ligase complex[5]. Electrophilic compounds, which contain molecules that are electron deficient, modify cysteine (Cys) residues on the surface of Keap1 causing a conformational change and the release of Nrf2. The transcription factor then accumulates in the nuclease and forms a heterodimer with small maf proteins before binding to electrophile response elements (EpREs) within promoter regions[5]. The inducible expression of genes such as glutathione-s-transferase Pi (GSTP), NAD(P)H quinone dehydrogenase (NQO1) and superoxide dismutase (SOD1) which have roles in detoxification processes, have all been attributed to the binding of Nrf2 in promoters[1],[6] .

However, although the Nrf2-Keap1 pathway is widely regarded as being responsible for the expression of antioxidant defence processes, the roles of multiple transcription factors in regulating shared downstream targets is becoming increasingly well documented. Redox-associated transcription factors including the aryl hydrocarbon receptor (AhR), hypoxia-inducible factor 1-alpha (HIF-1α), heat shock factor 1 (HSF1), nuclear-factor kappa-light-chain-enhancer of B cells (NFkB) and metal-transcription factor 1 (MTF1) have all been attributed to either sharing downstream target genes with Nrf2 or being involved in direct protein-protein interactions, identifying that the adaptive stress response acts in a gene regulatory network (GRN) (Chapter 2). The research presented in Chapter 2 of this thesis suggests that the activation of adaptive-stress response pathways through MTF1, AhR, HSF1 and HIF1-a at molecular initiating events (MIEs) can lead to similar outcome processes as Nrf2; this

suggests canalised responses can be reached where the same outcome is produced regardless of the inducer (Chapter 2). This hypothesis has implications for risk assessment methods in predictive toxicology by proposing outcomes to chemical toxicants based on transcription factor interactions. Such predictions are of a particular benefit to ecotoxicology where aquatic vertebrates are at high risk from pollutant exposures to toxicants predicted to act through adaptive-stress response pathways at molecular initiating events (MIEs).

The regulatory interactions that form the basis of the adaptive stress response GRN have seen little validation *in vivo*, where model organisms such as the *Danio rerio* (zebrafish) provide a means of establishing the transferability of the model across vertebrate groups and during embryonic development[7]. An exception is the AhR, traditionally associated with inducing phase I detoxification enzymes in the presence of dioxin-like chemicals such as poly-aromatic hydrocarbons (PAHs) by binding to xenobiotic response elements (XREs) in regulatory regions. In this case, chemicals that are able to activate both the release of Nrf2 from Keap1 and the binding of the AhR to the Aryl-hydrocarbon nuclear translocator (ARNT), are termed bifunctional inducers and up-regulate phase I and phase II detoxification processes [8]. Positive feedback interactions between the two factors have been established in mouse embryonic fibroblast (MEFs) and mouse hepatoma 1c1c7 (Hepa-1c1c7) cell lines where functional EpRE binding sites have been identified in the promoter of the AhR[8] and of XREs in Nrf2[9] respectively. Exposures in mice to the AhR inducer TCDD caused an upregulation and nuclear accumulation of Nrf2[10] . Evidence suggests this mechanism is conserved across vertebrate groups with zebrafish eleutheroembryos exposed to PCB-126 showing an upregulation of both *nrf2a/b* and AhR subtypes, the expression of which was removed in an *nrf2a* mutant line[11].

Validation of the outcomes of cross-talk between adaptive-stress response factors requires the level of concordance between responses to the inducers of seemingly different pathways to be assessed. This will lead to a better understanding of how inducers of adaptive stress-response pathways - in particular, chemical pollutants that are a significant exposure risk to aquatic vertebrates - can lead to adverse effects. To this end, prototypic inducers,

agonists of specific TFs, can be used to identify cross-talk between pathways[12] and have established the evolutionary conservation of Nrf2 across vertebrate species[12,13]; the Nrf2 agonists, diethylmaleate (DEM) and tert-butyl-1,4-hydroquinone (tBHQ), induce comparable gene expression responses across mammalian systems and zebrafish exposure scenarios[12]. These inducers initiate the release of Nrf2 through associating with Cys[151] on Keap1[5]. The mechanism for this is well documented for tBHQ, which is oxidised to tert-butylquinone (tBQ) by Cu^{2+} releasing superoxide and Cu^+. O_2^- is converted to hydrogen peroxide (H_2O_2) by Cu^+ releasing Cu^{2+} (Figure 4.1) [14]. Both tBQ and H_2O_2 can act on Keap1 to cause the release of Nrf2[15]. The upregulation of *gstp1* after 6-hour exposure to 30 µM tBHQ in 4 days post fertilisation (dpf) zebrafish larvae [2] has become a standard exposure time-period to measure Nrf2-mediated OSR, regardless of developmental stage. Differential expression patterns of antioxidant genes have since been established at 2, 4, 5, 6 and 7 dpf under a range of doses from 3 – 30 µM tBHQ [2,12].

.
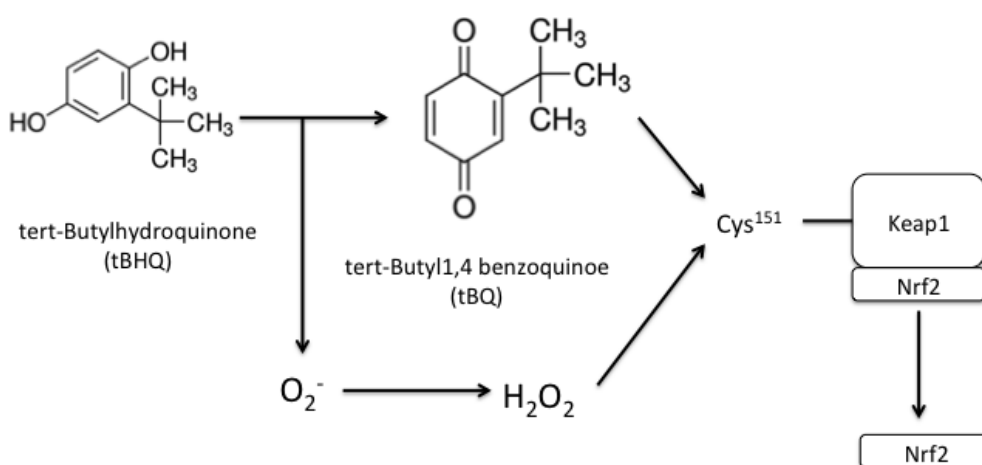
Figure 4.1. The activation of Nrf2 by tert-butylhydroquine (tBHQ). tBHQ is converted to tert-butyl-1,4-bezoquinone (tBQ) and superoxide. tBQ activates the release of Nrf2 from Keap1 by associating with Cys[151]. O_2^- is converted to H_2O_2, which can also act on Cys[151] releasing Nrf2.

This study aimed to validate the hypothesis that canalised responses occur following the induction of Nr2, AhR, HIF1a, MTF1 and HSF1 pathways under chemical induction across vertebrate species. Similarities in response outcomes

were determined across literature-derived studies based on exposures in fish species to a range of inducers of the selected adaptive stress-response factors. Furthermore, considering that Nrf2 is a central mediator in the adaptive stress response (Chapter 2), understanding its function as either an initiator or inhibitor of other response pathways will provide fundamental insights into the roles of redox-sensitive transcription factors within an integrated system. In addition, as Nrf2 is neither maternally deposited[16] nor essential for the function of normal development, comparing the developmental expression of redox-factors gives an indication of the capacity to mount an antioxidant response in the absence of Nrf2 in early development. Using the zebrafish as a model, the expression dynamics of *mtf1*, *hif1a* and *nfkb1* were examined in the context of nrf2 expression during embryogenesis, where vertebrates are suggested to be highly vulnerable to oxidative insult. In addition, the expression dynamics of these factors under the model Nrf2-inducer tBHQ was conducted as a factor of dose and time.

This study identified some similarities across exposures in fish species to a variety of inducers and the predicted modelling outcomes presented in Chapter 2. Further support for the modelling outcomes was provided through RT-qPCR analysis in zebrafish. This showed that *nrf2a* had the lowest level of detectable expression when compared to the expression of *hif1a*, *nfkb1* and *mtf1* in pre- 5 dpf zebrafish embryos. Zebrafish exposures to tBHQ at 2 dpf and 4 dpf showed different levels of the antioxidants *gstp1* and *sqstm1* suggesting variations in susceptibility between life stages. At 4 dpf, exposures identified responses to tBHQ for *mtf1* and *nfkb1* as a factor of dose and time. Differential expression between *keap1a/b* and *nrf2a/b* identified that the divergent function of these paralogs equates to variations in expression as a factor of chemical insult. The changes in gene expression profiles were correlated with increased pre-apoptotic cells in zebrafish larvae across exposure concentrations.

4.3 Materials and Methods

To validate the hypothesis that adaptive-stress response factors HIF-1α, HSF1, MTF, AhR and Nrf2 can lead to similar response outcomes. Firstly, validation for the outcomes was established in mammalian cell lines. Secondly, evidence for pathway activation in fish was determined from the literature. Through laboratory studies the developmental expression of,*hif1a*, *mtf1*, *nfkb1* and *nrf2*, was determined over embryogenesis using RT-qPCR. To identify the appropriate life-stage for transcript expression under the Nrf2 inducer, tBHQ, the downstream targets, *sqstm1* and *gstp1* were subsequently quantified at 2 and 4 dpf. Following this, the most responsive stage was selected to measure the expression of *hif1a*, *mtf1*, *nfk1b*, *nrf2a*, *nrf2b*, *keap1a* and *keap1b* over time and exposure concentrations. An apoptosis assay was conducted as a measure of adverse outcome. The experimental methods are shown in Figure 4.2.

4.3.1. Evidence of shared responses in fish-specific exposures.

To validate the results of Boolean model presented in chapter 2, the literature was searched for fish studies where gene expression had been measured under exposures to either prototypic inducers or environmental contaminants known to activate AhR MTF1, ER, Nrf2, HIF-1α and HSF1 pathways. Data was collated from both targeted (RT-qPCR/ WISH) and non-targeted approaches (RNA-seq/microarray) for recording gene expression and downstream target genes involved in the adaptive stress pathway (Figure 2.4) were searched. The expression profiles were compared to the results of the Boolean model presented in Chapter 2 and shown in the supplementary information (Figure 2.4).

4.3.1. Fish maintenance

Breeding stocks of WIK strain zebrafish were maintained at 28 ± 0.5 °C with a 12 hr light: dark cycle including a 30 min dawn-dusk transition period. Zebrafish were fed artemia twice daily. Embryos were collected from multiple tanks immediately following spawning washed with ISO water and mixed to avoid potential inter-tank effects. Following examination under a stereomicroscope, unfertilized embryos were removed prior to sorting into groups of 30. Embryos were maintained in petri-dishes with 50 ml of ISO water with full water changes conducted on a daily basis.

4.3.2. Developmental time series

Embryos were maintained in petri-dishes with 50 ml of ISO water and complete water changes were conducted on 24 hr basis. At each sampling time point, 3 pools of 15 embryos were harvested from dishes containing 30 embryos with four experimental replicates (n=12). Samples were immediately snap-frozen in liquid nitrogen and stored at -80 °C until RNA extraction.

4.3.3. Chemical exposures

tBHQ (97% purity) and analytical-grade dimethyl sulfoxide (DMSO) were purchased from Sigma-Aldrich Chemical co (Gillingham, UK). A stock solution was prepared by dissolving tBHQ in DMSO at a concentration of 300 mM and stored at 4 °C. Working stocks were made on the morning of each exposure in ISO water to give a final DMSO concentration of 0.01%. For RT-qPCR analysis, 30 larvae per concentration and time point were exposed in 50 ml of solution in glass cylindrical dishes. All experiments were carried out in triplicate and repeated 3 times.

At 2 and 4 dpf embryos/larvae were exposed in groups of 30 for 6 and 12 hrs to ISO water, 0.01% DMSO, 3 µM, 10 µM and 30 µM tBHQ. Samples were snap-frozen in liquid nitrogen and subsequently stored at -80 °C prior to RNA extraction.

For apoptosis analysis, acridine orange staining was conducted following exposures of 4 dpf larvae for 6 hrs in 10 ml of each concentration (ISO water, 0.01% DMSO, 3 µM, 10 µM and 30 µM tBHQ) within a 20 ml glass-beaker which was replicated three times (n = 9 pooled from 10 larvae per group).

4.3.4. Quantification of gene expression

Real-time quantitative PCR (RT-qPCR) was conducted to determine the relative expression of selected regulatory proteins involved in stress response pathways and downstream target genes. Primer sequences were sourced from the literature where possible (Table 4.1) or designed using Beacon Designer 7.0 software (Premier Biosoft Internations, Paulo Alto, CA) using zebrafish NCBI

RefSeq sequences for selected target genes. All primers were purchased from Eurofins Genomics (Ebersberg. Germany) and diluted to a working stock of 10 µM.

Annealing temperatures (Ta) were optimised using a temperature gradient and the specificity of each primer was confirmed by observing a single amplification product of expected size and Tm. RT-qPCR was run at 95°C for 15 minutes followed by 40 cycles of 95°C for 10 seconds and Ta 30 seconds. The primer specificity, detection range and amplification efficiency were established using a serial dilution of zebrafish cDNA at 4 dpf followed by melt-curve analysis. Over the detection range, the linear correlation (R^2) between mean cycle threshold (Ct) with the logarithm of the cDNA dilution >0.99.

Target Gene	Forward Primer (5'-3')	Reverse Primer (5'-3')	Ta (°C)	PCR efficiency
rpl8*	CCGAGACCAAGAAATCCAGAG	CCAGCAACAACACCAACAAC	59.5	104.20%
gstp1	ACGACAGTGAGGCTTCC	GAGGTGGTTGGGCAGAT	59.5	110.60%
sqstm1	GGCGTAAGATGAGACACT	GAGGCAGTAGCACCATT	59.5	109.01%
nrf2a	CGAGATGAGAACGGAAAGG	GAAGGAGGAAGGACAAAGC	57.5	107.30%
nrf2b*	GGCAGAGGGAGGAGGAGACCAT	AAACAGCAGGGCAGACAACAAGG	65	106.50%
keap1a	TACACCTTCGCACCAGAG	TCGCAGAGCACCTTCAG	59.5	93.80%
keap1b*	ACGGAGTGTAAGGCGGAG	ACCTGGCTGAAGTTCATG	61.5	90.40%
mtf1	GGATGAGGAAGGAGAAGA	ATGGTGTGGTGGATGTA	59.5	93.10%
nfkb1*	AGAGAGCGCTTGCGTCCTT	TTGCCTTTGGTTTTTCGGTAA	61	101.40%
hif1a	AACAACGCAAACAAATCCT	GTCACCTCAACCTCCTC	60	93.20%

Table 4.1. Forward and reverse primer sequences for regulatory proteins and downstream targets in zebrafish used in RT-qPCR analysis. Annealing temperatures (Ta), product size (base pairs), efficiency values and accession numbers. *rpl8* , *nfkb1*[17], *nrf2b*[18]*, keap1b*[19] primer sequence was derived from the literature.

Total RNA was extracted from whole homogenised zebrafish samples using TRI reagent following the manufacturer's guidelines with quality and concentration assessed using the NanoDrop ND -1000 Spectrophotometer (NanoDrop

Technologies, Wilmington, USA). 0.2 µg and 1 µg of total RNA from the developmental time series and exposure samples, respectively, were treated with RQ1 RNase-free DNase (Promega, Southampton, UK). The higher overall RNA yields in the exposure samples allowed to input more RNA. Complementary DNA (cDNA) was synthesised with M-MLV reverse transcriptase (Promega, Southampton, UK) and random hexamers (Eurofins Genomics) according to the manufacturer's instructions. cDNA was diluted 1:2 in HPLC water and stored at -20 °C.

Samples were run in triplicate using SYBR green chemistry (Bio-Rad Laboratories, Hercules, CA, USA) and an iCycler iQ Real-time Detection System (Bio-Rad, Hercules, CA, USA). A negative control was run on each plate to confirm the absence of cDNA contamination.

