

# **TOXICOGENOMIC BIOMARKER DISCOVERY OF AHR-MEDIATED TCDD-INDUCED HEPATOTOXICITY**

**By**

**Edward Dere**

**A DISSERTATION**

**Submitted to  
Michigan State University  
in partial fulfillment of the requirements  
for the degree of**

**DOCTOR OF PHILOSOPHY**

**Biochemistry and Molecular Biology**

**2010**

## **ABSTRACT**

### **TOXICOGENOMIC BIOMARKER DISCOVERY OF AHR-MEDIATED TCDD-INDUCED HEPATOTOXICITY**

**By**

**Edward Dere**

2,3,7,8 Tetrachlorodibenzo-*p*-dioxin (TCDD) is a ubiquitous environmental contaminant that causes a wide array of species-specific adverse biochemical and physiological responses, including increased tumor promotion, lethality and hepatotoxicity. Most, if not all of the effects elicited by TCDD are due to inappropriate changes in gene expression that are mediated through activation of the aryl hydrocarbon receptor (AhR). Although the mechanism of AhR gene regulation is well known, the full spectrum of targeted genes leading to the subsequent toxicological responses remains poorly understood. The objective of this research was to integrate disparate and complementary toxicogenomic approaches to identify putative biomarkers of TCDD-induced hepatotoxicity that would aide in reducing the uncertainties involved in cross-species and cross-model extrapolations.

*In vitro* microarray investigation of a mouse hepatoma cell line treated with TCDD identified complex temporal and dose-dependent gene expression responses. Comparative analysis with *in vivo* hepatic gene expression responses in mice identified a small subset of conserved genes with biological functions related to xenobiotic metabolism, consistent with the known responses observed *in vivo*. Furthermore, *in vitro* cross-species comparison using human, mouse, and rat hepatoma cell lines identified relatively few species-conserved gene expression

and is corroborates prior reports of species-specific TCDD-induced toxicities. Genome-wide computational identification and characterization of dioxin response elements (DREs) using a position weight matrix identified species-specific regulons in the promoter regions of targeted genes that may account for the observed species-divergent and -specific responses. In order to better understand the molecular mechanisms responsible for regulating the transcriptional responses and downstream hepatotoxicity, ChIP-chip analysis was performed to globally identify TCDD-induced AhR/DNA interactions in mouse hepatic tissue. Interestingly, integration of the DRE, ChIP-chip and gene expression analyses found that only ~32% of all TCDD-elicited hepatic gene expression responses are mediated by a DRE-dependent mechanism. These direct targets of AhR regulation have biological functions related to xenobiotic and lipid metabolism, which correspond with the physiological responses observed *in vivo*. The remaining transcriptional responses that are mediated through a DRE-independent mechanism illustrate the diverse regulatory role of the AhR. Collectively, these results have expanded our knowledge of the hepatic AhR regulatory network and provide insight into the species-conserved responses elicited by TCDD.

## **ACKNOWLEDGEMENTS**

Graduate school has been a tremendous learning experience of not only science and research, but also of life in general. This memorable experience has defined me scientifically and personally in ways that I never thought were possible. I would never have been able to accomplish this without the guidance, support and love from all those around me during my time at Michigan State University and I wish to thank you all.

First, I give many thanks to my advisor, Dr. Timothy Zacharewski. He has provided me with abundant support both in and outside of the lab. Without his initial faith in hiring me as a co-operative education student while as an undergraduate student, I would never have had the opportunity to be at Michigan State University.

Second, I would like to thank my Graduate Committee members: Drs. Kristina Chan, David DeWitt, Gregg Howe and Jack Watson for their availability, insight and unique perspectives in contributing to the success of my project during my committee meetings.

I would also like to acknowledge all my fellow lab members, both past and present, Darrell Boverhof, Josh Kwekel, Lyle Burgoon, Ania Kopec, Suntae Kim, Michelle Angrish, and many others, including numerous co-operative education students. Thank you all for putting up with me over the years. I've learned so much from each and every single one of you and never would have made it through graduate school without your support. I also would like to thank



members of the LaPres lab, especially Dr. John LaPres and Ajith Vengellur for all their help and insightful conversations about everything.

Finally, I could never have endured this experience without the encouragement, support and love from those closest to me. To my father and mother, my brother, sister-in-law, and even my little nephew Riley, I owe you all a tremendous debt for all that you have given me. And to May, you are the highlight of my graduate experience. You've always been there for me when I needed you most and I hope that I can somehow return the favor.

# TABLE OF CONTENTS

|                      |      |
|----------------------|------|
| LIST OF TABLES ..... | viii |
|----------------------|------|

|                       |   |
|-----------------------|---|
| LIST OF FIGURES ..... | x |
|-----------------------|---|

|                             |      |
|-----------------------------|------|
| LIST OF ABBREVIATIONS ..... | xiii |
|-----------------------------|------|

## CHAPTER 1

|                                                                                                                           |          |
|---------------------------------------------------------------------------------------------------------------------------|----------|
| <b>Review of the Literature: Toxicogenomics and TCDD-Induced Toxicity Mediated by the Aryl Hydrocarbon Receptor .....</b> | <b>1</b> |
| Introduction .....                                                                                                        | 2        |
| Toxicogenomics .....                                                                                                      | 2        |
| TCDD and its Elicited Effects .....                                                                                       | 3        |
| The Aryl Hydrocarbon Receptor .....                                                                                       | 5        |
| Conclusions .....                                                                                                         | 9        |
| References .....                                                                                                          | 10       |

## CHAPTER 2

|                                                      |           |
|------------------------------------------------------|-----------|
| <b>Rationale, Hypothesis and Specific Aims .....</b> | <b>16</b> |
| Rationale .....                                      | 17        |
| Hypothesis .....                                     | 17        |
| Specific Aims .....                                  | 17        |

## CHAPTER 3

|                                                                                                                                                               |           |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------|
| <b><i>In Vivo – In Vitro</i> Toxicogenomic Comparison of TCDD-Elicited Gene Expression in Hepa1c1c7 Mouse Hepatoma Cells and C57BL/6 Hepatic Tissue .....</b> | <b>19</b> |
| Abstract .....                                                                                                                                                | 20        |
| Introduction .....                                                                                                                                            | 21        |
| Materials and Methods .....                                                                                                                                   | 23        |
| Results .....                                                                                                                                                 | 27        |
| Discussion .....                                                                                                                                              | 48        |
| References .....                                                                                                                                              | 53        |

## CHAPTER 4

|                                                                                                                       |           |
|-----------------------------------------------------------------------------------------------------------------------|-----------|
| <b>Comparison of TCDD-Elicited Gene Expression in Human HepG2, Mouse Hepa1c1c7 and Rat H4IIE Hepatoma Cells .....</b> | <b>59</b> |
| Abstract .....                                                                                                        | 60        |
| Introduction .....                                                                                                    | 60        |
| Materials and Methods .....                                                                                           | 62        |
| Results .....                                                                                                         | 63        |
| Discussion .....                                                                                                      | 81        |

|                                                                                                                                                               |            |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------|------------|
| References.....                                                                                                                                               | 84         |
| <b>CHAPTER 5</b>                                                                                                                                              |            |
| <b>Genome-Wide Computational Analysis of Dioxin Response Element Location<br/>and Distribution in the Human, Mouse and Rat Genomes.....</b>                   | <b>88</b>  |
| Abstract.....                                                                                                                                                 | 89         |
| Introduction.....                                                                                                                                             | 90         |
| Methods.....                                                                                                                                                  | 92         |
| Results.....                                                                                                                                                  | 98         |
| Discussion.....                                                                                                                                               | 114        |
| References.....                                                                                                                                               | 119        |
| <b>CHAPTER 6</b>                                                                                                                                              |            |
| <b>Integration of Genome-Wide Computational DRE Search, AhR ChIP-chip and<br/>Gene Expression Analyses of TCDD-Elicited Responses in the Mouse Liver.....</b> | <b>125</b> |
| Abstract.....                                                                                                                                                 | 126        |
| Introduction.....                                                                                                                                             | 127        |
| Materials and Methods.....                                                                                                                                    | 129        |
| Results.....                                                                                                                                                  | 132        |
| Discussion.....                                                                                                                                               | 159        |
| References.....                                                                                                                                               | 163        |
| <b>CHAPTER 7</b>                                                                                                                                              |            |
| <b>Conclusions and Future Research.....</b>                                                                                                                   | <b>171</b> |
| Comparative Gene Expression Analysis.....                                                                                                                     | 172        |
| Global AhR Enrichment Analysis .....                                                                                                                          | 173        |
| AhR Interactions with Other Transcription Factors.....                                                                                                        | 174        |
| References.....                                                                                                                                               | 176        |