The housekeeping gene, *ribosomal protein 8 (rpl8)*, was selected as the PCR reference target based on previously reported stable expression patterns throughout development and under both silver and oestrogenic exposure in zebrafish[20,21]. Relative transcriptional levels were determined as follows:

$$RE = (E\ Ref)^{Ct\ Ref} / (E\ Target)^{Ct\ Target}$$

RE : Relative gene expression, Ref = Housekeeping gene, target = Gene of interest, E = PCR amplification efficiency

4.3.5. Acridine Orange (AO) staining as an indicator of apoptosis:

Acridine orange-hemi(zinc chloride) salt was purchased from Sigma Aldrich. An AO stock solution was made at a concentration of 1 mg/ml in milliQ water and stored at 4°C in the dark. On the day of the exposure, a working stock of AO was made at 1:100 dilution in ISO water. 10 larvae from each treatment were incubated for 45 min in 5 ml of AO in a 12 well-plate in the dark to avoid fluorescence quenching. Larvae were washed thoroughly with ISO water three times before being anesthetised with 0.4% tricaine and mounted on glass bottom dish in 0.07% low melting point agarose. Whole embryo images were taken using an ANDOR Zyla SCMDs camera on a light microscope (Olympus SZX16) with a Lumen 200 fluorescence illumination system under GFP and

RFP excitation. Images were analysed for fluorescence intensity using the ImageJ software. To quantitatively assess fluorescence intensity, average pixel brightness was measured within a standardised area for each identified target tissue (the hindbrain, gill, jaw and whole body). For each image, normalised intensity measures were determined through subtracting an average image background reading from five measurements.

4.3.6. Statistical Analysis:

Statistical analysis for all datasets was conducted using R studio (version 3.3.42). For all qPCR datasets, outliers were removed following Chauvenet's criterion[22] prior to further analysis. All data was tested for normality and homogeneity of variance before selection of parametric or non-parametric tests as appropriate.

Data that did not fit the standards of normality were transformed using Box-Cox transformations identifying the best dependant power transformation using the value of lambda. The stability of *rpl8* expression between control conditions and tBHQ exposure and over developmental time was assessed using a one-way ANOVA. For time series and acridine orange results, normalised datasets were assessed with a one-way ANOVA or Kruskal-wallis test where appropriate. Post-hoc tests on data that met the significance level in an ANOVA were conducted using pairwise-t-test using Bonferroni-holm p-adjustment correction. Non-parameteric tests were conducted using a Wilcoxon-signed rank test. For chemical exposures, generalized linear models using a gamma family and log link function were used to determine the effect of concentration, exposure time and when appropriate day, across experimental sampling points. Minimum-adequate model testing was carried out to select the final model. For all data, significant effects are determined where $p < 0.05$ and are represented graphically as the mean ± SEM.

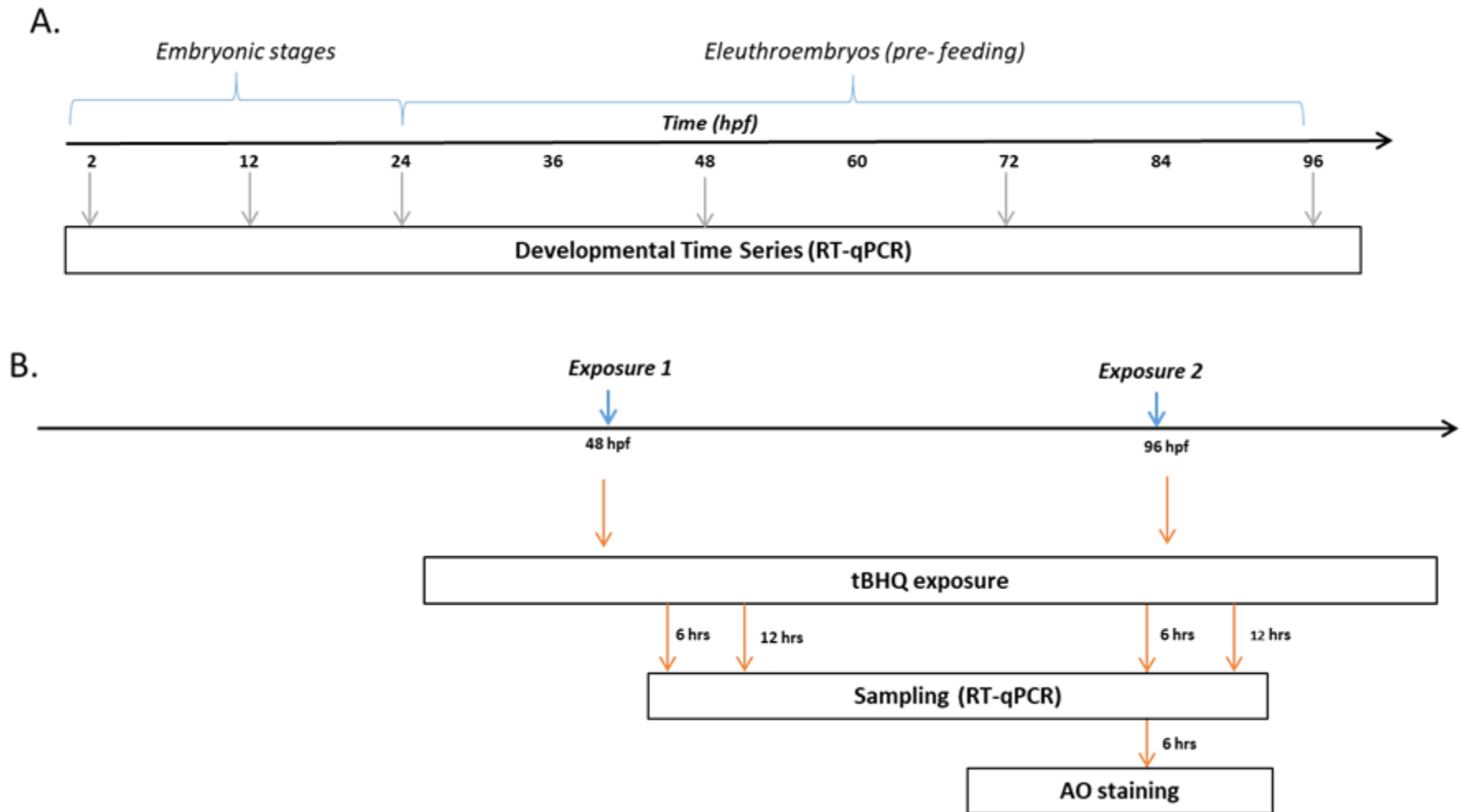**Figure 4.2. Schematic of experimental design**. A) In the first instance developmental expression of *nrf2a, nfkb1, hif1a* and *mtf1* was determined through RT-qPCR. B). The most responsive life-stage was selected following targeted exposures at 2 and 4 dpf. The acridine orange assay that indicates pre-apoptotic and apoptotic cells was used as an indicator of adverse effects of the exposures.

4.4 Results

4.3.1. Evidence of shared gene response pathways under exposures to selected pollutants in fish

The literature search compiled representative experimental studies that reported gene-expression responses across fish species under traditional inducers of AhR, HSF1, HIF-1α, MTF1, Nrf2 and ER pathways. The majority of studies were from laboratory strains of zebrafish (Table 4.2).

The literature search investigated two chemical pollutants that are predicted to act through AhR pathways at MIEs (Table 4.2). These represented an acute exposure to 2,3,7,8- Tetrachlorodibenzodoxin (TCDD) and a chronic exposure to Benzo(a)pyrene (BaP), where the progeny of adult zebrafish exposed to the chemical were investigated. Both studies showed significant fold increases in *cyp* genes, *nrf2a* and *nrf2b*, although this was development stage dependant; *nrf2a* and *nrf2b* were elevated in the 2 and 5 dpf treatments in TCDD[13] but at 4 dpf under BaP treatment[23]. This was coupled with an increase in expression of antioxidants for the BaP exposure but not for TCDD, where, in contrast to the Boolean model predictions (Figure 2.4), there was no increase in expression of the genes for the selected antioxidant targets[23],[13]. No other genes identified in the Boolean model (Figure 2.4) were differentially expressed in the RNA seq dataset for BaP exposure, but *nos1*, which acts similarly to *nos2* - a target in the Boolean model, and involved in cell-cycle arrest - was downregulated under exposure conditions[23].

Experimental evidence from hypoxia representing HIF1 activation were taken from exposures in adult flounder, zebrafish and ruffe (Table 4.2). These showed tissue-specific gene expression profiles for antioxidant defence genes. For example, in the *Platichthys flesus* (flounder), there was a significant increase in *hsp70* and *sod2* in the gills, as predicted from the model (Figure 2.4) but a decrease in *hsp70* in the heart[24]. Transcriptomic analysis of zebrafish exposures showed differential expression of *bcl2, mt2, gstp* and *mt* genes but a down-regulation of glycolytic processes with increasing levels of hypoxia,

therefore only partly supporting the modelling results for HIF1 activation (Figure 2.4).

For heat-stress, representing the activation of HSF1, studies were taken across zebrafish larval and adult exposures (Table 4.2). In adult exposures, there was no significant increase in antioxidant defence genes but an increase in *hsp70* in adult liver tissues after 7 day exposures to 32 degrees [25]. In larveal zebrafish, 48 hrs of cold stress initiated an increase in redox sensitive genes, metallothioneins and heat-shock chaperones, in keeping with the predicted modelling results in Figure 2.4. However, this contrasted with heat-stress where higher temperatures led to an increase in immune responses and cell death, which more closely resembles the response to p53 and NFkB activation (Figure 2.4) [25].

Representative studies for MTF1 activation covered a range of metals including cadmium, silver ions and silver nanoparticles (Table 4.2). Transgenic zebrafish containing a reporter construct for the heat shock element upstream of green fluorescent protein (GFP) showed an increased expression under exposures to $CuSO_4$ [26], indicating binding to the heat shock element under exposure conditions. Cadmium exposure induced a significant increase in *sod1, cat, hsf2* and *mtf1* at 2.5ug/l, in keeping with the predictions from the model (Figure 2.4), but a decrease in *nrf2* [27]. However, under 5ug/l Cd, there was no significant difference in the levels of gene expression compared to the control with the exception of *mt,* which was decreased[27].

RNA sequencing of exposures to Silver ions, bulk and NP contrasted with the model predictions (Figure 2.4) showing a general decrease in the levels of antioxidant genes that were differentially expressed from the control conditions[28]. This was the case for *gsta1* expression under NP and bulk exposures and for *fth1*, *hsp90*, *hif1a*, *sod1* and *mt2* under ions at 24 hpf[28]. At 48 hpf, *fth1* expression was significantly decreased under bulk and ion exposures but the levels of *gsta1* were significantly increased[28]. These results differed from acute exposures in zebrafish embryos which supported the model (Figure 2.4) and showed increases in *gstp*, *mt2* and *gstm1* expression at 24 and 48 hpf under citrate covered AgNPs[28].

The effects of the Nrf2 inducers diethyl malonate (DEM) and tBHQ were similar in zebrafish larvae and adults [13]. Increases in all targeted genes involved in antioxidant defence processes were observed under all exposures, the expression of which was significantly reduced in exposures using both Nrf2 morpholino (MO) and Nrf2 mutant lines [13,35] .

Adaptive stress response factors were predicted to be activated in regulatory cascades initiated by oestrogen receptor agonists (Figure 2.6). For pollutants known to cause endocrine disruption, selected studies covered chronic exposures in adult zebrafish and acute exposures in zebrafish larvae. Roundup and its active ingredient, glyphosate, which has been shown to activate ER-alpha [29], resulted in differences in gene expression of selected adaptive-stress response targets in the gonads of males and females[30]. In testis, there were significant differences in the expression of *sod1* and *cat* between treatment groups and these genes were down-regulated at 0.5 mg/l but upregulated in 0.05 and 10 mg/l roundup and 10 mg/l glyphosate exposures[30]. Their expression dynamics appeared to match those of cytochrome genes *cyp17a1* and *cyp11a1*, though this was not statistically significant. In the ovaries, there was no significant difference in expression of selected targets across the exposure range. Again, although not significant, there was an apparent increase in *gstp* from 0.05 – 10 mg/l in roundup treatments and in glyphosate at 10 mg/l[30]. This was coupled with an increase in *sod1* expression at 0.01-0.05 mg/l but a decrease in expression at 10mg/l under roundup[30]. There was a decrease in *cyp17a1* and *cyp11a1* in the ovary but an increase in *cyp19a1* in comparison to the control at 10mg/l roundup[30].

Acute larvae exposures to bisphenol-A (BPA) in zebrafish embryos between 48 hpf to 120 hpf showed concentration -dependent increases in *cyp1a1, gstp1, gstp2, cyp25b1, gstm.3* and *gstr* but a decrease in genes involved in glycolytic processes[31], therefore again only partly supporting modelling results (Figure 2.6). Exposures of fathead minnows to BPA, Bis(2-ethylhexyl) phthalate (DEHP) and Nonylephenol (NP) showed significant increases in *cyp1a1* and *gsta* under all exposures after 21 days with significant enrichment in the gene expression associated with the AhR pathways. These results from studies on endocrine

disrupting chemicals showed that induction was both dose and time dependent, with apparent hermetic response trajectories shown in the testis for antioxidant response genes. However, given that the model presented in Figure 2.6 shows a cyclic attractor, this means that gene expression is stabilised in a dynamic state; genes are not continuously activated in the attractor state for cascades initiated by oestrogen receptor alpha (ERa) and Era + oestrogen receptor beta (ERb) activation. The results from estrogen exposures therefore show some coherence with the modelling results.

Mixture exposures focused on the effect of metals under heat stress and hypoxia as well as mixtures of pharmaceutical compounds. Heat stress caused significant decreases in the expression of *sod1, hsf1* and *hsf2* at 36 degrees and a decrease in *nrf2* under cadmium exposures[27]. In the absence of heat stress, cadmium induced an increase in *mt* gene expression and the expression of *mtf1* and *hsp70* at higher concentrations[27].

The combination of hypoxia and copper caused a general level of decrease in the levels of antioxidants expressed under mixtures compared to copper alone[32]. Similarly, there were marked differences in Tributylitin (TBT) and BaP with a decrease in expression of AOs under mixtures compared to control conditions[33].

According to the Bradford Hill criteria [34], these results showed a moderate level of confidence in the model generated in Chapter 2 according to the guidance for assessing confidence in overall AOPs. The empirical evidence demonstrated expected changes in gene expression responses in most cases with some evidence inconsistent with predictions.

	Weight of evidence for canalisation							
	Chemical	Hyp. MIE	Exposure conditions	Species	Tissue	Target Genes	Effects	Ref.
PAHs	TCDD	AhR	2 nM at 1 ,2, 3, 4, 5 and 6 dpf for 6 hrs	*Zebrafish (TL line)*	*Whole body (larvae)*	*gstp, gclc, sod1, cyp1a, nqo1, nrf2a, hsp70, gadd45a*	Significant ↑ *cyp1a* at all time points and *nrf2a* at 2 and 5 dpf.	[13]
	Benzopyrene (BaP)	AHR, ER	Adults exposed to 50ug/l BaP for 7 days. Collected embryos exposed to 50ug/l BaP or control conditions and sampled at 3.3 and 96 hpf	*Zebrafish (AB line)*	Whole body (larvae)	RNA-seq	96 hpf ↑: *cyp1c1, cyp1c2, gpx1b, gsta1, gdtp1, hsp70, hsp90, nrf2a, nrf2b, prdx1, mt2, sqstm1* 96 hpf ↓: *nos1*	[23]
endocrine disruption	Glyphosate (Roundup)	unknown	Adult exposures to 0.01, 0.5, 10 mg/l Roundup for 21 days.	*Zebrafish (WIK line)*	Ovary/Testis	*cat, gstp1, sod1, gpx1a, cyp19a, cyp17a1, cyp11a1, esr2a, esr2b, esr1a*	No significant difference in levels of antioxidants but *gstp*↑ under 10 mg/l roundup. ↓ in *esr1a* (Ovary). In the testis,↑ in *cat* and *sod1* at high levels of R.	[30]
	Bisphenol- A (BPA), Di(2-ethylhexyl) phthalate (DEHP), 4-nonylphenol (NP)	ER	Adult exposures to 100ug/l of each chemical for 21 days.	*Fathead minnow (pimephales promelas)*	Liver (male)	Microarray + RT-qPCR	↑ *gsta1* and *cyp1a1* (RT-qPCR) under all exposure conditions	[35]
	Bisphenol- A (BPA)	ER	Exposures to control, 0.1, 1 and 4mg/l from 48 -120 dpf	*Zebrafish*	Whole body (larvae)	RNA-seq	↑ from low to high expression: *cyp1a1, gstp1, gstp2, cyp26b1, gstm.3, gstr.* ↓ from low to high expression: *gpid, pkr, pkmb, aldoaa, pdha1a, pgk1* (glycolytic processes)	[31]
Hypoxia	Hypoxia	HIF1A	mild (5.2 mg/l), moderate (3.6 mg/l), servere (1.5 mg/l)* (Laboratory and field)	*Ruffe (Gymnocephalus cernua)*	Gills, brain, heart	*hif1a, hsp70, sod2, gpx*	↑ *hsp70* and *sod2* in the gills. ↓*hif1a* and *hsp70* under severe hypoxia in the brain. In the field, ↓ *hsp70* under moderate hypoxia in the gills but ↑ in the heart. *Sod2* ↓ in the heart.	[24]

258

	Stressor	TF	Exposure	Organism	Tissue	Method	Findings	Ref
Heat stress	Hypoxia	HIF1A	mild (5.2 mg/l), moderate (3.6 mg/l), servere (1.5 mg/l)* (field)	*Flounder (Platichthys flesus).*	Gills, brain, heart	*hif1a, hsp70, sod2, gpx*	↓ *hif1a* and *sod2* in the gills. ↓ in *hif2a* in the brain and *hsp70* in the heart	[24]
	Hypoxia	HIF1A	Adult exposure from 40% to 10% air saturation for 25 days.	*Zebrafish*	Heart	Microarray	↑ in *bcl2, mt2, gst* and *mt* ↓ metabolic processes (downregualtion of *g6pd*).	[25]
	heat stress	HSF1	Adults exposed to 26°C and 36°C for 7 days.	*Zebrafish (AB line)*	Liver	*nrf2, sod, cat, hsf1, hsf2, hsp70, mtf1, mt, IL-6, il-1b, nfkb,*	*hsp70* ↑ 26°C to 36°C but there was no signficant change in expression of any other genes.	[27]
	heat stress	HSF1	Larvea exposed to temperature stress at 96 hpf (16°C, 28°C, 34°C) for 2 – 24hrs.	*Zebrafish (AB line)*	Whole body (larvae)	Microarray	↑ 16°C at 48 hpf: redox homeostasis, metal ion processes, response to heat. ↑ 34°C at 48 hpf: immune response, cell death, response to heat. ↓ 34°C at 48 hpf: DNA replication, DNA damage.	[36]
metals/ metal NPs	CuSo4	MTF1	250nm at 72 hpf for 3 hrs.	*Zebrafish (hsp70 promoter - TG)*	Somites	*hsp70*	Increase in gfp indicating increase in *hsp70* gene expression .	[26]
	Cd	MTF1	Adult exposures at 28 degrees to 2.5ug/l and 5 ug/l Cd + for 10 weeks.	Zebrafish *(AB line)*	*Liver*	*nrf2, sod, cat, hsf1, hsf2, hsp70, mtf1, mt, IL-6, il-1b, nfkb,*	↓ *nrf2a* at 2.5ug/l. ↑ *sod1, cat, hsf2* and *mtf1* at 2.5ug/l. ↓ in *mt* expression from 2.5 - 5ug/l	[27]
	Ag NP	MTF1	24 and 48 hpf (10nM AgNP and Ag bulk (0.6-1.6um)	*Zebrafish (WiK line)*	*Whole body (larvae)*	RNA-seq	↑ 24 hrs : *gpx4a, fth1a, hif1a1* ↓ 24 hrs: *gstal*, ↑ 48 hrs : *hsbp1* ↓ 48 hrs: *gsta1*	[20]

259

electrophiles	AgNO3	MTF1	24 and 48 hpf (10nM AgNP and Ag bulk (0.6-1.6um)	*Zebrafish (WiK line)*	*Whole body (larvae)*	RNA-seq	↓24 hrs: *hsp90, fth1, cdkn1a, gadph, hif1a1, hsbp1, prdx1, mt2, sod1, gpx1a, cdkn3, hif1a* ↑ 48 hrs: *hspb1* ↓ 48 hrs: *fth1a*	[20]
	AgB	MTF1	24 and 48 hpf (10nM AgNP and Ag bulk (0.6-1.6um)	*Zebrafish (WiK line)*	*Whole body (larvae)*	RNA-seq	↑ 24 hrs: *hsp70* ↓ 24 hrs: *gsta1, cyp1b* ↑ 48 hrs: *hsbp1, gpx4a, gsta1, hsp47* ↓ 48 hrs: *fth1a*	[20]
	AgNPCi, AgBCi, AgNO3	MTF1	Exposure to AgNPCi (500ug/l), AgBCi (500ug/l)and AgNO3 (20ug/l) at 24, 48, 72, 120 hpf	*Zebrafish (WiK line)*	*Whole body (larvae)*	*gstp, fth1, hmox1, gstm1, mt2*	All genes ↑ yolk sac at 24 hpf except *hmox1* and *fth1*. *mt2* and *gstp1* ↑ in the head under AgNP at 48 hpf and at 96 hpf. *gstm1* ↑ in the head at 96 hpf under all treatments. *gstm1* and *fth1* ↓ at 120 hpf in the head but ↑ in the yolk sac.	[28]
	tert-butyl hydroquinone (tBHQ)	Nrf2	10 uM at 1 ,2, 3, 4, 5 and 6 dpf for 6 hrs	*Zebrafish (TL line)*	*Whole body (larvae)*	*gstp, gclc, sod1, cyp1a, nqo1, nrf2a, hsp70, gadd45a*	↑ in *gstp* (8 & 10 fold) at 1 and 2 dpf, *gclc* at 1, 2 and 4 , *nrf2a* at 1 and 2, *hsp70* (fold increase >200) and *gadd45a* (fold increase >10) at all time points	[13]
	H2O2	Nrf2	1mM H2O2 for 6 hrs at 5 dpf	*Zebrafish (AB strain + Nrf2 mutant line)*	*Whole body (larvae)*	*sod1, prdx1, txn1, gpx1b*	Exposure to H2O2 in wild-type and mutant lines. Under H2O2, ↑ *prdx1, txrn1, gclc* and *gpx1* under all conditions with lower increases in *nrf2* mutant compared to Wt.	[37]
	diethylmaleate (DEM)	Nrf2	100uM DEM	*Zebrafish (AB line)*	Whole body *(larvae)*	*gstp, gsta, mgst3b, prdx1, frrsc1, gclc, gclm, hmox1, txnrd1,*	All genes ↑ under DEM exposures but this was prevented in an Nrf2 MO.	[38]

260

Mixtures	Cd + heat stress	MTF1	Adult zebrafish exposed to 2.5ug/l, 5ug/l of Cd under normal (26°C) (for 10 weeks) prior to heat-stressed (36°C) conditions for 7 days.	*Zebrafish (AB line)*	Liver	nrf2, sod, cat, hsf1, hsf2, hsp70, mtf1, mt, IL-6, il-1b, nfkb,	↓ 36°C at 2.5ug/l : *nrf2, sod1, hsf1, hsf2* ↓ 36°C at 5ug/l : *nrf2* ↑ 26°C at 2.5 ug/l: *mt* ↑ 26°C at 5 ug/l : *hsp70, mtf1* .	[27]
	Hypoxia + Cu2+	HIF1a, MTF1	embryo exposures at 4-28, 28-52, 52-76 and 76-100 hpf to 0.024 mg Cu/L under hypoxia (43.2% 0.55 air sat.) or Normoxia (98.9% 0.22 air sat).	Zebrafish (WIK line)	Whole body (*larvae*)	mt2, cat, sod1, gstp1, gsta1, gpx1a	↑ 28-52 hrs normoxia: *gstp1* ↓ 28-52 hrs: *cat* Sig. ↑ *mt2* and gstp1 between normoxia and hypoxia. ↑normoxia: *gpx1aa,* gsta1 *and gstp*. Hypoxia ↓ sig. expression under Cu2+ exposure compared to normoxia.	[32]
	tributylin (TBT) and benzo[a]pyrene (BaP).	ER, AhR	500ug/l BaP, 10ng/l TBT or 0.03 nM TBT + 2uM BaP from 0-0.5 hpf to 72 hpf.	Zebrafish (TU line)	Whole body (*larvae*)	*gstp1, cyp1a, cyp1c, gpx1b, gpx1a.*	↑ *gstp1, cyp1c* and *cyp1a* in BaP but not TBT. ↑ *gstp1* in BaP + TBT but this was lower than in BaP alone.	[33]

* mg/l O2.** gene expression measurements by whole mount in situ hybridisation (WISH).

Table 4.2. Gene expression profiles across teleost fish species under adaptive-stress response inducers. Evidence from teleost fish exposures to stress-response inducers hypothesised to act through AhR, ER, HIF1a, HSF1, MTF1 and NRF2 at molecular initiating events. Results are through RT-qPCR, RNA-seq, microarrays or Whole-mount in situ hybridisation (WISH) as indicated.

4.4.1. Expression of stress-responsive transcription factors over developmental time:

To establish the ontogeny of expression of *mtf1*, *hif1a* and *nfkb1* in comparison with *nrf2* in zebrafish RT-qPCR was conducted at 2, 6, 12, 24, 48, 72 and 96 hours post fertilisation (hpf). The transcript level of *nrf2b* was below the level of detection in the ontogeny analysis. For all genes, the lowest expression levels recorded were between 12 and 24 hpf (statistically different to all other time-points for *mtf1*; Figure 4.3).

Transcript levels of *nrf2a* decreased from 2 to 12 hpf before increasing to 96 hpf (Figure 4.3). Expression of *hif1a*, initially increased from 2 to 6 hpf and relative levels were significantly lower at 12 hrs (Figure 4.3). *hif1a* was expressed at a higher level between 24 to 96 hpf, with the greatest change in expression between 72 and 96 hpf. There was no significant difference in expression levels for *nfkb1* throughout time points with a large variability in the data (Figure 4.3). *mtf1* was expressed at the highest level at 2 hpf, and decreased between 2 to 24 hpf and remained low at 96 hpf. The expression of *mtf1* was significantly reduced at 6 hpf compared to 2 hpf and showed a relatively steady state of expression from 48 hpf onwards (Figure 4.3A).

At 2 hpf, transcript levels were higher for *nfkb1* and *mtf1* than *hif1a* and *nrf2a*. The expression of *hif1a* was comparatively higher than *nrf2a* and *mtf1* at 96 hpf. Transcript levels of *nfkb1* were higher than *nrf2a*, *hif1a* and *mtf1* for all time points studied whereas *nrf2a* had the lowest transcript levels (Figure 4.3B).
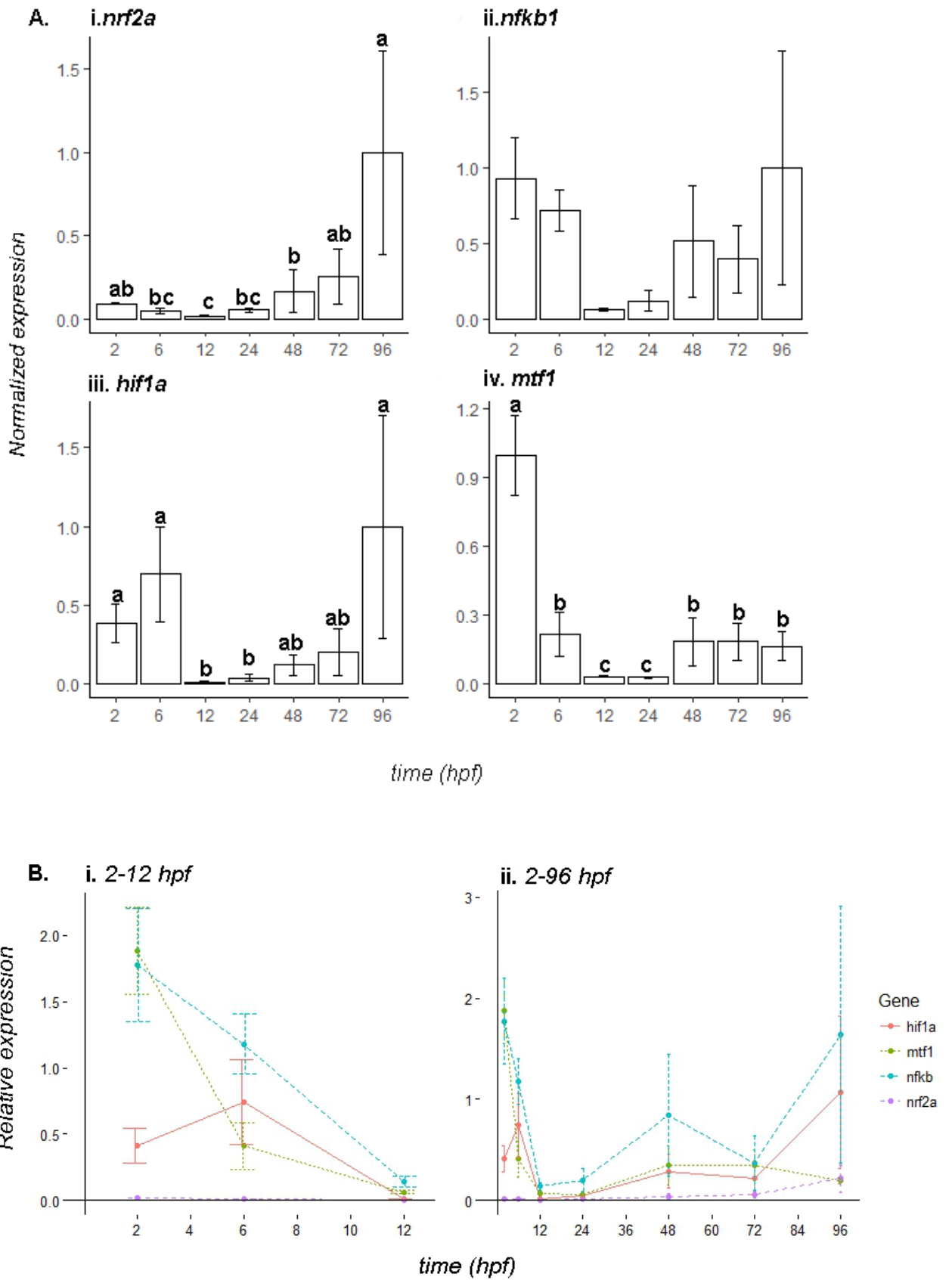
Figure 4.3. Developmental expression of *nrf2a*, *hif1a*, *nfkb1* and *mtf1*.

Zebrafish embryos were sampled in pools of 15 at 2, 6, 12, 24, 48, 72 and 96 hpf (n = 8 -12). RT-QPCR analysis of gene expression was conducted and relative mRNA

expression was calculated by normalising expression to *rpl8*. A). To accurately assess gene expression ontogeny throughout developmental stages, relative transcript abundance was determined and the results were normalised to highest relative transcript level, i). *nrf2a* was normalised to the expression of *nrf2a* at 96 hpf. Ii). *nfkb1* was normalised to the expression of *nfkb1* at 96 hpf. Iii). *hif1a* was normalised to the expression of *hf1a* at 96 hpf. Iv). *mtf1* was normalised to the expression of *mtf1* at 2 hpf. Error bars represent ± SEM with letters indicating significant differences between treatment groups where shared letters show no significant difference (p < 0.05). B). Transcript abundance for all genes relative to *rpl8* across time points at i). 2-12 hpf and ii). 2-96 hpf, respectively.