## LIST OF TABLES

|                                                                                                                                                                                           |     |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| <b>Table 1.</b><br>Gene names and primer sequences for QRT-PCR.....                                                                                                                       | 28  |
| <b>Table 2.</b><br>Classification of common differentially regulated temporal gene expression responses to TCDD in both <i>in vitro</i> and <i>in vivo</i> models.....                    | 39  |
| <b>Table 3.</b><br>Examples of TCDD-elicited gene expression responses unique to Hepa1c1c7 cells.....                                                                                     | 43  |
| <b>Table 4.</b><br>Examples of TCDD-elicited gene expression responses unique to C57BL/6 hepatic tissue.....                                                                              | 44  |
| <b>Table 5.</b><br>Gene coverage of species-specific cDNA microarray platforms and number of differentially regulated genes.....                                                          | 64  |
| <b>Table 6.</b><br>Gene coverage of species-specific Agilent microarray platforms and number of differentially regulated genes.....                                                       | 70  |
| <b>Table 7.</b><br>List of common genes identified as differentially expressed by TCDD treatment from whole genome Agilent microarray analysis. ....                                      | 73  |
| <b>Table 8.</b><br><i>Bona fide</i> DRE sequences used to construct the revised position weight matrix. ....                                                                              | 93  |
| <b>Table 9.</b><br>Distribution of DRE cores, putative DREs and putative DRE densities across the human, mouse and rat genomes. ....                                                      | 100 |
| <b>Table 10.</b><br>Chromosomal density of putative DREs (per Mbp) within the intergenic and intragenic DNA regions of the human, mouse and rat genomes. ....                             | 103 |
| <b>Table 11.</b><br>Chromosomal density of putative DREs (per Mbp) within the 10kb upstream, 5' and 3' UTRs, and CDS regions of RefSeq sequences in the human, mouse and rat genomes..... | 106 |

|                                                                                                                                              |     |
|----------------------------------------------------------------------------------------------------------------------------------------------|-----|
| <b>Table 12.</b><br>Analysis of DRE core and putative DRE containing RefSeq sequences and genes in the<br>human, mouse and rat genomes. .... | 110 |
| <b>Table 13.</b><br>Distribution and density analysis of TCDD-induced AhR enriched regions in the mouse<br>genome. ....                      | 135 |
| <b>Table 14.</b><br>List of AhR enriched regions identified by ChIP-chip analysis at 2 hrs confirmed by<br>ChIP-PCR. ....                    | 140 |
| <b>Table 15.</b><br>Distribution of DRE cores in AhR enriched regions. ....                                                                  | 144 |
| <b>Table 16.</b><br>Significantly over-represent transcription factor module families in TCDD-induced AhR<br>enriched regions. ....          | 148 |
| <b>Table 17.</b><br>Distribution and AhR enrichment and DRE analyses of differentially expressed genes<br>elicited by TCDD. ....             | 155 |
| <b>Table 18.</b><br>Functional enrichment analysis of differentially regulated genes with AhR enrichment using<br>DAVID. ....                | 157 |

# LIST OF FIGURES

|                                                                                                                                                                                                                                                                    |    |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| <b>Figure 1.</b><br>Aryl hydrocarbon receptor signaling mechanism. ....                                                                                                                                                                                            | 7  |
| <b>Figure 2.</b><br>Number of genes differentially regulated ( $P(t) > 0.9999$ and $ \text{fold change}  > 1.5\text{-fold}$ ) as measured by microarray analysis for the (A) time-course and (B) and dose-response studies in mouse hepatoma Hepa1c1c7 cells. .... | 29 |
| <b>Figure 3.</b><br>Hierarchical clustering of the differentially regulated gene lists for A) temporal and B) dose-response microarray studies in mouse hepatoma Hepa1c1c7 cells. ....                                                                             | 31 |
| <b>Figure 4.</b><br><i>K</i> -means clustering of temporally differentially regulated genes <i>in vitro</i> . ....                                                                                                                                                 | 33 |
| <b>Figure 5.</b><br>Comparison of common significant <i>in vitro</i> and <i>in vivo</i> TCDD-elicited time-dependent gene expression changes. ....                                                                                                                 | 36 |
| <b>Figure 6.</b><br>Comparison of Hepa1c1c7 cell and C57BL/6 hepatic tissue basal gene expression. ....                                                                                                                                                            | 46 |
| <b>Figure 7.</b><br>Quantitative real-time PCR verification of <i>in vitro</i> and <i>in vivo</i> microarray results. ....                                                                                                                                         | 47 |
| <b>Figure 8.</b><br>Number of TCDD-elicited differentially expressed genes in human HepG2, mouse Hepa1c1c7, and rat H4IIE. ....                                                                                                                                    | 65 |
| <b>Figure 9.</b><br>Cross-species comparison of TCDD-elicited temporal gene expression responses using cDNA microarrays. ....                                                                                                                                      | 67 |
| <b>Figure 10.</b><br>QRT-PCR verification of the conserved induction of CYP1A1 across the human HepG2, mouse Hepa1c1c7, and rat H4IIE cell lines. ....                                                                                                             | 68 |
| <b>Figure 11.</b><br>Cross-species comparison of TCDD-elicited gene expression responses at 24 hrs using 44×4K Agilent microarrays. ....                                                                                                                           | 72 |

|                                                                                                                                                                                                                                              |     |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| <b>Figure 12.</b><br>QRTPCR verification of examples of species-specific orthologous gene expression responses identified from whole-genome microarray analysis at 24 hrs in the human HepG2, mouse Hepa1c1c7 and rat H4IIE cell lines. .... | 75  |
| <b>Figure 13.</b><br>Comparative analysis of GSTA5 orthologs. ....                                                                                                                                                                           | 78  |
| <b>Figure 14.</b><br>Comparison of the previously published position weight matrix (PWM) and conservation index (C <sub>i</sub> ) for dioxin response elements (DREs) with the revised PWM. ....                                             | 94  |
| <b>Figure 15.</b><br>Defining the various genomic regions used for DRE location analysis. ....                                                                                                                                               | 96  |
| <b>Figure 16.</b><br>Visualization of DRE sequence locations in the UCSC Genome Browser for human CYP1A1 and CYP1A2 gene regions and adjacent intergenic regions. ....                                                                       | 102 |
| <b>Figure 17.</b><br>Distribution of putative DREs in the regions 10 kb upstream to 5 kb downstream of a TSS for all RefSeq sequences.....                                                                                                   | 108 |
| <b>Figure 18.</b><br>Frequency of putative DREs within known human, mouse and rat genes. ....                                                                                                                                                | 112 |
| <b>Figure 19.</b><br>Summary of AhR enrichment within <i>Cyp1a1</i> genic region at 2 and 24 hrs. ....                                                                                                                                       | 133 |
| <b>Figure 20.</b><br>Characterization of TCDD-induced AhR enriched regions at 2 and 24 hrs (FDR < 0.01).....                                                                                                                                 | 137 |
| <b>Figure 21.</b><br>TCDD-induced AhR enrichment (FDR < 0.01) densities in the proximal promoter (10 kb upstream and 5 kb downstream of a TSS) at 2 hrs (A) and 24 hrs (B). ....                                                             | 138 |
| <b>Figure 22.</b><br>Confirmation of hepatic TCDD-induced AhR enrichment identified by ChIP-chip analysis (FDR < 0.01) at 2 hrs by ChIP-PCR. ....                                                                                            | 141 |
| <b>Figure 23.</b><br>Mapping TCDD-induced AhR enriched regions (FDR < 0.01) with DRE locations. ....                                                                                                                                         | 145 |