4.4.2. Expression of gstp1 and sqstm1 between 2 and 4 dpf under tBHQ exposure:

Both *gstp1* and *sqstm1* showed a concentration-dependent increase in transcript levels over tBHQ exposure concentrations (Figure 4.4). The fold change in *gstp1* expression was greatest after 12 hrs exposure at both 2 dpf and at 4 dpf. There was no significant response in gstp expression under exposure to 3 µM tBHQ and DMSO at all time points with the exception of the 6 hr exposure at 2 dpf where there was a 2-fold increase in expression (statistically significant). *gstp1* expression was the highest at 4 dpf where *gstp1* mRNA levels were between 8- and 20-fold higher than in DMSO exposures (Table S4.2).

For *sqstm1*, the highest levels of expression were at 6 hrs post exposure for the 4 dpf life stage, which was 11-fold higher than DMSO. There was no significant difference in the levels of *sqstm1* expression between the lowest (3 µM) tBHQ exposure concentration and DMSO control with the exceptionn of the 12 hr exposure period of the 4 dpf life stage (Table S4.2).

The fold-change in expression was lower at 10 µM exposures for *gstp1* than for *sqstm1* in comparison to control conditions for all exposures tested. There was a significant effect of developmental stage of gene expression for both *gstp1* and *sqstm1.*
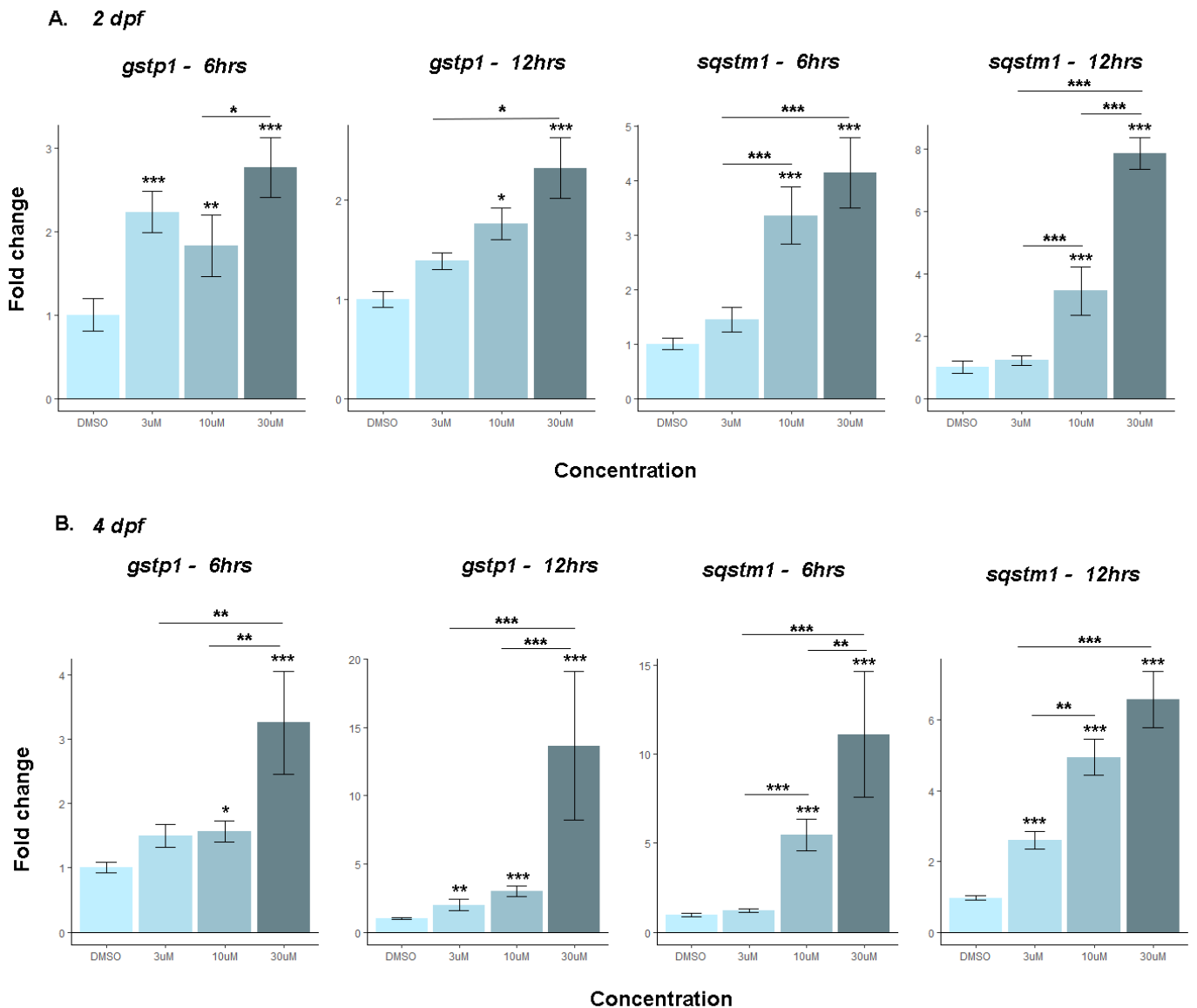
Figure 4.4. Relative expression of nrf2 target genes gstp1 and sqstm1 following 6 and 12 hr exposures to tBHQ at 2 dpf and 4 dpf. Effects of tBHQ treatment at 3, 10 and 30 µM on the gene expression of *gstp1* and *sqstm1* measured using RT-qPCR and normalised to the house keeping gene *rpl8*. Graphs show the mean fold-change relative to the solvent control (DMSO 0.01%) with error bars as ± SEM and significance recorded at 0.05 *, 0.001 ** and 0.0001***. Pooled samples of 30 embryos were exposed in triplicate to each concentration with 3 experimental replicates (n = 5 – 9).

4.4.3. Concentration-dependent expression of stress-responsive transcription factors under tBHQ.

Following the findings on the (relative) transcript levels of *sqstm1* and *gstp1* illustrated in Figure 4, the 4 dpf life-stage was selected for further analysis of stress-responsive transcription factor expression following 6 and 12 hr exposures to tBHQ (Figure 4.5).

Keap1a was the only regulatory protein to show a concentration-dependent response at both time points. Following the 6 hr exposure, there was a significant difference in expression of *keap1a* for exposure to 10 µM and 30 µM tBHQ relative to the control (Figure 4.5A). Gene-transcription of *keap1a* was upregulated between 6 and 12 hr exposures for the 3 µM tBHQ treatment. For the 10 µM and 30 µM tBHQ exposure groups, whilst the average fold change across exposures did not differ significantly, there was a reduced variability across samples between 6 hr and 12 hr time points.

Keap1b showed a variable expression pattern across concentration and time with no significant difference in expression recorded compared to DMSO across all time points (Table S4.3)**.**

Nrf2a showed no significant difference in expression across concentrations following the 6 hr tBHQ exposure. However, the expression of *nrf2a* increased after 12 hr exposure times across all concentrations (Figure 4.5A). For *nrf2b*, there was a high level of variation in expression at 6 hrs with the greatest recorded change at the lowest level (3 µM) of tBHQ exposure in comparison to 10 µM tBHQ, the latter which was not significantly different from DMSO control. After12 hr exposure there was no difference in expression between concentration groups (Table S4.3).

Mtf1 had a significantly lower level of expression after a 6hr exposure to 10 µM and 30 µM tBHQ compared to DMSO. Whilst there was no significant difference in expression of *mtf1* after 6 hr exposure to 3 µM tBHQ compared to DMSO this contrasted with the exposure to 3 µM tBHQ at the 12 hr time point where the highest fold change in gene expression was seen compared with all groups

(Figure 4.5B). Similarly, whilst transcript levels for *mtf1* were downregulated for exposure to 10 µM tBHQ, at 12 hrs there was an upregulation of *mtf1* compared with the DMSO control (Table S4.3).

The gene expression of *hif1a* was higher at 3 µM tBHQ compared with that for 10 µM tBHQ after the 6 hr exposure but not significantly different from DMSO. The variability in expression of *hif1a* was highest in the 30 µM tBHQ exposure after the 6 hr exposurs. After 12 hrs, there was no change in the transcript levels of *hif1a* compared to the control for both the 3 µM and 10 µM tBHQ exposures (Figure 4.5B). The down-regulation of hif1a under 30 µM tBHQ was not significant compared to the DMSO control (Table S4.1)**.**

There was no change in expression of *nfkb1* across the different exposure concentrations following 6 hr exposures to tBHQ. However, after 12 hrs, there was a significant difference in expression for the exposure at 30 µM tBHQ (Figure 4.5B), which was also significantly higher than DMSO control, 3 µM and 10 µM tBHQ exposure groups for this time-point (Table S4.3).

A. Bar charts showing Fold Change versus Concentration (DMSO, 3uM, 10uM, 30uM) for genes nrf2a - 6hrs, nrf2a - 12hrs, nrf2b - 6hrs, nrf2b - 12hrs, keap1b - 6hrs, keap1b - 12hrs, keap1a - 6hrs, keap1a - 12hrs, nfkb1 - 6hrs, nfkb1 - 12hrs, hif1a - 6hrs, hif1a - 12hrs, mtf1 - 6hrs, mtf1 - 12hrs.

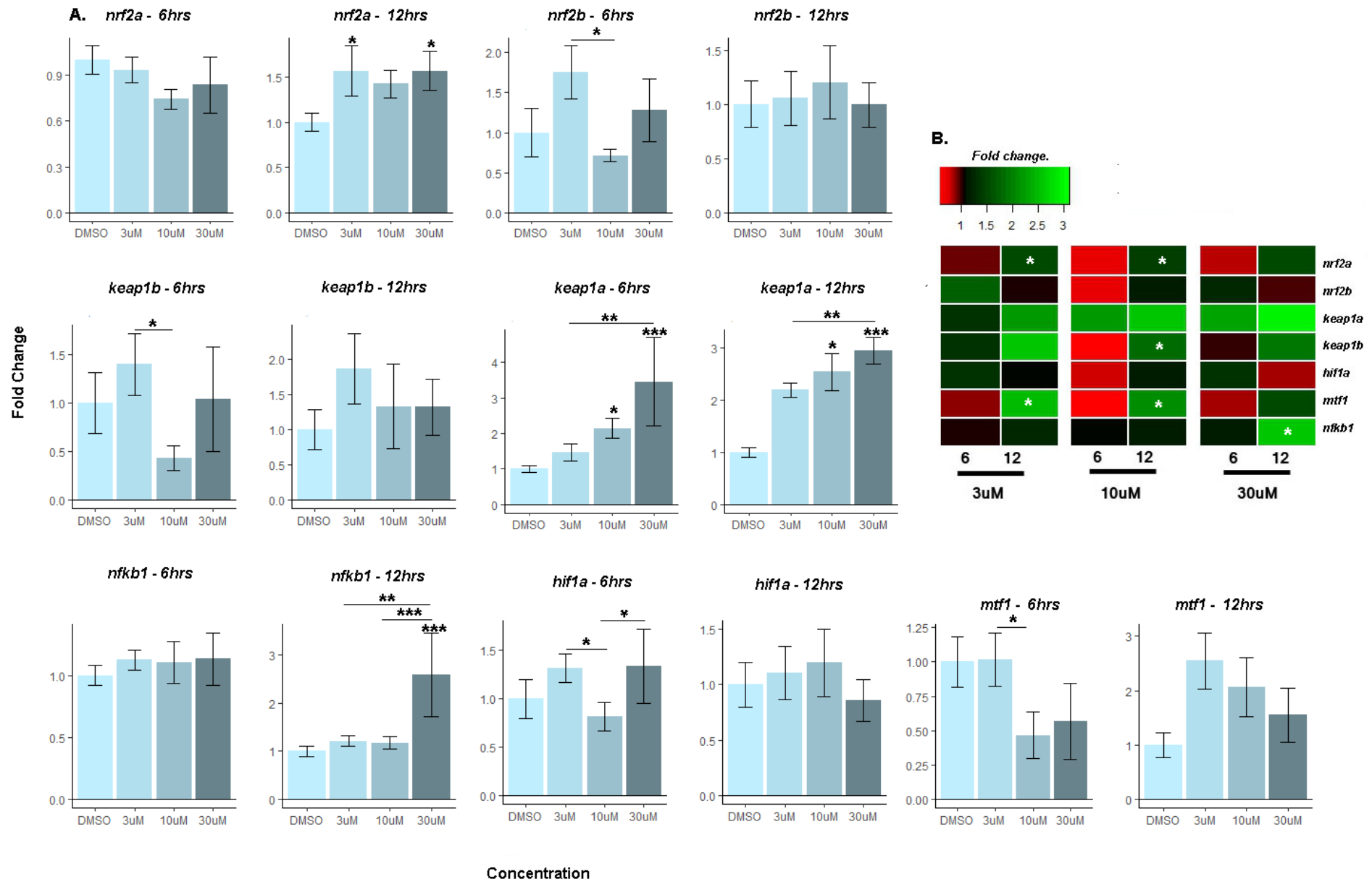B. Heatmap of Fold change for nrf2a, nrf2b, keap1a, keap1b, hif1a, mtf1, nfkb1 at 6 and 12 hrs for 3uM, 10uM, 30uM.

Figure 4.6: Relative expression of genes encoding for oxidative stress associated regulatory proteins after 6 and 12 hr exposures to tBHQ at 4 dpf. A). Effects of tBHQ treatment at 3 µM, 10 µM and 30 µM on the gene expression of *keap1a*, *keap1b*, *nrf2a*, *nrf2b*, *mtf1*, *hif1a* and *nfkb1* measured using RT-QPCR and normalised to the housekeeping gene *rpl8*. Pooled samples of 30 embryos were exposed in triplicate to each concentration with 3 experimental replicates (n = 5-9). Graphs show the mean fold-change relative to the solvent control (DMSO 0.01%) with error bars as ± SEM and significance recorded at 0.05 *, 0.001 ** and 0.0001***. All statistical tests were conducted using generalized linear models B) Heatmap of RT-qPCR results of the expression of genes encoding for oxidative stress associated regulatory proteins as a factor of concentration and time under tBHQ exposure. The mean relative expression of biological replicates (7-9) are indicated as normalized values to DMSO and asterisks indicate significant differences between time points (p<0.05).