|                                                                                                                 |     |
|-----------------------------------------------------------------------------------------------------------------|-----|
| <b>Figure 24.</b>                                                                                               |     |
| <i>De novo</i> motif analysis of intragenic (A) and intergenic (B) AhR enriched regions lacking a DRE core..... | 149 |

|                                                                                        |     |
|----------------------------------------------------------------------------------------|-----|
| <b>Figure 25.</b>                                                                      |     |
| Mapping TCDD-induced AhR enriched regions (FDR < 0.01) and DRE analysis to genes. .... | 151 |

|                                                                                                                                                                                                                                                                   |     |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| <b>Figure 26.</b>                                                                                                                                                                                                                                                 |     |
| Circos plots integrating DRE analysis, AhR enrichment (2 hrs; FDR < 0.01) and heatmaps for hepatic differential gene expression responses ( $ \text{fold change}  \geq 1.5$ and $P1(t) > 0.999$ ) induced by TCDD across the genome (A) and chromosome 9 (B)..... | 153 |



## LIST OF ABBREVIATIONS

|                  |                                                                             |
|------------------|-----------------------------------------------------------------------------|
| 3MC              | 3-methylcholanthrene                                                        |
| AHH              | aryl hydrocarbon hydroxylase                                                |
| AhR              | aryl hydrocarbon receptor                                                   |
| ANOVA            | analysis of variance                                                        |
| ARNT             | aryl hydrocarbon nuclear translocator                                       |
| bHLH             | basic-helix-loop-helix                                                      |
| CDS              | coding sequence                                                             |
| ChIP             | chromatin immunoprecipitation                                               |
| ChIP-chip        | ChIP coupled with genome tiling microarrays                                 |
| ChIP-PCR         | ChIP coupled with PCR                                                       |
| ChIP-seq         | ChIP coupled with next-generation sequencing                                |
| CHX              | cycloheximide                                                               |
| C <sub>i</sub>   | conservation index                                                          |
| Cyp1a1           | cytochrome P450, family 1, subfamily a, polypeptide 1                       |
| DLC              | dioxn-like compounds                                                        |
| DMSO             | dimethyl sulfoxide                                                          |
| DRE              | dioxin response element                                                     |
| dUTP             | deoxyuridine triphosphate                                                   |
| EC <sub>50</sub> | effective concentration causing 50% of the maximal response                 |
| ED <sub>50</sub> | effective dose causing 50% of the maximal response                          |
| FDR              | false discovery rate                                                        |
| GO               | Gene Ontology                                                               |
| HAH              | halogenated aromatic hydrocarbon                                            |
| HSP              | heat shock protein                                                          |
| IARC             | International Agency for Research on Cancer                                 |
| ICCVAM           | Interagency Coordination Committee on the Validation of Alternative Methods |
| IgG              | Immunoglobulin G                                                            |
| LD <sub>50</sub> | lethal dose for 50% of the population                                       |
| MA               | moving average                                                              |
| MS               | matrix similarity                                                           |

|        |                                              |
|--------|----------------------------------------------|
| PAS    | PER-ARNT-SIM                                 |
| PCB    | polychlorinated biphenyl                     |
| PCDD   | polychlorinated-dibenzo- <i>p</i> -dioxin    |
| PCDF   | polychlorinated-dibenzofuran                 |
| PCR    | polymerase chain reaction                    |
| POP    | persistent organic pollutant                 |
| PWM    | position weight matrix                       |
| QRTPCR | quantitative real-time PCR                   |
| TCDD   | 2,3,7,8-tetrachlorodibenzo- <i>p</i> -dioxin |
| TF     | transcription factor                         |
| TFBS   | transcription factor binding site            |
| TSS    | transcription start site                     |
| UCSC   | University of California Santa Cruz          |
| UTR    | untranslated region                          |

# CHAPTER 1

# **CHAPTER 1**

## **REVIEW OF THE LITERATURE: TOXICOGENOMICS AND TCDD- INDUCED TOXICITY MEDIATED BY THE ARYL HYDROCARBON RECEPTOR**

### **INTRODUCTION**

Dioxins and other related environmental persistent organic pollutants (POPs) continue to be public concerns due to their potentially adverse effects in ecological wildlife and humans [1-3]. These compounds trigger a signal transduction pathway that lead to various physiological responses, including homeostatic perturbations and cellular responses such as proliferation, differentiation, apoptosis and necrosis. The activation of the aryl hydrocarbon receptor (AhR) signaling pathway by dioxin is responsible for inducing metabolizing enzymes that are involved with detoxifying and/or biotransforming various xenobiotics. Although the AhR signaling pathway is well understood, the full spectrum of AhR-mediated responses remain largely unknown. Advancements in AhR research through the incorporation of toxicogenomics hopes to further expand the current understanding of the AhR regulatory network.

### **TOXICOGENOMICS**

Technological advances in microarray technology have revolutionized the field of toxicology and have contributed to the emergence of toxicogenomics [4]. Microarrays can simultaneously profile the gene expression responses across entire genomes to provide comprehensive insight into the mechanisms of toxicity for drugs, natural products, commerce

chemicals and environmental pollutants as well as their mixtures, which supports drug development and quantitative risk assessment [5-9]. Using an integrative systems biology approach by combining common endpoints of traditional toxicology, such as changes in body and organ weights, and changes in blood chemistry, with global gene expression signatures, transcriptional responses can be phenotypically anchored to those physiological responses. Furthermore, comparison of elicited gene expression profiles with databases containing signatures of known toxicants can aid in identifying biomarkers of exposure and toxicity that can be used in high-throughput screening programs.

There are many contributing factors in addition to elicited gene expression responses that can influence a toxic outcome, including DNA-protein interactions, DNA methylation and post-translational modifications. To this end, toxicogenomics will continue to evolve and incorporate additional high-throughput bioassays, such as DNA methylation and chromatin immunoprecipitation microarrays (ChIP-chip) and next-generation sequencing (ChIP-seq), which will provide further mechanistic insight into toxicity. Predictive biomarkers will integrate all disparate and complementary responses, and allow for further stratification of the population to identify sensitive groups, which could then be treated more effectively while minimizing the risk of unacceptable toxicities. These biomarkers will be mechanistically based and anchored to the adverse effect, which is expected to further minimize uncertainties in the source-to-outcome continuum and extrapolations between across models (*in vitro* to *in vivo*) and species (rodent to human).

## **TCDD AND ITS ELICITED EFFECTS**

2,3,7,8-Tetrachlorodibenzo-*p*-dioxin (TCDD) and other related halogenated aromatic hydrocarbons (HAHs) are widespread, persistent and bioaccumulative environmental

contaminants [1-3]. Polychlorinated biphenyls (PCBs), such as 3,3',4,4',5-pentachlorobiphenyl (PCB126), were widely used as heat transfer fluids, organic diluents and plasticizers. Many dioxin-like compounds (DLC), including polychlorinated-dibenzo-*p*-dioxins (PCDDs), dibenzofurans (PCDFs) and alkylated PCDFs, are by-products inadvertently created during common industrial processes, including the production of organochlorine pesticides, bleaching of wood pulp and waste incineration processes [10]. TCDD is considered to be the most toxic HAH and has been used as a model compound to study their mechanism of action [10].