4.4.4. Detection of pre-apoptotic and apoptotic cells following tBHQ exposure: AO staining indicating the level of pre-apoptotic cells showed a concentration-dependent increase following exposure to tBHQ (Figure 4.6)*.* These responses appeared to be concentration dependent for the exposures between 3 and 30 µM tBHQ with the jaw, hindbrain and gills identified as target tissues. Pre-apoptotic cells were identified in the jaw after exposure to 3 µM, 10 µM and 30 µM tBHQ. In the 10 µM tBHQ exposure, the occurrence of pre-apoptotic cells increased in the gills and hindbrain. At 30 µM tBHQ exposure, whole body fluorescence intensity indicating apoptosis was significantly different from the control conditions. The expression of apoptotic cells in the gill was significantly higher in embryos exposed to 10 µM and 30 µM tBHQ.
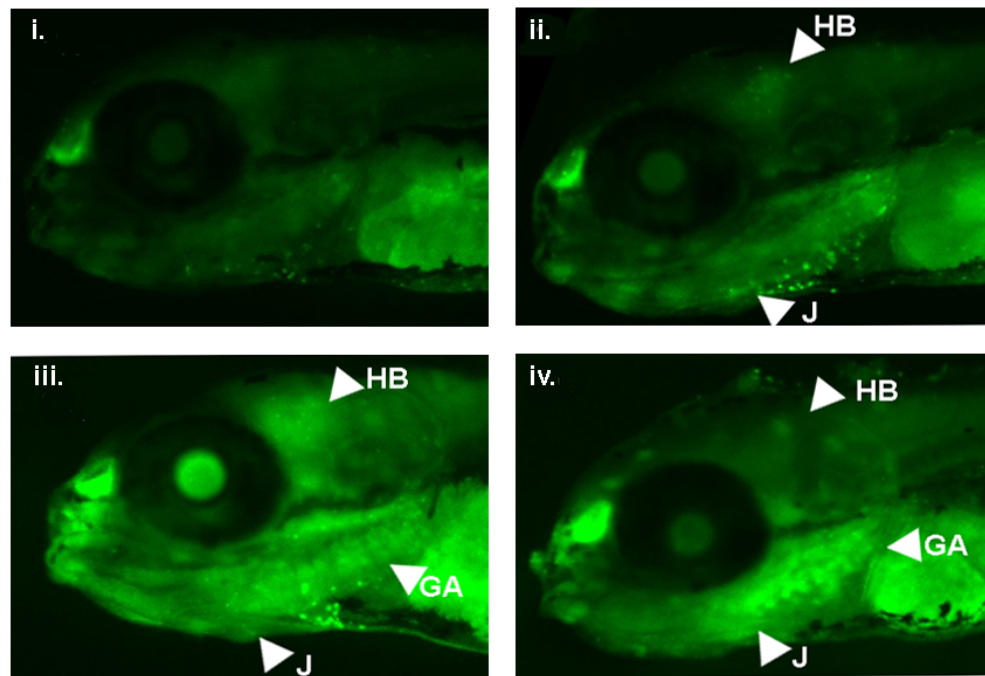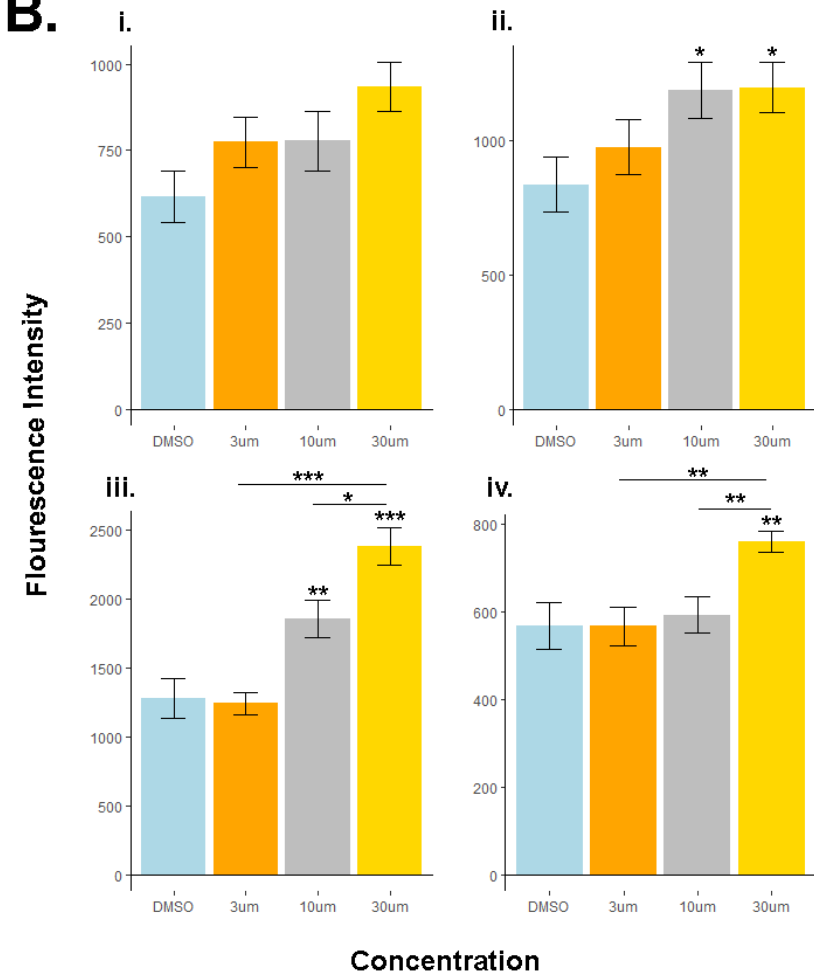
Figure 4.7: Acridine orange staining as an indication of apoptosis in 4 dpf zebrafish eleutheroembryos following 6 hr exposure to tBHQ. A) tissue-specific fluorescence intensity in i) DMSO, ii) 3 µM iii) 10 µM and iv) 30 µM. Fluorescence indicated with an arrow with HB = hindbrain, J = Jaw and GA = gill arches. B) Mean fluorescence intensity of tissue specific and whole-body acridine orange

staining with background readings subtracted (n = 15 for all concentrations except DMSO where n = 10) for i) hindbrain, ii) jaw iii) gill arches iv) whole-body fluorescence. Error bars indicate ±SEM and significance is recorded at 0.05 *, 0.001 **.

4.5 Discussion:

This research sort to identify the level of confidence in the GRN produced in Chapter 2 by compiling representative evidence of gene-expression responses across exposures to traditional inducers of adaptive-stress response pathways. Despite the importance of Nrf2 from both human health and environmental perspectives, little is currently known about the interactions it shares with other regulatory proteins in response to oxidative insult or throughout developmental time. This research provides an ontological expression analysis of key transcription factors that are involved in regulating stress response pathways: *mtf1*, *hif1a* and *nfkb1* in the zebrafish. In addition, the inducible expression of these factors under the Nrf2 inducer tBHQ was investigated to indicate potential regulatory interactions between factors at the transcriptional level.

4.5.1. Validation of modelling outcomes across exposure studies in teleost fish:

The attractor reached by AhR, HIF1a, HSF1, MTF1 and Nrf2 in the boolean model is characterised by the activation of Nrf2, AhR and HIF1a transcription factors and target genes involved in xenobiotic metabolism, antioxidant defence processes, heat-shock chaperones and angiogenesis. The comparisons of gene-expression responses across exposures to chemical toxicants in fish species showed antioxidant and cytochrome p450 genes as widespread biomarkers in targeted gene expression analysis. The expression of genes such as *gstp1, hsp70* and *cyp1b*, which were recorded under exposure scenarios to activators of Nrf2, HIF1a, HSF1 and MTF1, supports the results of the model generated in Chapter 2 and therefore provides evidence of the connectivity between stress-response pathways being conserved in fish species. However, given that not all the presented data conformed to the outcomes of model simulations, this indicates the complexity of stress response pathways with responses showing a dependence on tissue, life-stage and exposure concentration.

The results highlight clear differences in tissue-specific expression patterns of antioxidant defence genes when measured under the same exposure scenarios for tissues such as the gills, liver, heart and brain (Table 4.2). For example, while the target gene *hsp70*, regulated by multiple stress-response factors, was

upregulated in the gills under hypoxia, it was downregulated in the heart. In addition, whole-mount *in situ* hybridisation (WISH) experiments identified tissue-restricted expression patterns of antioxidant defence genes in zebrafish larvae under DEM exposures. Given that antioxidant defence genes have specific functions in mitigating against oxidative stress, this outcome reflects that fact that tissues have different basal levels of antioxidants. As it has also been shown that Nrf2 has tissue restricted expression patterns [38], this could result in different susceptibility to toxicants between tissues, therefore reflecting changes in the regulatory architecture. Additional information, such as the levels of TFs within tissues, should be integrated into the model generated in Chapter 2 in order to provide predictions on specific tissue-level affects.

This result highlights the difficulty in establishing complexities in response processes from whole-organism sequencing methods, where changes in gene expression within individual tissues may not be identified in whole-body samples. The expression of antioxidants was gene-specific in the literature analysis and whilst some targets were upregulated, not all of the predicted targets were differentially regulated.

The representative studies showed differential expression of TFs for BaP, Cd and AgNP exposures where increases in transcript levels of *nrf2a, nrf2b, mtf1, hsf1* and *hif1a* were shown. Interestingly, the downregulation of *hif1a* under Ag bulk at 24 hrs in zebrafish exposures was coupled with a downregulation in *fth1, mt2* and *sod-* all target genes in the boolean model attractor which are predicted to be activated (therefore upregulated) at the same timepoint. This differs from exposures to cadmium where *nrf2* was downregulated but targets *sod1* and *cat* were upregulated along with *hsf2* and *mtf1*. Given that the burden of TFs is different between tissues, this could result in changes in the levels of mRNA transcripts, which do not reflect the level of protein and therefore, the degree of TF activity. In addition, the model does not show an increased regulation of *mtf1* under the predicted simulations with metals and this suggests additional regulatory links are missing.

The most support for the boolean model was provided in responses to cold-stress in zebrafish larvae which showed gene expression responses in

accordance with the modelling outcomes for HSF1 (Figure 2.4). However, gene-expression profiles reached under heat-stress were more similar to the NFkB attractor in chapter 2, showing activation of immune response processes and cell death. Given that the wider literate supports higher levels of stress with NFkB and p53 activation [39], this could suggest that heat stress caused more severe stress-response processes than cold stress. These results show a good level of confidence in the model outcomes.

Of the studies assessed in this chapter, there are marked differences in the levels of gene expression in mixtures compared to single exposures; the results were inconsistent with the modelling outcomes. Under mixtures exposure scenarios, where the MIEs activated are not clearly defined, decreases in the expression of adaptive stress response factors was observed compared to single exposures.

This highlights a key issue when validating the model in Chapter 2; it is difficult to assess similarities in gene-expression profiles across chemical exposures as each pollutant exerts different severities of response depending on dose. There therefore needs to be a continuous measure across exposures such as glutathione (GSH) as a comparative indicator of redox status, which could aid in validating the Boolean model results by setting a standard to which outcomes can be compared with more confidence. In addition, this analysis assumes the MIEs of inducers based on traditional associations - it is not known if the selected chemicals and stressors activate additional TFs upon exposure, which could also account for some of the inconsistencies in the reported results in this chapter and the model outcomes.

4.5.3. Variation exists in the developmental expression of redox related transcription factors nrf2a, mtf1, nfkb1 and hif1:

To begin to understand the roles of redox-sensitive transcription factors during early life development, the expression patterns of *nrf2a*, *hif1a*, *mtf1* and *nfkb1* were established during embryogenesis in zebrafish via RT-qPCR. At 2 hpf, a development stage that is transcriptionally silent and thus a period when the embryo is reliant on the translation of maternally deposited mRNA[16,40], all

transcription factors had higher relative expression levels compared to *nrf2a* . The high transcript abundances of *hif1a, mtf1* and *nfkb1* suggests these genes were maternally deposited and this is supported by previous findings for these factors. For *mtf1*, high relative transcript levels pre-gastrulation have been shown using whole-mount *in situ* hybridisation (WISH) with an oscillating expression pattern from the envelope layer at 2 hpf to the sphere stage[40]. The results of this study showed that *nfkb1* had a high level of expression prior to 24 hpf whereas *hif1a* fluctuated between a high transcript abundance at 2 hpf and a low level at 72 hpf both of which are supported by the results of previous studies [41].

The finding of expression of *nrf2a* only after 24 hpf in this study is again supported by previous reports which show an increasing expression pattern from 24 hpf to 96 hpf [18]. The strong agreement between the results recorded in the study reported here and of those in the literature would indicate little difference between the wild-type strain used in this study with other zebrafish strains (AB and Tupfel/Long fin mutation (TL) strains respectively) in the developmental expression of the selected stress responsive transcription factors.

This study identified significant variation in the mRNA levels between transcription factors relative to *rpl8* over developmental time. Firstly, the levels of *nrf2a* transcripts were low in comparison to all other transcription factors throughout the developmental windows investigated. No developmental role of *nrf2a* has so far been established whereas all other factors have been associated with regulating genes involved in morphogenesis. MTF1 -/- mice are embryonically lethal suggesting the factor is essential in vertebrate development. *hif1a* has been shown to be essential for neural crest migration zebrafish [42] which occurs during somatogenesis, 13-14 hpf [43]. For the NFkB family of transcription factors, essential roles in inhibiting apoptotic process during gastrulation[44], where cell-turnover rates are high, have been shown in addition to regulating targets involved in notochord development[45]. Whilst the levels of *nfkb1* fluctuated, there was no significant difference in expression pattern and a large level of variation between samples within time-points.

Whilst the ability to mount an OSR pre-24 hpf was not assessed, the finding that *nfkb1* and *hif1a*, both of which are able to regulate antioxidant genes in mammals, suggests a potential mechanism for the upregulation of antioxidants under chemical insult when the expression of *nrf2a* is low. The susceptibility to oxidative stress is high in early embryos which develop from reduced to oxidised states between 3 hpf and 48 hpf [46]. This correlates with the requirement for reactive oxygen species (ROS) which act as signalling molecules in early developmental processes[7] but also leads to an increased susceptibility to oxidative stress during early life stages. In this case, *gstp1* has been shown to be maternally deposited but becomes depleted in embryos after 6 hpf where its expression is inhibited by the extra-cellular-signal-related kinase (Erk) – cAMP response element binding protein (Creb) pathway[47]. Both HIF-1α and NFkB pathways can regulate GSTP1, suggesting potential mechanisms to counteract inhibitory affects during early development. However, it is also possible that only low levels of *nrf2a* are necessary to bind to EpRE regions to initiate transcription, providing transcriptional regulation even in early development.

Further research is needed to identify if these redox-associated factors are involved in regulating antioxidant defence genes in embryos prior to the 24 hpf stage. However, it should also be considered that the role of a factor might change over developmental time especially if there is a dependence on specific accessory proteins to regulate the expression of individual genes.

4.5.2. The level of inducible expression of gstp1 and sqstm1 is dependent on developmental stage:

The inducible expression of the Nrf2 targets, *gstp1* and *sqstm1* under tBHQ exposure were analysed as indicators of Nrf2 activation at both 2 dpf and 4 dpf. In both cases, there was a positive correlation between the fold change in *gstp1* and *sqstm1* expression and increasing tBHQ concentration following both 6 hr and 12 hr exposure conditions. This showed the greatest increase in the upregulation of *gstp1* in comparison to DMSO at the 4 dpf period compared to 2 dpf, as well as the highest expression of *sqstm1* following 6 hr exposures in 4 dpf embryos (Figure 4.4). These results are in keeping with the GRN predictions in Chapter 2 (Figure 2.4).

The increase in inducible expression of these antioxidants could be related to total GSH concentration which has been shown to decrease in embryos treated with low-levels of tBHQ at 4 dpf but not at 2 dpf [48] . *gstp1* has been identified as having the highest binding efficiency to GSH compared to other members of the gst family in zebrafish[49] which could therefore result in a greater redox imbalance at 4 dpf compared to 2 dpf where there was a lower fold change between both *sqstm1* and *gstp1*.

The results reported in this study contrast with previous findings which identified that lower doses of tBHQ resulted in a significant induction of *gstp1* expression at 2 dpf and 4 dpf[5]. The reduced expression observed in this study indicates potential variations in AB and WIK strain susceptibility to tBHQ and potentially, oxidative stress, between zebrafish lines.

4.5.3. Paralogs in the nrf2-keap1 pathway under tBHQ had divergent expression patterns:

The inducible expression of Nrf2 and Keap1 paralogs was investigated under tBHQ exposure to establish any changes in the transcriptional regulation of the pathway. This research showed comparatively different expression dynamics of *keap1a* and *keap1b* over both time and increasing concentrations to tBHQ. Whilst *keap1a* showed a concentration-dependent increase in expression at 6 hrs post-exposure, this differed for *keap1b*, which had a stochastic expression pattern (Figure 4.5). This is conflicting with other data reporting that both genes are transcribed at the same level under xenobiotic exposure scenarios [11]*,* [50]*.* However, each factor has been shown to have divergent functional roles and this could relate to the results in Figure 4.4; *keap1a*, but not *keap1b*, is targeted by tBHQ, which could cause a reduction in protein abundance.

Increasing concentrations of tBHQ were coupled with a higher fold change in *keap1a* expression (Figure 4.5) suggesting an inhibitory mechanism for *nrf2a* release under increasing levels of oxidative stress. In support of this, whilst Nrf2 has been identified as being self-regulating[51], this study established no

significant change in expression of *nrf2a* or *nrf2b* at both 6 and 12 hr exposure time periods (Figure 4.5).

As this research showed *keap1a* is upregulated during exposures to tBHQ, this suggests depletion in the levels of free Nrf2 proteins within the organism. However, if this was the case, it was not reflected by decreased levels of the antioxidants *sqstm1* and *gstp1* which both showed a concentration-dependent change in gene expression (Figure 4.4). However, the levels of Nrf2 protein required to initiate a regulatory response under biological stress are unknown and it is possible that small-fluctuations in *nrf2a* induction are sufficient to regulate gene transcription but would not be identified in RT-qPCR of whole-body samples.

4.5.4. Transcript levels of stress-response transcription factors mtf1 and nfkb1, but not hif1a, were affected by tBHQ exposures:

The expression of *mtf1*, *hif1a* and *nfkb1* under tBHQ exposure was assessed as an indicator of regulatory connectivity between Nrf2 proteins. Of these factors, *mtf1* was upregulated at 3 µM and 10 µM tBHQ after a 12 hr exposure period (Figure 4.5). Whilst the mechanisms for the regulation of *mtf1* transcription have not been established, the higher expression levels after a prolonged exposure suggest an increase in the transcription of downstream target genes. These targets include metallothioneins, which are also regulated by Nrf2 transcription factors and have roles in detoxification processes. Given that an upregulation of *mtf1* was also shown in the literature analysis (Table 4.2), this also indicates additional regulatory processes are involved in the expression of *mtf1* that are not included in the model presented in Chapter 2. As *mtf1* was not upregulated under 30 µM for any time point analysed, this suggests that *mtf1* regulation could be influenced by the internal GSH:GSSG concentrations, which would be specific to the time-period and exposure concentration.

This study also identified a greater expression of *nfkb1* following 12 hr exposures at 30 µM tBHQ (Figure 4.5). Interactions between NFkB and Nrf2 have been identified at the protein-level and not at the gene-transcription

level[52], suggesting a downstream mechanism is involved in the upregulation of *nfkb1* at the 4 dpf developmental period. However, as *sqstm1* transcription was upregulated under increasing concentrations of tBHQ, this suggests a mechanism for initiating the NFkB pathway during shorter exposure periods. It is important to highlight that this study only investigated the expression of one NFkB transcription factor and changes in the regulation of other components, including NFkB2, Rela and Relb, in the pathway may occur under tBHQ exposure, leading to changes in the transcriptional regulation of downstream targets regulated by NFkB dimers. The upregulation of NFkB at higher concentrations suggests NFkB activation, and therefore the initiation of a different regulatory cascade to Nrf2 as predicted by Chapter 2 (Figure 2.4). Whole-genome sequencing would indicate if the outcome processes regulated by NFkB (e.g. immune response processes) are activated at higher levels of tBHQ exposure.

There was no significant change in the levels of *hif1a* in the exposure conditions analysed. Increases in *hif1a* transcription have been widely documented under hypoxic conditions across vertebrates *in vivo* [41] which suggests that neither *hif1a* activation or hypoxic conditions occurred under tBHQ exposure. From the modelling results presented in Chapter 2, it is predicted that HIF1a would be initiated in the downstream regulatory processes caused by Nrf2 activation. As changes in hif1a expression where not seen, this does not necessarily indicate that HIF1a was not activated under exposure conditions. It is possible that the levels of *hif1a* transcription caused by tBHQ exposure are low and therefore not identifiable in whole-body samples.

4.5.5. Increases in pre-apoptotic cells following tBHQ exposure:

Pre-apoptotic cells, as indicated by increases in fluorescence intensity, were observed under AO staining in the jaw, hindbrain and gill arches correlating to tissues where genes involved in antioxidant defence and phase II metabolism are expressed in zebrafish embryos. In this case*, gstp1*, and the antioxidant, ferritin 1 (*fth1*), have been shown to be expressed in the gills[38] and hindbrain[28] respectively and these tissues had a significant increase in

fluorescence intensity compared to control conditions in this research. This is supported by the increases in *gstp1* expression under tBHQ exposures as shown in the RT-qPCR analysis. In addition, an increase in fluorescence intensity was observed in the jaw at 10 µM and 30 µM tBHQ, which is a major target tissue for 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD) exposures during development. TCDD is a prototypic inducer of the aryl hydrocarbon receptor which has been established as interacting in a positive feedback loop with Nrf2 in zebrafish[11]. This therefore suggests the induction of antioxidant defence processes occurred within these tissues.

However, fluorescence following AO staining is an indicator of pre-apoptotic cells, indicating that antioxidant defence processes were overwhelmed in the identified tissues. In this case, the induction of apoptosis is pre-empted by the upregulation of inflammatory genes, a process mediated by NFkB. Although this research showed no significant change in *nfkb1* expression until 12 hr exposures at the highest concentration of tBHQ, there was a significant increase in *sqstm1* across all concentrations which is a known induce or the NFkB pathway[53]. This suggests that the protein-protein interaction whereby *sqstm1* activates NFkB signalling pathways through interacting with TNF receptor associated factor (TRAF6) [52] , is not supported by significant changes in the basal transcript levels of the *nfkb1*. These findings indicate that that the NFkB pathway was active at lower exposure concentrations where the fluorescence observed in the gills was not correlated with a change in the expression of *nfkb1*. Given that AO staining indicates pre-apoptotic cells, this is also in line with the outcomes of model simulations with p53 initiated in the start state (Figure 2.4) which was the only regulatory cascade to result in the activation of cell-death response genes.

4.5.6. Further research:

This study established that changes in gene expression of the transcription factors *nfkb1* and *mtf1* occurred under prolonged exposures to the Nrf2 inducer, tBHQ, indicating potential cross-talk with the Nrf2 pathway. In addition, considering that antioxidant genes have been identified to be shared targets in mammals, the higher transcript levels of *hif1a* and *nfkb1* prior to the expression of *nrf2a* at 24 hpf identified in this study indicates a potential mechanism

whereby antioxidant genes could be regulated under oxidative insult in early development. Further research is necessary to fully establish these regulatory links firstly through repeating the chemical exposures and image analysis conducted in this study using an Nrf2 mutant line, as has been developed by Mukaigasa et al [37]. This would establish if the expression of *mtf1, nfkb1* or *hif1a* under chemical induction is altered at 4 dpf in a Nrf2 knock-out, suggesting if compensatory mechanisms for detoxification processes have occurred. WISH analysis could also be used to support the identification of target tissues, allowing greater precision in gene-expression analysis by targeting specific tissues.

The findings reported in this chapter could be further supported by conducting RNA sequencing across selected time-points following the initial induction of Nrf2 to establish changes in transcript expression, and identify progression in regulatory cascades. The Assay for Transposase Accessible Chromatin using sequencing (ATAC-seq), which identifies regulatory regions by indicating areas of open chromatin in the DNA sequence, would further support the establishment of regulatory processes. This technique can identify the regulatory activity of multiple transcription factors by searching the returned DNA for transcription factor binding sites, giving an overall view of the transcriptional regulation at the time-point analysed.
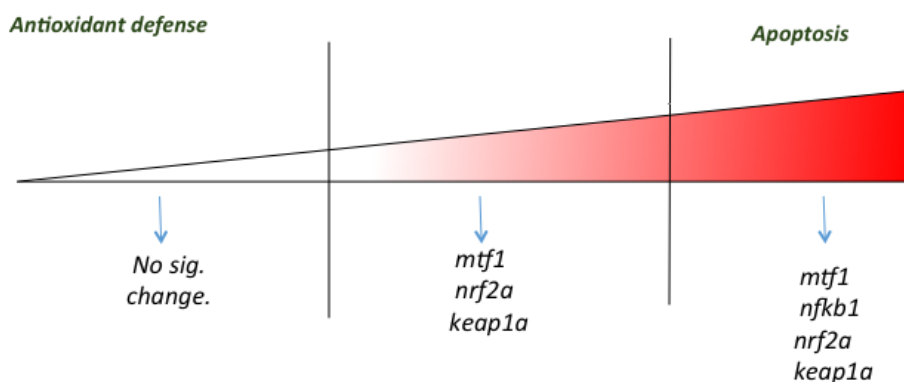
Figure 4.8: Schematic of experimental results summarising changes in transcript expression under tBHQ exposures over time. Changes in transcript expression of selected regulatory targets where red indicates prolonged exposure time periods and higher exposure concentrations.

4.6 Conclusion:

This chapter identified evidence of coherence between stress-response factors and the outcome processes that are reached following induction from a range of inducers in fish exposures. However, there was also a need for more clarity in the modelling results, particularly considering differences in the tissue-expression of target genes. The identification that TF targets such as *mtf1* were upregulated under exposures in the literature suggests additional regulatory interactions have not yet been defined in the adaptive stress response. This was supported by the *in vivo* analysis that showed that the expression of *nfkb1* and *mtf1* are elevated in response to tBHQ after prolonged exposure periods. The observation of pre-apoptotic cells in the jaw, gill arches and hindbrain suggest these as likely target tissues for tBHQ response processes. Further research is necessary to fully establish the regulatory connectivity between pathways. These studies should seek to identify if compensatory responses, leading to the upregulation of transcription factors, occur in the absence of Nrf2, and to establish protein-DNA binding events using whole genome sequencing techniques.

Chapter 4: Identifying the expression dynamics of redox-sensitive transcription factors and downstream target genes under the Nrf2-inducer, tBHQ.

4.7 Supplementary information:

This supplementary information contains:

Table S4.1: Generalised linear model for the relationship between chemical treatment, exposure duration and life-stage (dpf) between RT-qPCR transcript expression.

Table S4.2: Generalised linear model for the relationship between chemical treatment and exposure duration between RT-qPCR transcript expression after 6 hr and 12 hr exposures to tBHQ

Table S4.1: Generalised linear model for the relationship between chemical treatment, exposure duration and life-stage (dpf) between RT-qPCR transcript expression. Transcript expression of *sqstm1* and *gstp1* after 6 and 12 hr exposures to tBHQ (3 uM, 10 uM and 30 uM) at 2 dpf and 4 dpf. normalised to DMSO (0.01%). Minimum adequate models (F-value) is shown with significant codes *p<0.05, **p<0.001, ***p<0.0001. NS= Not significant.

Gene	Concentration	Concentration: Time	Concentration: Day	Concentration :Time :Day
gstp1	3	0.095	0.1163	0.1314
	10	0.3219	0.2396	0.10221
	30	0.3243	0.2698	0.00249**
sqstm1	3	0.16313	0.08614	0.04480*
	10	0.65861	0.375	0.63395
	30	0.00458	0.0019	0.00463**

Table S4.2: Generalised linear model for the relationship between chemical treatment and exposure duration between RT-qPCR transcript expression after 6 and 12 hr exposures to tBHQ. tBHQ exposures at 3 uM, 10 uM and 30 uM at 4 dpf. Coefficient of the minimum adequate model (F-value) is shown with significant codes *p<0.05, **p<0.001, ***p<0.0001. NS = Not significant.

Gene	Concentration	6 hrs	12 hrs	Concentration /Time
keap1a	3	0.375	0.375	NS
	10	0.76*	0.76*	NS
	30	1.24***	1.24***	NS
keap1b	3	0.33	0.63	0.13*
	10	-0.85	0.82	0.04*
	30	0.04	0.27	NS
nrf2a	3	-0.07	0.5	0.51*
	10	-0.3	3.54*	0.65*
	30	-0.18	0.45*	0.75***
nrf2b	3	0.25	0.053	NS
	10	-0.33	0.82	NS
	30	-0.0054	0.0053	NS
hif1a	3	0.18	0.0012	NS
	10	-0.43	0.048	NS
	30	0.42	0.42	NS
mtf1	3	0.02	0.93	0.019*
	10	-0.76	0.73	-1.48***
	30	-0.57	0.44	0.02*
nfkb1	3	-0.12	0.19	NS
	10	0.1	0.16	NS
	30	-0.23	0.95***	-0.82***

4.8 References:

1. Chorley BN, Campbell MR, Wang X, Karaca M, Sambandan D, Bangura F, Xue P, Pi J, Kleeberger SR, Bell DA. 2012. Identification of novel NRF2-regulated genes by ChIP-Seq: influence on retinoid X receptor alpha. *Nucleic Acids Res.* 40, 7416–29. (doi:10.1093/nar/gks409)

2. Kobayashi M, Itoh K, Suzuki T, Osanai H, Nishikawa K, Katoh Y, Takagi Y, Yamamoto M. 2002. Identification of the interactive interface and phylogenic conservation of the Nrf2-Keap1 system. *Genes Cells* 7, 807–20.

3. Li J, Johnson D, Calkins M, Wright L, Svendsen C, Johnson J. 2005. Stabilization of Nrf2 by tBHQ confers protection against oxidative stress-induced cell death in human neural stem cells. *Toxicol. Sci.* 83, 313–28. (doi:10.1093/toxsci/kfi027)

4. Sykiotis G, Bohmann D. 2010 .Stress-Activated Cap'n'collar Transcription Factors in Aging and Human Disease. *Sci Signal* 3, 1–45. (doi:10.1126/scisignal.3112re3.Stress-Activated)

5. Kobayashi M, Li L, Iwamoto N, Nakajima-Takagi Y, Kaneko H, Nakayama Y, Eguchi M, Wada Y, Kumagai Y,Yamamoto M. 2009 .The Antioxidant Defense System Keap1-Nrf2 Comprises a Multiple Sensing Mechanism for Responding to a Wide Range of Chemical Compounds. *Mol. Cell. Biol.* 29, 493–502. (doi:10.1128/MCB.01080-08)

6. Malhotra D, Portales-Casamar E, Singh A, Srivastava S, Arenillas D, Happel C, Shyr C, Wakabayashi N, Kensler TW, Wassermann W, Biswal S. 2010. Global mapping of binding sites for Nrf2 identifies novel targets in cell survival response through chip-seq profiling and network analysis. *Nucleic Acids Res.* 38, 5718–5734. (doi:10.1093/nar/gkq212)

7. Timme-Laragy AR, Goldstone J V, Imhoff BR, Stegeman JJ, Hahn ME, Hansen JM. 2013 .Glutathione redox dynamics and expression of glutathione-related genes in the developing embryo. *Free Radic. Biol. Med.* 65, 89–101. (doi:10.1016/j.freeradbiomed.2013.06.011)

8. Shin S, Wakabayashi N, Misra V, Biswal S, Lee GH, Agoston ES, Yamamoto M, Kensler TW. 2007 .NRF2 Modulates Aryl Hydrocarbon Receptor Signaling: Influence on Adipogenesis. *Mol. Cell. Biol.* 27, 7188–7197. (doi:10.1128/MCB.00915-07)

9.	Miao W, Hu L, Scrivens PJ, Batist G. 2005 .Transcriptional regulation of NF-E2 p45-related factor (NRF2) expression by the aryl hydrocarbon receptor-xenobiotic response element signaling pathway: Direct cross-talk between phase I and II drug-metabolizing enzymes. *J. Biol. Chem.* 280, 20340–20348. (doi:10.1074/jbc.M412081200)

10.	Wang L, He X, Szklarz GD, Bi Y, Rojanasakul Y, Ma Q. 2013 .The aryl hydrocarbon receptor interacts with nuclear factor erythroid 2-related factor 2 to mediate induction of NAD(P)H: Quinoneoxidoreductase 1 by 2,3,7,8-tetrachlorodibenzo-p-dioxin. *Arch. Biochem. Biophys.* 537, 31–38. (doi:10.1016/j.abb.2013.06.001)

11.	Rousseau ME, Sant KE, Borden LR, Franks DG, Hahn ME, Timme-Laragy AR. 2015 .Regulation of Ahr signaling by Nrf2 during development: Effects of Nrf2a deficiency on PCB126 embryotoxicity in zebrafish (Danio rerio). *Aquat. Toxicol.* 167, 157–171. (doi:10.1016/j.aquatox.2015.08.002)