The primary route of exposure to TCDD is through the diet, but other sources of constant exposure include the air and soil [11, 12]. Environmental levels of dioxins in the U.S. have continually declined in recent decades due largely in part to government imposed emission regulations, advancements in pollution control technologies specific to controlling dioxin discharges, and voluntary actions of industries to reduce and/or prevent dioxin release. However, dioxins released from “*backyard burning*” have risen dramatically in recent years and is now the primary source of environmental dioxins in the U.S [13]. Although the environmental levels of dioxin are steadily decreasing, concern still remains due to its chemical structure and lipophilic nature that make TCDD very resistant to metabolism. It is estimated that the half-life of TCDD in humans is between 7 and 10 years [14], which may contribute to the sustained activation of the AhR and downstream toxicities.

TCDD and other DLCs elicit a multitude of toxic and biochemical responses including immunotoxicity, dermal toxicity, lethality, wasting syndrome, tumor promotion, hepatotoxicity, teratogenicity, modulation of diverse enzyme activities and alteration of endocrine response pathways [2, 15]. Many of these biological and toxic effects of TCDD occur in a tissue-, sex-, age- or species-specific manner [16-18]. Even within closely related rodent models, there are

wide differences in their response to TCDD exposure [19]. For example, LD<sub>50</sub> (lethal dose for 50% of the population) values range from 1 µg/kg in the guinea pig, the most sensitive species, to > 1000 µg/kg in hamster, the most resistant [20].

Although the effects of TCDD in rodent models are well documented, significantly less is known regarding the effects in humans. Our current understanding of the human effects are based on limited epidemiological observations from populations accidentally exposed to TCDD, including Vietnam residents and war veterans exposed to Agent Orange, and neighboring residents of a chemical plant in Sveso, Italy. The acute effects of TCDD include the onset of chloracne, transient liver toxicity, fatigue, general weakness, and weight loss [21, 22]. TCDD's persistent nature and resistance to metabolism in the body allows for long-term effects that can remain years following a massive exposure. Symptoms include an increased risk of atherosclerosis [23, 24], hypertension and ischemic heart disease [22, 25], neurological abnormalities [21], diabetes [26, 27], hormonal perturbations [28, 29], and increased incidences of cancer [22, 30]. Although there is limited evidence supporting the carcinogenic effects of TCDD in humans, it remains classified as a known human carcinogen by the International Agency for Research on Cancer (IARC) [10] based on sufficient evidence in animal models and extensive mechanistic data from studies involving humans and animals [30].

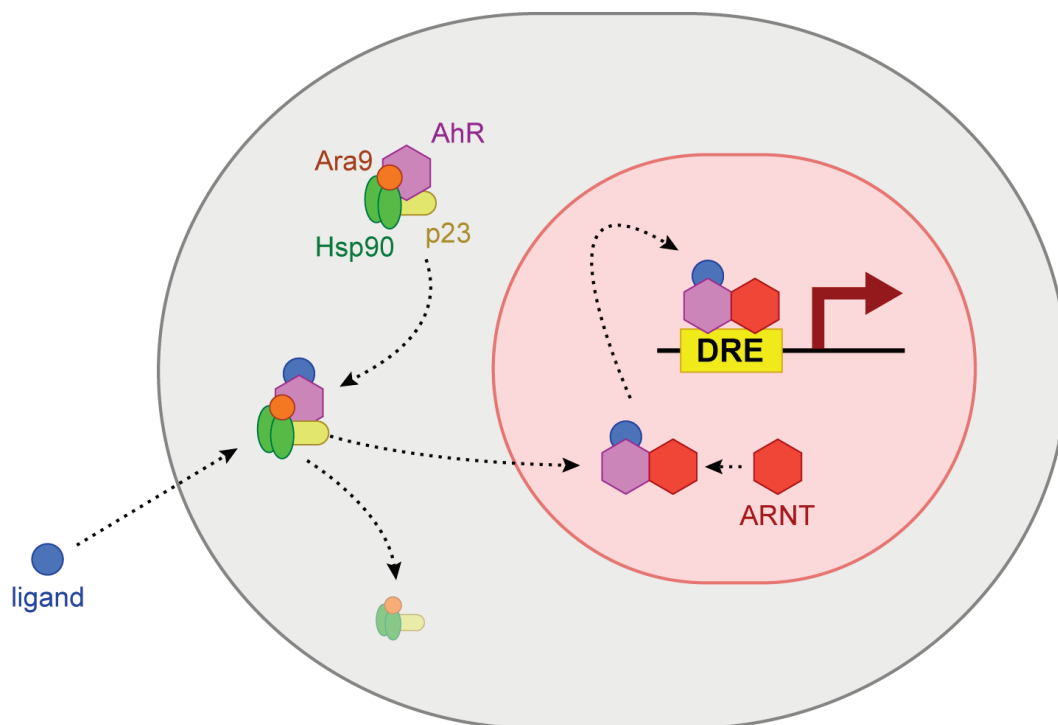
## **THE ARYL HYDROCARBON RECEPTOR**

Early research into the potential mechanisms of TCDD-elicited toxicity revealed that both TCDD and 3-methylcholanthrene (3MC) induced aryl hydrocarbon hydroxylase (AHH) activity but with different potencies [31]. Furthermore, examination of a series of halogenated aromatic compounds revealed a strong correlation between their structure-AHH induction and structure-toxicity relationships. Based on these data, Poland and co-workers hypothesized that a

ligand-binding protein or receptor was the initial cellular target of TCDD and, subsequently, using radiolabeled-TCDD identified the AhR in the hepatic cytosol from C57BL/6 mice [32]. Additional research demonstrated that the AhR is present in multiple tissues and species, and that it shares many characteristics with members of the nuclear hormone receptor superfamily as a ligand-activated transcription factor [33].

The AhR is classified as a member of the basic-helix-loop-helix PER-ARNT-SIM (bHLH-PAS) family of transcription factors [2, 16, 34, 35]. Members of this family have important roles as sensors for different environmental stimuli, such as hypoxia and exogenous chemical insult [36]. TCDD elicits a broad spectrum of gene responses, but the best characterized responses are those belonging to the “AhR gene battery”, which include phase I and II xenobiotic metabolizing enzymes, *CYP1A1*, *CYP1A2*, NAD(P)H: quinone oxidoreductase (*NQO1*), aldehyde dehydrogenase 3 (*ALDH3A1*), and UDP glucuronosyltransferase (*UGT1A6*) [37]. Unliganded AhR exists within the cytosol as a complex with other proteins that stabilize the receptor and maintain its proper cellular localization (Figure 1). The complex consists of a Hsp90 dimer [38], the co-chaperone protein p23 [39] and the immunophilin-like AIP/ARA9/XAP2 protein [40-42]. Ligand binding to the AhR causes a conformational change that results in the dissociation of the chaperone proteins and translocation of the activated receptor into the nucleus. Within the nucleus, the activated AhR heterodimerizes with the aryl hydrocarbon receptor translocator (ARNT), another member of the bHLH-PAS family of transcription factors. The activated heterodimer is then able to bind specific regulatory elements, called dioxin response elements (DREs), within the promoter region of target genes to regulate transcriptional events, which ultimately result in the observed toxic and biochemical responses [43].





**Figure 1. Aryl hydrocarbon receptor signaling mechanism.**

For interpretation of the references to color in this and all other figures, the reader is referred to the electronic version of this dissertation. In the absence of ligand, the aryl hydrocarbon receptor (AhR) is sequestered in the cytoplasm bound to heat shock protein 90 (Hsp90), ARA9 and p23. Ligand binding results in a conformational change in the receptor, dissociation of chaperone proteins and translocation to the nucleus where it forms a heterodimer with the aryl hydrocarbon receptor nuclear translocator (ARNT), another member of the bHLH-PAS family. This heterodimer binds specific DNA elements, termed dioxin response elements (DREs), leading to changes in gene expression.