12.	Hahn ME, McArthur A, Karchner S, Franks D, Jenny M, Timme-Laragy A, Stegemen J, Woodin BR, Cipriano MJ, Linney E. 2014. The Transcriptional Response to Oxidative Stress during Vertebrate Development: Effects of tert-Butylhydroquinone and 2,3,7,8-Tetrachlorodibenzo-p-Dioxin. *PLoS One* 9, 11, e113158. (doi:10.1371/journal.pone.0113158)

13.	Harbeitner RC, Hahn ME, Timme-Laragy AR. 2013 .Differential sensitivity to pro-oxidant exposure in two populations of killifish (Fundulus heteroclitus). *Ecotoxicology* 22, 387–401. (doi:10.1007/s10646-012-1033-x)

14.	Wang XJ, Hayes JD, Higgins LG, Wolf CR, Dinkova-kostova AT. 2010 Activation of the NRF2 Signaling Pathway by Copper-Mediated Redox Cycling of Para- and Ortho-Hydoquinones. *Chem. Biol.* 17, 75–85. (doi:10.1016/j.chembiol.2009.12.013)

15.	Zhang DD, Hannink M. 2003 .Distinct Cysteine Residues in Keap1 Are Required for Keap1-Dependent Ubiquitination of Nrf2 and for Stabilization of Nrf2 by Chemopreventive Agents and Oxidative Stress. *Mol. and Cel. Bio.* 23, 8137–8151. (doi:10.1128/MCB.23.22.8137)

16.	Williams LM, Timme-Laragy AR, Goldstone J V, McArthur AG, Stegeman JJ, Smolowitz RM, Hahn ME. 2013. Developmental expression of the

Nfe2-related factor (Nrf) transcription factor family in the zebrafish, Danio rerio. *PLoS One* 8,10, e79574. (doi:10.1371/journal.pone.0079574)

17. Oyarbide U, Iturria I, Rainieri S, Pardo MA. 2015 .Use of gnotobiotic zebrafish to study Vibrio anguillarum pathogenicity. *Zebrafish* 12, 71–80. (doi:10.1089/zeb.2014.0972)

18. Timme-Laragy AR, Karchner SI, Franks DG, Jenny MJ, Harbeitner RC, Goldstone J V, McArthur AG, Hahn ME. 2012. Nrf2b, novel zebrafish paralog of oxidant-responsive transcription factor NF-E2-related factor 2 (NRF2). *J. Biol. Chem.* 287, 4609–27. (doi:10.1074/jbc.M111.260125)

19. Li L, Kobayashi M, Kaneko H, Nakajima-Takagi Y, Nakayama Y, Yamamoto M. 2008. Molecular evolution of Keap1: Two Keap1 molecules with distinctive intervening region structures are conserved among fish. *J. Biol. Chem.* 283, 3248–3255. (doi:10.1074/jbc.M708702200)

20. Aerle R Van, Lange A, Moorhouse A, Paszkiewicz K, Ball K, Johnston BD, Booth T, Tyler CR, Santos EM. 2013 .Molecular Mechanisms of Toxicity of Silver Nanoparticles in Zebrafish Embryos. *Environ. Sci. Technol.* 16, 8005–8014. (doi:10.1021/es401758d)

21. Filby AL, Tyler CR. 2007 .Appropriate 'housekeeping' genes for use in expression profiling the effects of environmental estrogens in fish. *BMC Mol. Biol.* 8, 10. (doi:10.1186/1471-2199-8-10)

22. Chauvenet W. 1863 .A Manual of Spherical and Practical Astronomy: 2: Theory and Use of Astronomical instruments. *Lippincott*

23. Fang X, Corrales J, Thornton C, Clerk T, Scheffler BE, Willett KL. 2015 Transcriptomic changes in zebrafish embryos and larvae following benzo[a]pyrene exposure. *Toxicol. Sci.* 146, 395–411. (doi:10.1093/toxsci/kfv105)

24. Tiedke J, Thiel R, Burmester T. 2014. Molecular response of estuarine fish to hypoxia: a comparative study with ruffe and flounder from field and laboratory. *PLoS One* 9, e90778. (doi:10.1371/journal.pone.0090778)

25. Marques IJ, Leito JTD, Spaink HP, Testerink J, Jaspers RT, Witte F, van den Berg S, Bagowski CP. 2008 .Transcriptome analysis of the response to chronic constant hypoxia in zebrafish hearts. *J. Comp. Physiol. B.* 178, 77–92. (doi:10.1007/s00360-007-0201-4)

26. Seok Sh, Park JH, Baek MW, Lee HY, Kim DJ, Uhm HM, Hong JJ, Na YR, Jin BH, Ryu DY, Park JH, . 2006.Specific activation of the human

HSP70 promoter by copper sulfate in mosaic transgenic zebrafish. *J. Biotechnol.* 126, 406–13. (doi:10.1016/j.jbiotec.2006.04.029)

27. Zhu QL, Guo SN, Yuan SS, Lv ZM, Zheng JL, Xia H. 2017. Heat indicators of oxidative stress, inflammation and metal transport show dependence of cadmium pollution history in the liver of female zebrafish. *Aquat. Toxicol.* 191, 1–9. (doi:10.1016/j.aquatox.2017.07.010)

28. Osborne OJ, Mukaigasa K, Nakajima H, Stolpe B, Romer I, Philips U, Lynch I, Mourabit S, Hirose S, Lead JR, Kudoh T, Tyler CR. 2016 .Sensory systems and ionocytes are targets for silver nanoparticle effects in fish. *Nanotoxicology* 10, 1276–1286. (doi:10.1080/17435390.2016.1206147)

29. Mesnage R, Phedonos A, Biserni M, Arno M, Balu S, Corton JC, Ugarte R, Antoniou MN. 2017 .Evaluation of estrogen receptor alpha activation by glyphosate-based herbicide constituents. *Food Chem. Toxicol.* 108, 30–42. (doi:10.1016/J.FCT.2017.07.025)

30. Webster TU, Laing L, Florance H, Santos EM, 2013. Effects of glyphosate and its formulation, Roundup®, on reproduction in zebrafish (Danio rerio). *Sci. Technol.* 48, 1271–1279. (doi: 10.1021/es404259h./Environ.Sci.Technol.)

31. Martínez R, Esteve-Codina A, Herrero-Nogareda L, Ortiz-Villanueva E, Barata C, Tauler R, Raldúa D, Piña B, Navarro-Martín L. 2018. Dose-dependent transcriptomic responses of zebrafish eleutheroembryos to Bisphenol A. *Environ. Pollut.* 243, 988–997. (doi:10.1016/J.ENVPOL.2018.09.043)

32. Fitzgerald JA, Jameson HM, Dewar Fowler VH, Bond GL, Bickley LK, Uren Webster TM, Bury NR, Wilson RJ, Santos EM. 2016 .Hypoxia Suppressed Copper Toxicity during Early Development in Zebrafish Embryos in a Process Mediated by the Activation of the HIF Signaling Pathway. *Environ. Sci. Technol.* 50, 4502–4512. (doi:10.1021/acs.est.6b01472)

33. Huang L, Zuo Z, Zhang Y, Wang C. 2015 .Toxicogenomic analysis in the combined effect of tributyltin and benzo[a]pyrene on the development of zebrafish embryos. *Aquat. Toxicol.* 158, 157–164. (doi:10.1016/J.AQUATOX.2014.10.024)

34. Becker RA, Ankley GT, Edwards ST, Kennedy SW, Linkov I, Meek B,

Sachana M, Segner H, Van Der Burg B, Vileneuve DL, Watanabe H, Barton-Maclaren TS. 2015 .Increasing Scientific Confidence in Adverse Outcome Pathways: Application of Tailored Bradford-Hill Considerations for Evaluating Weight of Evidence. *Regul. Toxicol. Pharmacol.* 72, 514–537. (doi:10.1016/J.YRTPH.2015.04.004)

35. Zare A, Henry D, Chua G, Gordon P, Habibi HR. 2018. Differential Hepatic Gene Expression Profile of Male Fathead Minnows Exposed to Daily Varying Dose of Environmental Contaminants Individually and in Mixture. *Front. Endocrinol. (Lausanne).* 9, 749. (doi:10.3389/fendo.2018.00749)

36. Long Y, Li L, Li Q, He X, Cui Z. 2012. Transcriptomic Characterization of Temperature Stress Responses in Larval Zebrafish. *PLoS One* 7, e37209. (doi:10.1371/journal.pone.0037209)

37. Mukaigasa K, Nguyen LTP, Li L, Nakajima H, Yamamoto M, Kobayashi M. 2012. Genetic Evidence of an Evolutionarily Conserved Role for Nrf2 in the Protection against Oxidative Stress. *Mol. Cell. Biol.* 32, 4455–4461. (doi:10.1128/MCB.00481-12)

38. Nakajima H, Nakajima-Takagi Y, Tsujita T, Akiyama SI, Wakasa T, Mukaigasa K, Kaneko H, Tamaru Y,Yamamoto M, Kobayashi M. 2011. Tissue-restricted expression of Nrf2 and its target genes in zebrafish with gene-specific variations in the induction profiles. *PLoS One* 6, e26884. (doi:10.1371/journal.pone.0026884)

39. Wardyn JD, Ponsford AH, Sanderson CM. 2015. Dissecting molecular cross-talk between Nrf2 and NF-κB response pathways. *Biochem. Soc. Trans.* 43, 621–6. (doi:10.1042/BST20150014)

40. Hong S-K, Levin CS, Brown JL, Wan H, Sherman BT, Huang DW, Lempicki R a, Feldman B. 2010 .Pre-gastrula expression of zebrafish extraembryonic genes. *BMC Dev. Biol.* 10, 42. (doi:10.1186/1471-213X-10-42)

41. Rojas DA, Perez-Munizaga DA, Centanin L, Antonelli M, Wappner P, Allende ML, Reyes AE. 2007 .Cloning of hif-1α and hif-2α and mRNA expression pattern during development in zebrafish. *Gene Expr. Patterns* 7, 339–345. (doi:10.1016/j.modgep.2006.08.002)

42. Barriga EH, Maxwell PH, Reyes AE, Mayor R. 2013. The hypoxia factor Hif-1 controls neural crest chemotaxis and epithelial to mesenchymal

transition. *Cell Biol.* 201, 759–776. (doi:10.1083/jcb.201212100)

43. Schilling TF, Kimmel CB. 1994. Segment and cell type lineage restrictions during pharyngeal arch development in the zebrafish embryo. *Development* 120, 483–94.

44. Liu X, Huang S, Ma J, Li C, Zhang Y, Luo L. 2009 .NF-kappaB and Snail1a coordinate the cell cycle with gastrulation. *J. Cell Biol.* 184, 805–15. (doi:10.1083/jcb.200806074)

45. Correa RG, Tergaonkar V, Ng JK, Izpisua-belmonte JC, Verma IM, Dubova I. 2004 .Characterization of NF-kB/ IkB Proteins in Zebra Fish and Their Involvement in Notochord Development Characterization of NF-kB/ IkB Proteins in Zebra Fish and Their Involvement in Notochord Development. *Mol. Cell. Biol.* 24, 5257–5268. (doi:10.1128/MCB.24.12.5257)

46. Sant KE, Hansen JM, Williams LM, Tran NL, Goldstone J V., Stegeman JJ, Hahn ME, Timme-Laragy A. 2017. The role of Nrf1 and Nrf2 in the regulation of glutathione and redox dynamics in the developing zebrafish embryo. *Redox Biol.* 13, 207–218. (doi:10.1016/j.redox.2017.05.023)

47. Hrubik J, Glisic B, Fa S, Pogrmic-majkic K, Andric N. 2016 .Erk-Creb pathway suppresses glutathione- S -transferase pi expression under basal and oxidative stress conditions in zebrafish embryos. *Toxicol. Lett.* 240, 81–92. (doi:10.1016/j.toxlet.2015.10.013)

48. Sant KE, Hansen JM, Williams LM, Tran NL, Goldstone J V., Stegeman JJ, Hahn ME, Timme-Laragy A. 2017 .The role of Nrf1 and Nrf2 in the regulation of glutathione and redox dynamics in the developing zebrafish embryo. *Redox Biol.* 13, 207–218. (doi:10.1016/j.redox.2017.05.023)

49. Glisic B, Mihaljevic I, Popovic M, Zaja R, Loncar J, Fent K, Kovacevic R, Smital T. 2015. Characterization of glutathione-S-transferases in zebrafish (Danio rerio). *Aquat. Toxicol.* 158, 50–62. (doi:10.1016/j.aquatox.2014.10.013)

50. Zheng JL, Yuan SS, Wu CW, Lv ZM, Zhu AY. 2017. Circadian time-dependent antioxidant and inflammatory responses to acute cadmium exposure in the brain of zebrafish. *Aquat. Toxicol.* 182, 113–119. (doi:10.1016/j.aquatox.2016.11.017)

51. Kwak M-K, Itoh K, Yamamoto M, Kensler TW. 2002 .Enhanced expression of the transcription factor Nrf2 by cancer chemopreventive

agents: role of antioxidant response element-like sequences in the nrf2 promoter. *Mol. Cell. Biol.* 22, 2883–92. (doi:10.1128/MCB.22.9.2883-2892.2002)

52. Duran A, Linares JF, Galvez AS, Wikenheiser K, Flores JM, Diaz-Meco MT, Moscat J. 2008. The Signaling Adaptor p62 Is an Important NF-κB Mediator in Tumorigenesis. *Cancer Cell* 13, 343–354. (doi:10.1016/j.ccr.2008.02.001)

53. Seibold K, Ehrenschwender M. 2015. p62 regulates CD40-mediated NFκB activation in macrophages through interaction with TRAF6. *Biochem. Biophys. Res. Commun.* 464, 330–335. (doi:10.1016/j.bbrc.2015.06.153)

Chapter 5

General Discussion

5.1 Introduction

Adaptive stress-response pathways were reviewed collectively for the first time in 2009 [1]. Since then, sequencing technologies such as Chromatin Immunoprecipitation (ChIP) and ATAC-seq have significantly advanced our understanding of TF-DNA associations (Figure 1.6) and in doing so, enabled the characterisation of gene regulatory networks (GRNs) that control biological processes. This thesis demonstrates a novel approach for predicting gene-expression outcomes resulting from the activation of adaptive stress-responses by generating and modelling a GRN of interactions between pathways. Chapter 2 of this thesis combined the available knowledge on mammalian adaptive-stress response transcription factor (TF) interactions from experimental evidence. Simulations of this model identified that the nuclear factor (erythroid-derived 2)-like 2 (Nrf2), aryl hydrocarbon receptor (AhR), heat shock factor 1 (HSF1), hypoxia inducible factor 1 alpha (HIF-1α) and metal-transcription factor 1 (MTF1) led to the same outcome processes whereas nuclear factor kappa-light-chain-enhancer of activated -B cells (NFkB) and tumor protein p53 (p53) caused distinct responses (Figure 2.4). The efficiency of transcription factor binding site (TFBS) predictions to identify target genes using mammalian-derived motifs was assessed in Chapter 3 as an indicator of the level of network conservation across teleost fish species. This demonstrated that whilst mammalian-based matrices were able to predict binding sites in expected target genes, fish-specific binding motifs identified putative sequences that were distinct from mammalian predictions. Chapter 4 set out to validate the outcomes of the GRN model generated in Chapter 2 across teleost fish through comparing gene-expression profiles from different exposure scenarios and through *in vivo* exposures to the Nrf2 agonist, tert-butylhydroquinone (tBHQ) in zebrafish early life stages.

This thesis advances the current knowledge of the adaptive stress response in toxicology by:

a) Predicting that Nrf2, MTF1, AhR, HSF1 and HIF-1α activate the same outcome processes through molecular interactions.

b) Identifying that target genes where transcription factor binding sites (TFBS) have diverged in fish-species can be predicted using mammalian motifs.

c) Supporting the GRN generated in Chapter 2 by identifying analogous gene-expression responses across chemical inducers in fish species and providing evidence of interactions between Nrf2 and adaptive stress response factors in zebrafish early life-stages.

This discussion explores how the key findings of this thesis expand our current knowledge of the adaptive-stress response across vertebrates and chemical exposures. In doing so, this thesis highlights the need to re-evaluate the approaches taken when analysing gene-expression datasets from toxicology studies with the aim of integrating this knowledge into adverse outcome pathway frameworks (AOPs).