*In vivo* studies have demonstrated the necessary requirement for the AhR signaling pathway in mediating the observed TCDD-induced toxicity responses. Mice carrying low binding-affinity AhR alleles are less susceptible than other mice to the toxic effects classically induced by TCDD [44]. Additionally, AhR-null mice exhibit resistance to prototypical toxicities of TCDD and other related compounds [45]. Studies with mice carrying mutations within the nuclear localization/DRE binding domains as well as with mice harboring a hypomorphic ARNT allele fail to display TCDD-induced toxicity responses [46, 47]. In addition to its role in mediating toxicity responses, the AhR/ARNT signaling pathway has been implicated in having critical roles in normal development, differentiation and growth, as evidenced by abnormalities in the liver, heart, thymus and immune system of AhR-null mice. Moreover, mice expressing a constitutively active AhR display increased hepatocarcinogenesis, which has further implicated AhR activation in tumor promotion [48].

The AhR binds DNA at DREs containing the substitution intolerant 5'-GCGTG-3' core sequence to regulate transcription [49-53]. Ultra-violet cross-linking [54] and site selection experiments [43] indicate that the AhR occupies the 5'-TNGC half-site, while ARNT contacts the GTG-3' half-site. Furthermore, strong evidence indicates that the 5'- and 3'- flanking nucleotides play important roles in modulating the DNA-binding affinity and enhancer functionality of the AhR/ARNT heterodimer [52, 55-57]. Although AhR binding at *bona fide* functional DREs has been demonstrated for genes including those of the AhR gene battery, genome location analysis of AhR-DNA interactions using chromatin immunoprecipitation tiling arrays have found that approximately 50% of AhR binding sites across the genome occur independently of a DRE core [58-60]. Furthermore, other studies have reported an alternate

functional response element (DRE-II) containing the 5'-CATGN<sub>6</sub>C[T|A]TG-3' sequence that is capable of recruiting the activated AhR/ARNT heterodimer complex [61, 62].

## **CONCLUSIONS**

Although the mechanism of AhR-mediated changes in gene expression has been well characterized, the full spectrum of targeted genes has not been identified, which contributes to the poor understanding of TCDD-induced toxicity. Furthermore, recent data suggest that there are potentially alternate mechanisms of AhR signaling that are independent of DRE binding. Collectively, these data indicate the complexity behind AhR-mediated transcriptional regulation and requires further research in order to expand our current understanding of the AhR regulatory network. Additional characterization of the mechanism of AhR signaling will provide an enhanced foundation for toxicogenomic-based biomarker discovery of TCDD-induced toxicity.

## REFERENCES

## REFERENCES

1. Parzefall W: **Risk assessment of dioxin contamination in human food.** *Food Chem Toxicol* 2002, **40**:1185-1189.
2. Denison MS, Heath-Pagliuso S: **The Ah receptor: a regulator of the biochemical and toxicological actions of structurally diverse chemicals.** *Bulletin of environmental contamination and toxicology* 1998, **61**:557-568.
3. Mandal P: **Dioxin: a review of its environmental effects and its aryl hydrocarbon receptor biology.** *J Comp Physiol [B]* 2005, **175**:221-230.
4. Nuwaysir E, Bittner M, Trent J, Barrett J, Afshari C: **Microarrays and toxicology: the advent of toxicogenomics.** *Mol Carcinog* 1999, **24**:153-159.
5. Leighton J: **Application of emerging technologies in toxicology and safety assessment: regulatory perspectives.** *Int J Toxicol* 2005, **24**:153-155.
6. MacGregor JT: **SNPs and chips: genomic data in safety evaluation and risk assessment.** *Toxicol Sci* 2003, **73**:207-208.
7. Olden K: **Genomics in environmental health research--opportunities and challenges.** *Toxicology* 2004, **198**:19-24.
8. Tennant R: **The National Center for Toxicogenomics: using new technologies to inform mechanistic toxicology.** *Environ Health Perspect* 2002, **110**:A8-10.
9. Waring J, Ulrich R: **The impact of genomics-based technologies on drug safety evaluation.** *Annu Rev Pharmacol Toxicol* 2000, **40**:335-352.
10. **IARC Working Group on the Evaluation of Carcinogenic Risks to Humans: Polychlorinated Dibenzo-Para-Dioxins and Polychlorinated Dibenzofurans.** Lyon, France, 4-11 February 1997. *IARC Monogr Eval Carcinog Risks Hum* 1997, **69**:1-631.
11. Liem AK, Fürst P, Rappe C: **Exposure of populations to dioxins and related compounds.** *Food Addit Contam* 2000, **17**:241-259.
12. Larsen JC: **Risk assessments of polychlorinated dibenzo- p-dioxins, polychlorinated dibenzofurans, and dioxin-like polychlorinated biphenyls in food.** *Mol Nutr Food Res* 2006, **50**:885-896.
13. U.S.EPA: **An Inventory of Sources and Environmental Releases of Dioxin-Like Compounds in the United States for the Years 1987, 1995, and 2000.** Washington, DC; 2006.

14. Neuberger M, Landvoigt W, Derntl F: **Blood levels of 2,3,7,8-tetrachlorodibenzo-p-dioxin in chemical workers after chloracne and in comparison groups.** *Int Arch Occup Environ Health* 1991, **63**:325-327.
15. Krishnan V, Porter W, Santostefano M, Wang X, Safe S: **Molecular mechanism of inhibition of estrogen-induced cathepsin D gene expression by 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD) in MCF-7 cells.** *Mol Cell Biol* 1995, **15**:6710-6719.
16. Poland A, Knutson JC: **2,3,7,8-tetrachlorodibenzo-p-dioxin and related halogenated aromatic hydrocarbons: examination of the mechanism of toxicity.** *Annu Rev Pharmacol Toxicol* 1982, **22**:517-554.
17. Boutros PC, Yan R, Moffat ID, Pohjanvirta R, Okey AB: **Transcriptomic responses to 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD) in liver: comparison of rat and mouse.** *BMC Genomics* 2008, **9**:419.
18. Boverhof DR, Burgoon LD, Tashiro C, Sharratt B, Chittim B, Harkema JR, Mendrick DL, Zacharewski TR: **Comparative toxicogenomic analysis of the hepatotoxic effects of TCDD in Sprague Dawley rats and C57BL/6 mice.** *Toxicol Sci* 2006, **94**:398-416.
19. Harris M, Moore J, Vos J, Gupta B: **General biological effects of TCDD in laboratory animals.** *Environ Health Perspect* 1973, **5**:101-109.
20. Schwetz BA, Norris JM, Sparschu GL, Rowe UK, Gehring PJ, Emerson JL, Gerbig CG: **Toxicology of chlorinated dibenzo-p-dioxins.** *Environ Health Perspect* 1973, **5**:87-99.
21. Pelclová D, Urban P, Preiss J, Lukás E, Fenclová Z, Navrátil T, Dubská Z, Senholdová Z: **Adverse health effects in humans exposed to 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD).** *Rev Environ Health* 2006, **21**:119-138.
22. Sweeney MH, Mocarelli P: **Human health effects after exposure to 2,3,7,8-TCDD.** *Food Addit Contam* 2000, **17**:303-316.
23. Dalton TP, Kerzee JK, Wang B, Miller M, Dieter MZ, Lorenz JN, Shertzer HG, Nerbert DW, Puga A: **Dioxin exposure is an environmental risk factor for ischemic heart disease.** *Cardiovasc Toxicol* 2001, **1**:285-298.
24. Martin JV: **Lipid abnormalities in workers exposed to dioxin.** *Br J Ind Med* 1984, **41**:254-256.
25. Vena J, Boffetta P, Becher H, Benn T, Bueno-de-Mesquita HB, Coggon D, Colin D, Flesch-Janys D, Green L, Kauppinen T, et al: **Exposure to dioxin and nonneoplastic mortality in the expanded IARC international cohort study of phenoxy herbicide and chlorophenol production workers and sprayers.** *Environ Health Perspect* 1998, **106 Suppl 2**:645-653.