5.2 False-positive associations could be widespread in adaptive-stress response pathways.

Chemical compounds and physiological stressors are historically associated with activating distinct molecular pathways and causing the regulation of discriminant sets of target genes. This has led to an approach in risk assessment that emphasises a need to identify MIEs from which chemicals are compiled into separate classes based on the TF, and therefore the pathway(s), they activate [2]. Associations are then made from these classes connecting the defined molecular initiating events (MIEs) with observed adverse effects across different levels of biological organisation, such as the tissue and whole organism level. The thesis questions the reliability of this approach by showing that the activation of Nrf2, AhR, HIF1a, HSF1 and MTF1 can cause the same outcome processes to be regulated through molecular interactions (Chapter 2). It also showed that the oestrogen receptor pathway can activate adaptive stress response processes at the molecular level, therefore causing the regulation of genes not traditionally associated with oestrogenic responses as highlighted in the model in Chapter 2 and the experimental evidence presented in Chapter 4.

Risk assessment methods therefore need to adopt an approach that moves away from placing chemical toxicity in the context of a single molecular

pathway, instead viewing toxicity as acting through a broader molecular regulatory network.

In showing this, the study provides evidence that MIEs in adaptive stress response pathways can be incorrectly inferred from gene expression datasets. For example, considering that the activation of antioxidant defence and xenobiotic genes was shown to be a shared outcome in Chapter 2, the AhR and Nrf2—traditional TF regulators for these outcome processes —could be incorrectly defined as the MIEs for specific chemical inducers. This is particularly likely considering that antioxidant defence and xenobiotic response genes were common biomarkers in targeted gene-expression assays collated in the literature, often in the absence of target genes for other outcome processes (Chapter 4). Such a result is perhaps more problematic for p53 and NFkB mediated response processes, where antioxidant defence and xenobiotic genes are active along with outcomes associated with higher levels of stress, such as cell-death and inflammatory responses (Figure 2.4).

The results from the model simulations also highlighted that more information can be gained from establishing regulatory cascades rather than just MIEs independently. Considering that TFs can control the expression of multiple target genes, it is difficult to establish which processes will be activated following chemical exposures. This is particularly the case for TFs such as NFkB where gene expression profiles can be defined as early, middle or late[3]. By modelling interactions within a GRN, this method allows for predictions to be made across gene expression profiles based on time and on interactions with other regulatory factors which are likely to influence the responses of target genes such as through sharing targets between TFs as shown in Chapter 2.

As it is likely that not all regulatory interactions between factors have been characterised for the GRN generated in Chapter 2 particularly for MTF1, which showed unpredicted expression patterns in the *in vivo* tBHQ exposure and experimental evidence in Chapter 4, additional interactions could impact the regulatory cascades shown in model simulations. The GRN generated in Chapter 2 is likely to only be a partial representation of the interactions between stress response processes. Using GRN approaches at the basis of AOPs must

therefore allow for networks to be updated in order for processes and interactions to be better understood. For example, as demonstrated for oestrogen receptors in Chapter 2, additional regulatory interactions with receptor-mediated pathways can exist which can initiate the adaptive stress response network at the molecular level.

5.3 Regulatory cascades in toxicology need to be better integrated into risk assessment methods.

AOPs provide only qualitative predictions of toxicity based on observed responses from molecular, tissue, whole organism and population levels. In doing so, this leaves knowledge gaps in how MIEs can lead to specific adverse effects with little integration of regulatory cascades in AOP frameworks. The GRN model generated in this thesis provides a means of establishing which genes are activated within a regulatory cascade therefore identifying the links between MIEs and key events, an essential part of generating and understanding AOPs.

However, validating the model in Chapter 2 according to the Bradford Hill criteria, which provides confidence levels to AOPs based on the weight-of-evidence (WOE), presented a challenge in Chapter 4. This was in part due to a lack of consistency between sampling and exposure time points. The compiled evidence in Chapter 4 demonstrated that gene-expression responses are reported over varying time-points under acute exposures resulting in different reported outcomes across target genes. Early responses to chemical toxicants are dynamic and this information can be lost through only using a single sampling time point. In some cases, experimental aims set out to identify the first responders to exposures yet there is little consistency in the time frames that are considered suitable for this analysis. For example, in zebrafish early life stages, acute exposure times ranged between 3 to 24 hours (Chapter 4), a window that leads to considerable differences in gene expression profiles as demonstrated when responses differed to tBHQ exposure following 6 and 12 hrs exposures (Chapter 4).

This is problematic as the rate of transcriptional regulation and the associated changes in gene expression are time dependent, and therefore observed changes in gene expression can form either the primary responses to chemical inducers or the latter outcomes of the regulatory cascade. This created a challenge when trying to use pre-existing data to validate the model presented in Chapter 2 in terms of validating the predicted changes in regulatory state. This is important as the stages prior to and including the gene expression responses (the regulatory cascade), which causes an adverse effect in an organism at the tissue level, need to be determined to predict responses computationally across chemical exposures.

5.4 The adaptive stress response GRN in an evolutionary and developmental context

Developmental toxicology is complex with organisms undergoing rapid changes in cellular conditions that increase susceptibility to chemical toxicants [4]. The approach taken in this thesis was based on techniques adopted in developmental biology context, where GRNs are established from prior to knowledge to identify the regulatory factors controlling morphogenesis[5]. GRNs that control early patterning processes are highly conserved, with the same TFs controlling organogenesis across phyla [4]. This section discusses the application of the results from this thesis with the aims set out by the review by Leung et al. 2017, which calls for a need to combine evolutionary and developmental biology to determine toxicity responses across species[4].

The GRN generated in this thesis is not specific to a tissue or cellular condition and this therefore allows its integration with developmental networks, where adaptive stress response TFs are necessary for developmental processes[4]. In fact, NFkB and Bcl2 TFs are considered consensual cell-cell signaling pathways in morphogenesis [4]. Perturbations to developmental GRNs make some tissues at risk from damage if targeted by chemical compounds that activate TFs that are involved in developmental patterning. For example, TCDD, an AhR agonist, can cause jaw deformities in developing zebrafish, which results from AhR regulating the expression of *sox9b* under basal conditions [6]. Given the results in Chapter 2, there is also the indication, for example, that chemicals

that are not AhR agonists can cause the same affect through downstream response trajectories leading to AhR activation. In addition, as the expression of TFs *nrf2a*, *mtf1*, *hif1a* and *nfkb1* in Chapter 4 were significantly different throughout development there is the potential that levels of TF activators could affect the inducible gene expression responses across early life stages.

However, it is necessary to consider how much the model generated in Chapter 2 can be used to predict responses to chemicals across species. Adaptive stress response pathways are highly conserved, evolving prior to the divergence of proteasomes and deuterostomes[4], and much like developmental GRNs, from this alone it can be predicted that the GRN generated in this thesis will be conserved across vertebrates. However, differences in toxicity profiles have been observed from a developmental context for chemicals such as the oestrogen receptor agonist bisphenol A (BPA), something that is predicted to be explained through evolutionary history [4].

With this in mind, it was unsurprising in Chapter 3 that divergences existed in TFBS across fish species compared to mammalian counterparts. Although this did not affect the ability to identify target genes significantly using mammalian motifs, fish-specific TFBS did identify novel sites (Chapter 3). Given the vast evolutionary distribution of fish species, it should also not be assumed that all species will share the same responses to toxicants and validation of the identified TFBS in the analysis in Chapter 3 may provide a means of establishing differences in toxicity response profiles. However, the results from Chapter 4 did show moderate confidence in the model results in terms of biological plausibility and the empirical evidence [2], with few results not supporting the model outcomes. This suggests the network was reasonably conserved from mammals to fish but more explicit evidence is needed at the TF-DNA binding level to confirm this result.

The GRN generated in Chapter 2 goes further in identifying the toxicity responses from an evolutionary context. For example, despite the same adverse outcome being reached, different MIEs are predicted to be activated across distantly related species as a result of developmental drift [4].

Considering that the model in this thesis showed that the activation of different TFs at MIEs could lead to the same outcome processes, the GRN provides a potential mechanism for defining how the activation of different pathways can lead to these responses at the molecular level [4]. As such, the results of Chapter 2 identify canalisation; a theory originally proposed in developmental biology that shows that despite differences in genotype the same phenotype can be reached [7]. Canalisation is a measure of evolutionary robustness [7] and considering the early evolutionary development of adaptive-stress response pathway and the evidence provided in this thesis, this provides strong evidence that the concept can be applied in a toxicology setting. At the time of writing, this is the first known application of this theory to an adaptive stress response and toxicology setting.

5.5 Future Research:

Considering the concept of canalisation, evolutionary conservation and a need to further validate the model presented in Chapter 2, this suggests that data from mammalian cell lines can be used to inform on predictions to chemical toxicants in an adaptive stress response context in fish species. A significant majority of toxicology studies conducted on fish *in vivo* that use gene expression datasets or RT-qPCR as biomarkers do not start with an overarching aim of producing datasets to inform on or be integrated into a computational modelling frameworks. However, this information does exist for exposures (often to pharmaceutical compounds) in mammalian cell lines and databases such as Cmap [8] and TG-GATES [9] provide time-course data across exposure concentrations. However, these datasets are often from exposures conducted in cancer cell lines that have different molecular components to normal cells, such as altered expression levels of p53 [10], which may affect the dynamics of network responses. Despite this, further support for the modelling results presented in this thesis should certainly include a comprehensive analysis of the similarities in gene-expression profiles from the available mammalian cell line datasets. The results from this analysis could then be compared to exposures in fish species as a means of establishing the conservation of adverse outcomes and molecular cascades across inducers.

As discussed in Chapter 4, there is a need to use a general marker for the level of stress, such as the GSH:GSSG content to indicate the redox status, to provide an indicator of the cellular state at the time of sampling. This could then be directly associated with the TFs and gene expression profiles that are active to better determine if responses are analogous across chemical exposures and species as predicted by the model and the literature.

Future research into the adaptive-stress response GRN should be primarily aimed at validating the key links within the network, particularly establishing the influence of dose, time, tissue and life-stage on how responses progress through regulatory cascades. One key step in this will be in advancing the model from a boolean setting to one which is based on ordinary differential equations and therefore able to provide quantitative predictions on gene response outcomes. This is a particularly clear step considering the results of chapter 4 where tissue restricted expression of antioxidants and of Nrf2 were observed.

The generation of an ODE model requires more experimental information than is currently unavailable including TF abundance and rates of transcription, translation and degradation which all influence the levels of gene expression [11]. It is also necessary to establish where TFs are activated within an organism over exposure times and this can be shown through high-throughput screening methods in transgenic zebrafish for the TFs of interest in the adaptive stress response.

5.6 Conclusion

The seminal report "Toxicity testing in the 21st century" by the EPA provided a key step in moving toxicology away from laboratory settings and into computational modelling frameworks [12]. While progress has been made in this context, such as the development of the AOP knowledge base, there are still vast gaps in generating robust predictive toxicology tools. This thesis shows that the GRN for the adaptive stress response can be applicable to AOP development, with pathways and TFBS highly conserved across vertebrate groups.

This thesis identified a number of factors that were analogous between developmental concepts and toxicology analysis. The concept of canalisation, where the same phenotype is shown regardless of the genotype, was predicted for responses controlled by stress response factors in Chapter 2. The model identified common responses and cohesion between outcome processes under adaptive stress response factors, identifying the need to place MIEs in the context of regulatory networks in order to understand responses across inducers and across evolutionary history. In doing so, this thesis identifies a need for using GRN approaches as evidence to support the development of AOPs across species.

5.7 References:

1. Simmons SO, Fan CY, Ramabhadran R. 2009 Cellular stress response pathway system as a sentinel ensemble in toxicological screening. *Toxicol. Sci.* 111, 202–225. (doi:10.1093/toxsci/kfp140)

2. Becker RA, Ankley GT, Edwards ST, Kennedy SW, Linkov I, Meek B, Sachana M, Segner H, Van Der Burg B, Vileneuve DL, Watanabe H, Barton-Maclaren TS. 2015 Increasing Scientific Confidence in Adverse Outcome Pathways: Application of Tailored Bradford-Hill Considerations for Evaluating Weight of Evidence. *Regul. Toxicol. Pharmacol.* 72, 514–537. (doi:10.1016/J.YRTPH.2015.04.004)

3. Iwanaszko M, Brasier AR, Kimmel M. 2012 The dependence of expression of NF-κB-dependent genes: statistics and evolutionary conservation of control sequences in the promoter and in the 3' UTR. *BMC Genomics* 13, 182. (doi:10.1186/1471-2164-13-182)

4. Leung MCK, Procter AC, Goldstone J V, Foox J, Desalle R, Mattingly CJ, Siddall ME, Timme-laragy AR. 2017. Applying evolutionary genetics to developmental toxicology and risk assessment. *Reprod. Toxicol.* 69, 174–186. (doi:10.1016/j.reprotox.2017.03.003)

5. Peter I, Davidson E. 2011. Evolution of Gene Regulatory Networks that Control Embryonic Development of the Body Plan. *Cell* 144, 970–985. (doi:10.1016/j.cell.2011.02.017.Evolution)

6. Jenny MJ, Karchner SI, Franks DG, Woodin BR, Stegeman JJ, Hahn ME. 2009 Distinct roles of two zebrafish AHR repressors (AHRRa and AHRRb) in embryonic development and regulating the response to 2,3,7,8-Tetrachlorodibenzo-p-dioxin. *Toxicol. Sci.* 110, 426–441. (doi:10.1093/toxsci/kfp116)

7. Siegal ML, Bergman A. 2002 Waddington's canalization revisited: Developmental stability and evolution. *Proc. Natl. Acad. Sci.* 99, 10528–10532. (doi:10.1073/pnas.102303999)

8. Subramanian A, Narayan R, Corsello SM, Peck DD, Natoli TE, Lu X, Gould J, Dais J, Tubelli AA, Asiedu JK, Lahr DL, Hirschmann JE, Liu Z, Donahue M, Julian B, Khan M, Wadden D, Smith IC, Lam D, Liberzon A, Tobar C, Bagul M, Orzechowski M, Enache OM, Piccioni F, Johnson SA, Lyons NJ, Beger AH, Shamii AF, Brooks AN, Vrcic A, Flynn C, Rosains J, Takeda D, Hu R, Davidson D, Lamb J, Ardelie K, Hogstorm L, Greenside P, Gray NS, Clemons PA, Silver

S, Wu X, Zhao WN, Read-Button W, WuX, Haggarty SJ, Ronco LV, Schreiber SL, Doench JG, Bitterker JA, Root DE, Wong B, Golub TR. 2017. A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles. *Cell* 171, 1437–1452.e17. (doi:10.1016/J.CELL.2017.10.049)

9. Nyström-Persson J, Natsume-Kitatani Y, Igarashi Y, Satoh D, Mizuguchi K. 2017. Interactive Toxicogenomics: Gene set discovery, clustering and analysis in Toxygates. *Sci. Rep.* 7, 1390. (doi:10.1038/s41598-017-01500-1)

10. Ozaki T, Nakagawara A. 2011. Role of P53 in Cell Death and Human Cancers. *cancers,* 3, 1, 994-1013

11. Parmar K, Blyuss KB, Kyrychko YN, Hogan SJ. 2015. Time-Delayed Models of Gene Regulatory Networks. *Comput. Math. Methods Med.*, 347273. (doi:10.1155/2015/347273)

12. Krewski D, Acosta D, Andersen M, Bailar III J, Boekelhedie K, Brent R, Charneley G, Cheung V, Green S, Kelsey K, Kerkvliet N, Li A, McCray L, Meyer O, Patterson RD, Pennie W, Scala R, Soloman G, Stephens M, Yager J, Zeise L . 2007 Toxicity Testing in the 21St Century: a Vision and a Strategy. *Natl. Acad. Sci.* 13, 51–138. (doi:10.1080/10937404.2010.483176.TOXICITY)