26. Henriksen GL, Ketchum NS, Michalek JE, Swaby JA: **Serum dioxin and diabetes mellitus in veterans of Operation Ranch Hand.** *Epidemiology* 1997, **8**:252-258.
27. Remillard RBJ, Bunce NJ: **Linking dioxins to diabetes: epidemiology and biologic plausibility.** *Environ Health Perspect* 2002, **110**:853-858.
28. Birnbaum LS, Cummings AM: **Dioxins and endometriosis: a plausible hypothesis.** *Environ Health Perspect* 2002, **110**:15-21.
29. Johnson E, Shorter C, Bestervelt L, Patterson D, Needham L, Piper W, Lucier G, Nolan C: **Serum hormone levels in humans with low serum concentrations of 2,3,7,8-TCDD.** *Toxicol Ind Health* 2001, **17**:105-112.
30. Steenland K, Bertazzi P, Baccarelli A, Kogevinas M: **Dioxin revisited: developments since the 1997 IARC classification of dioxin as a human carcinogen.** *Environ Health Perspect* 2004, **112**:1265-1268.
31. Poland A, Glover E: **Comparison of 2,3,7,8-tetrachlorodibenzo-p-dioxin, a potent inducer of aryl hydrocarbon hydroxylase, with 3-methylcholanthrene.** *Molecular Pharmacology* 1974, **10**:349-359.
32. Poland A, Glover E, Kende AS: **Stereospecific, high affinity binding of 2,3,7,8-tetrachlorodibenzo-p-dioxin by hepatic cytosol. Evidence that the binding species is receptor for induction of aryl hydrocarbon hydroxylase.** *J Biol Chem* 1976, **251**:4936-4946.
33. Evans RM: **The steroid and thyroid hormone receptor superfamily.** *Science* 1988, **240**:889-895.
34. Burbach K, Poland A, Bradfield C: **Cloning of the Ah-receptor cDNA reveals a distinctive ligand-activated transcription factor.** *Proc Natl Acad Sci U S A* 1992, **89**:8185-8189.
35. Ema M, Sogawa K, Watanabe N, Chujoh Y, Matsushita N, Gotoh O, Funae Y, Fujii-Kuriyama Y: **cDNA cloning and structure of mouse putative Ah receptor.** *Biochem Biophys Res Commun* 1992, **184**:246-253.
36. Gu Y, Hogenesch J, Bradfield C: **The PAS superfamily: sensors of environmental and developmental signals.** *Annu Rev Pharmacol Toxicol* 2000, **40**:519-561.
37. Nebert D, Petersen D, Fornace A: **Cellular responses to oxidative stress: the [Ah] gene battery as a paradigm.** *Environ Health Perspect* 1990, **88**:13-25.
38. Perdew G: **Association of the Ah receptor with the 90-kDa heat shock protein.** *J Biol Chem* 1988, **263**:13802-13805.

39. Kazlauskas A, Poellinger L, Pongratz I: **Evidence that the co-chaperone p23 regulates ligand responsiveness of the dioxin (Aryl hydrocarbon) receptor.** *J Biol Chem* 1999, **274**:13519-13524.
40. Ma Q, Whitlock J: **A novel cytoplasmic protein that interacts with the Ah receptor, contains tetratricopeptide repeat motifs, and augments the transcriptional response to 2,3,7,8-tetrachlorodibenzo-p-dioxin.** *J Biol Chem* 1997, **272**:8878-8884.
41. Carver L, Bradfield C: **Ligand-dependent interaction of the aryl hydrocarbon receptor with a novel immunophilin homolog in vivo.** *J Biol Chem* 1997, **272**:11452-11456.
42. Meyer B, Pray-Grant M, Vanden Heuvel J, Perdew G: **Hepatitis B virus X-associated protein 2 is a subunit of the unliganded aryl hydrocarbon receptor core complex and exhibits transcriptional enhancer activity.** *Mol Cell Biol* 1998, **18**:978-988.
43. Hankinson O: **The aryl hydrocarbon receptor complex.** *Annu Rev Pharmacol Toxicol* 1995, **35**:307-340.
44. Okey AB, Vella LM, Harper PA: **Detection and characterization of a low affinity form of cytosolic Ah receptor in livers of mice nonresponsive to induction of cytochrome P1-450 by 3-methylcholanthrene.** *Molecular Pharmacology* 1989, **35**:823-830.
45. Gonzalez F, Fernandez-Salguero P: **The aryl hydrocarbon receptor: studies using the AHR-null mice.** *Drug Metab Dispos* 1998, **26**:1194-1198.
46. Walisser J, Bunger M, Glover E, Harstad E, Bradfield C: **Patent ductus venosus and dioxin resistance in mice harboring a hypomorphic Arnt allele.** *J Biol Chem* 2004, **279**:16326-16331.
47. Bunger M, Moran S, Glover E, Thomae T, Lahvis G, Lin B, Bradfield C: **Resistance to 2,3,7,8-tetrachlorodibenzo-p-dioxin toxicity and abnormal liver development in mice carrying a mutation in the nuclear localization sequence of the aryl hydrocarbon receptor.** *J Biol Chem* 2003, **278**:17767-17774.
48. Moennikes O, Loeppen S, Buchmann A, Andersson P, Ittrich C, Poellinger L, Schwarz M: **A constitutively active dioxin/aryl hydrocarbon receptor promotes hepatocarcinogenesis in mice.** *Cancer Res* 2004, **64**:4707-4710.
49. Denison MS, Fisher JM, Whitlock JP: **The DNA recognition site for the dioxin-Ah receptor complex. Nucleotide sequence and functional analysis.** *The Journal of biological chemistry* 1988, **263**:17221-17224.
50. Fujisawa-Sehara A, Sogawa K, Yamane M, Fujii-Kuriyama Y: **Characterization of xenobiotic responsive elements upstream from the drug-metabolizing cytochrome P-450c gene: a similarity to glucocorticoid regulatory elements.** *Nucleic Acids Res* 1987, **15**:4179-4191.



51. Neuhold LA, Shirayoshi Y, Ozato K, Jones JE, Nebert DW: **Regulation of mouse CYP1A1 gene expression by dioxin: requirement of two cis-acting elements during induction.** *Molecular and Cellular Biology* 1989, **9**:2378-2386.
52. Shen ES, Whitlock JP: **Protein-DNA interactions at a dioxin-responsive enhancer. Mutational analysis of the DNA-binding site for the liganded Ah receptor.** *J Biol Chem* 1992, **267**:6815-6819.
53. Yao EF, Denison MS: **DNA sequence determinants for binding of transformed Ah receptor to a dioxin-responsive enhancer.** *Biochemistry* 1992, **31**:5060-5067.
54. Bacsí SG, Reisz-Porszasz S, Hankinson O: **Orientation of the heterodimeric aryl hydrocarbon (dioxin) receptor complex on its asymmetric DNA recognition sequence.** *Molecular Pharmacology* 1995, **47**:432-438.
55. Gillesby BE, Stanostefano M, Porter W, Safe S, Wu ZF, Zacharewski TR: **Identification of a motif within the 5' regulatory region of pS2 which is responsible for AP-1 binding and TCDD-mediated suppression.** *Biochemistry* 1997, **36**:6080-6089.
56. Lusska A, Shen E, Whitlock JP: **Protein-DNA interactions at a dioxin-responsive enhancer. Analysis of six bona fide DNA-binding sites for the liganded Ah receptor.** *J Biol Chem* 1993, **268**:6575-6580.
57. Sun YV, Boverhof DR, Burgoon LD, Fielden MR, Zacharewski TR: **Comparative analysis of dioxin response elements in human, mouse and rat genomic sequences.** *Nucleic Acids Res* 2004, **32**:4512-4523.
58. Kinehara M, Fukuda I, Yoshida K-I, Ashida H: **High-throughput evaluation of aryl hydrocarbon receptor-binding sites selected via chromatin immunoprecipitation-based screening in Hepa-1c1c7 cells stimulated with 2,3,7,8-tetrachlorodibenzo-p-dioxin.** *Genes Genet Syst* 2008, **83**:455-468.
59. Ahmed S, Valen E, Sandelin A, Matthews J: **Dioxin increases the interaction between aryl hydrocarbon receptor and estrogen receptor alpha at human promoters.** *Toxicol Sci* 2009, **111**:254-266.
60. Pansoy A, Ahmed S, Valen E, Sandelin A, Matthews J: **3-methylcholanthrene induces differential recruitment of aryl hydrocarbon receptor to human promoters.** *Toxicological sciences : an official journal of the Society of Toxicology* 2010.
61. Sogawa K, Numayama-Tsuruta K, Takahashi T, Matsushita N, Miura C, Nikawa J-i, Gotoh O, Kikuchi Y, Fujii-Kuriyama Y: **A novel induction mechanism of the rat CYP1A2 gene mediated by Ah receptor-Arnt heterodimer.** *Biochem Biophys Res Commun* 2004, **318**:746-755.
62. Boutros PC, Moffat ID, Franc MA, Tijet N, Tuomisto J, Pohjanvirta R, Okey AB: **Dioxin-responsive AHRE-II gene battery: identification by phylogenetic footprinting.** *Biochem Biophys Res Commun* 2004, **321**:707-715.

## CHAPTER 2

## **CHAPTER 2**

### **RATIONALE, HYPOTHESIS AND SPECIFIC AIMS**

#### **RATIONALE**

TCDD is a ubiquitous environmental contaminant that elicits a broad spectrum of toxic and biochemical responses in a tissue-, sex-, age-, and species-specific manner, and include wasting syndrome, tumor promotion, teratogenesis, hepatotoxicity and modulation of gene expression. Most, if not all of these effects are due to inappropriate changes in gene expression mediated by the AhR, a ligand-activated transcription factor. Despite decades of continuous research, the mechanism responsible for the full spectrum of elicited toxic effects remains largely unknown. The objective of this study is to further characterize the AhR regulatory network by using a comprehensive toxicogenomic approach that incorporates genome-wide identification of dioxin response elements (DREs) and analyses of TCDD-elicited gene expression responses and AhR interactions with the genome to enhance our knowledge of AhR-mediated transcriptional regulation.

#### **HYPOTHESIS**

Toxicogenomic approaches can identify a set of genes to be used as predictive biomarkers of AhR-mediated hepatotoxicity.

#### **SPECIFIC AIMS**

In order to test the hypothesis, a comprehensive understanding of the AhR's transcriptional regulatory role elicited by TCDD is required. Therefore, the approaches that will

be used to test this hypothesis involve the use of gene expression and chromatin immunoprecipitation microarrays, and *in silico* DRE analysis that will:

1. Assess the predictive capabilities of the Hepa1c1c7 *in vitro* system in modeling *in vivo* mouse hepatotoxicity responses elicited by TCDD.
2. Characterize the conserved gene expression responses elicited by TCDD in hepatoma cell lines across three separate species; mouse Hepa1c1c7, rat H4IIE, and human HepG2 cell lines.
3. Computationally locate and characterize all DREs in the human, mouse, and rat genomes using a position weight matrix.
4. Characterize the *in vivo* interaction of the AhR with the genome elicited by TCDD in the mouse liver.

## CHAPTER 3

Dere E, Boverhof DR, Burgoon LD, Zacharewski TR: **In Vivo-In Vitro Toxicogenomic Comparison of TCDD-Elicited Gene Expression in Hepa1c1c7 Mouse Hepatoma Cells and C57BL/6 Hepatic Tissue.** *BMC Genomics* 2006, 7:80

## CHAPTER 3

# IN VIVO – IN VITRO TOXICOGENOMIC COMPARISON OF TCDD- ELICITED GENE EXPRESSION IN HEP1C1C7 MOUSE HEPATOMA CELLS AND C57BL/6 HEPATIC TISSUE

### ABSTRACT

*In vitro* systems have inherent limitations in their capacity to model whole-organism gene responses, which must be identified and appropriately considered when developing predictive biomarkers of *in vivo* toxicity. Systematic comparison of *in vitro* and *in vivo* temporal gene expression profiles was conducted to assess the ability of Hep1c1c7 mouse hepatoma cells to model hepatic responses in C57BL/6 mice following treatment with 2,3,7,8-tetrachlorodibenzo-*p*-dioxin (TCDD). Gene expression analysis and functional gene annotation indicate that Hep1c1c7 cells appropriately modeled the induction of xenobiotic metabolism genes *in vivo*. However, responses associated with cell cycle progression and proliferation were unique to Hep1c1c7 cells, consistent with the cell cycle arrest effects of TCDD on rapidly dividing cells. In contrast, lipid metabolism and immune responses, representative of whole-organism effects *in vivo*, were not replicated in Hep1c1c7 cells. These results identified inherent differences in TCDD-mediated gene expression responses between these models and highlighted the limitations of *in vitro* systems in modeling whole-organism responses, and additionally identified potential predictive biomarkers of toxicity.

## INTRODUCTION

Advances in microarray and related technologies continue to revolutionize biomedical research and are being incorporated into toxicology and risk assessment. These technologies not only facilitate a more comprehensive elucidation of the mechanisms of toxicity, but also support mechanistically-based quantitative risk assessment [1-5]. In addition, these technologies are being used to develop predictive toxicity screening assays to screen drug candidates with adverse characteristics earlier in the development pipeline in order to prioritize resources and maximize successes in clinical trials [6-8]. Comparable screening strategies are also being proposed to rank and prioritize commercial chemicals, natural products, and environmental contaminants that warrant further toxicological investigation. Traditionally, rodent models or surrogates for ecologically-relevant species are typically used in regulatory testing. However, public and regulatory pressure, especially in Europe, seek to minimize the use of animals in testing [9]. Similar policies in the US, such as the ICCVAM Authorization Act of 2000, provide guidelines to facilitate the regulatory acceptance of alternative testing methods. These initiatives combined with the need to assess an expanding list of drug candidates and commercial chemicals for toxicity, have increased demand for the development and implementation of high-throughput *in vitro* screening assays that are predictive of toxicity in humans and ecologically-relevant species.

Various *in vitro* hepatic models including the isolated perfused liver, precision cut liver slices, isolated primary liver cells, and a number of immortalized liver cell lines, have been used as animal alternatives [10]. In addition to providing a renewable model, *in vitro* systems are a cost-effective alternative and are amenable to high-throughput screening. These models, particularly immortalized cell lines, also allow for more in-depth biochemical and molecular investigations, such as over-expression, knock-down, activation or inhibition strategies, thus

further elucidating mechanisms of action. However, inherent limitations in the capacity of cell cultures to model whole-organism responses must also be considered when identifying putative biomarkers for high-throughput toxicity screening assays, and elucidating relevant mechanisms of toxicity that support quantitative risk assessment. Despite several *in vitro* toxicogenomic reports [11-13], few have systematically examined the capacity of *in vitro* systems to predict *in vivo* gene expression profiles in response to chemical treatment [10, 14].

2,3,7,8-Tetrachlorodibenzo-*p*-dioxin (TCDD) is a widespread environmental contaminant that elicits a number of adverse effects including tumor promotion, teratogenesis, hepatotoxicity, and immunotoxicity as well as the induction of several metabolizing enzymes [15]. Many, if not all of these effects, are due to alterations in gene expression mediated by the aryl hydrocarbon receptor (AhR), a basic-helix-loop-helix-PAS (bHLH-PAS) transcription factor [15, 16]. Ligand binding to the cytoplasmic AhR complex triggers the dissociation of interacting proteins and results in the translocation of the ligand-bound AhR to the nucleus where it heterodimerizes with the aryl hydrocarbon receptor nuclear translocator (ARNT), another member of the bHLH-PAS family. The heterodimer then binds specific DNA elements, termed dioxin response elements (DREs), within the regulatory regions of target genes leading to changes in expression that ultimately result in the observed responses [17]. Although the role of AhR is well established, the gene regulatory pathways responsible for toxicity are poorly understood and warrant further investigation to assess the potential risks to humans and ecologically relevant species.

Hepa1c1c7 cells and C57BL/6 mice are well-established models routinely used to examine the mechanisms of action of TCDD and related compounds. In this study, TCDD-elicited temporal gene expression effects were systematically compared in order to assess the ability of Hepa1c1c7 cells to replicate C57BL/6 hepatic tissue responses. Our results indicate



that several phase I and II metabolizing enzyme responses are aptly reproduced. However, many responses were model-specific and reflect inherent *in vitro* and *in vivo* differences that must be considered in mechanistic studies and during the selection of biomarkers for developing toxicity-screening assays.

## **MATERIALS AND METHODS**

### ***CULTURE AND TREATMENT OF CELL LINES***

Hepalcl7 wild-type and c4 ARNT-deficient cell lines (gifts from O. Hankinson, University of California, Los Angeles, CA) were maintained in phenol-red free DMEM/F12 media (Invitrogen, Carlsbad, CA) supplemented with 5% fetal bovine serum (FBS) (Hyclone, Logan, UT), 2.5 µg/mL amphotericin B (Invitrogen), 2.5 µg/mL amphotericin B (Invitrogen), 50 µg/mL gentamycin (Invitrogen), 100 U/mL penicillin and 100 µg/mL streptomycin (Invitrogen).  $1 \times 10^6$  cells were seeded into T175 culture flasks (Sarstedt, Newton, NC) and incubated under standard conditions (5% CO<sub>2</sub>, 37°C). Time-course studies were performed with wild-type and c4 mutant cells where both were dosed with either 10 nM TCDD (provided by S. Safe, Texas A&M University, College Station, TX) or DMSO (Sigma, St. Louis, MO) vehicle and harvested at 1, 2, 4, 8, 12, 24 or 48 hrs. Additional untreated control cells were harvested at the time of dosing (i.e., 0 hrs). For the dose-response study, wild-type cells were treated with DMSO vehicle or 0.001, 0.01, 0.1, 1.0, 10 or 100 nM TCDD and harvested at 12 hrs. The treatment and harvesting regimen for cell culture studies are illustrated in Additional file 1.

### ***ANIMAL TREATMENT***

The handling and treatment of female C57BL/6 mice has been previously described [18]. Briefly, immature ovariectomized mice were orally gavaged with 30 µg/kg TCDD for the time-

course study and sacrificed at 2, 4, 8, 12, 18, 24, 72 or 168 hrs after treatment. For the dose-response study, mice were treated with 0.001, 0.01, 0.1, 1, 10, 100 or 300 µg/kg TCDD and sacrificed 24 hrs after dosing. Animals were sacrificed by cervical dislocation and tissue samples were removed, weighed, flash frozen in liquid nitrogen, and stored at -80°C until further use.

### ***RNA ISOLATION***

Cells were harvested by scraping in 2.0 mL of Trizol Reagent (Invitrogen). Frozen liver samples (approximately 70 mg) were transferred to 1.0 mL of Trizol Reagent and homogenized in a Mixer Mill 300 tissue homogenizer (Retsch, Germany). Total RNA from each study was isolated according to the manufacturer's protocol with an additional acid phenol:chloroform extraction. Isolated RNA was resuspended in the RNA Storage Solution (Ambion Inc., Austin, TX), quantified ( $A_{260}$ ), and assessed for purity by determining the  $A_{260}/A_{280}$  ratio and by visual inspection of 1.0 µg on a denaturing gel.

### ***MICROARRAY EXPERIMENTAL DESIGN***

Changes in gene expression were assessed using customized cDNA microarrays containing 13,362 features representing 8,284 unique genes. For the time-course study, TCDD-treated samples were compared to time-matched vehicle controls using an independent reference design [19]. In this design, treated Hepa1c1c7 cell or hepatic tissue samples were compared to the corresponding time-matched vehicle control with two independent labelings (dye swaps; Additional file 2). Four replicates of this design were performed, each using independent cell culture samples or different animals. Dose-response changes in gene expression were analyzed using a common reference design in which samples from TCDD-treated cells or mice were co-hybridized with a common vehicle reference (i.e., independent DMSO treated Hepa1c1c7 cell samples, hepatic samples from independent sesame oil treated C57BL/6 mice) using two

independent labelings (Additional file 2). Four replicates with two independent labelings were performed for both *in vitro* and *in vivo* samples. Co-hybridizations of untreated Hepa1c1c7 cells and hepatic tissue from C57BL/6 mice were performed to investigate differences in basal gene expression levels between models (Additional file 2). Four replicates were performed with two independent labelings per sample (dye swap).

More detailed protocols regarding the microarray assay, including microarray preparation, labeling of the cDNA probe, sample hybridization, and washing can be obtained from the dbZach website (<http://dbZach.fst.msu.edu>). Briefly, polymerase chain reaction (PCR) amplified cDNAs were robotically arrayed onto epoxy-coated glass slides (Schott-Nexterion, Duryea, PA) using an Omnigrid arrayer (GeneMachines, San Carlos, CA) equipped with 48 (4 × 12) Chipmaker 2 pins (Telechem) at Michigan State University's Research Technology Support Facility (<http://genomics.msu.edu>). Total RNA (30 µg) was reverse-transcribed in the presence of Cy3- or Cy5-deoxyuridine triphosphate (dUTP) to create fluorescence-labeled cDNA, which was purified using a Qiagen PCR kit (Qiagen, Valencia, CA). Cy3 and Cy5 samples were mixed, vacuum dried, and resuspended in 48 µL of hybridization buffer (40% formamide, 4× SSC, 1% sodium dodecyl sulfate [SDS]) with 20 µg polydA and 20 µg of mouse COT-1 DNA (Invitrogen) as competitor. This probe mixture was heated at 95°C for 3 min and hybridized on the array under a 22 × 60 mm LifterSlip (Erie Scientific Company, Portsmouth, NH) in a light-protected and humidified hybridization chamber (Corning Inc., Corning, NY) for 18-24 hrs in a 42°C water bath. Slides were then washed, dried by centrifugation, and scanned at 635 nm (Cy5) and 532 nm (Cy3) on an Affymetrix 428 Array Scanner (Santa Clara, CA). Images were analyzed for feature and background intensities using GenePix Pro 5.0 (Molecular Devices, Union City, CA).

## ***MICROARRAY DATA QUALITY ASSURANCE, NORMALIZATION AND ANALYSIS***

Microarray data were first passed through a quality assurance protocol prior to further analysis to ensure consistently high quality data throughout the dose-response and time-course studies prior to normalization and further analysis [20]. All the collected data were then normalized using a semi-parametric approach [21]. Empirical Bayes analysis was used to calculate posterior probabilities ( $P_1(t)$  value) of activity on a per gene and time point or dose group basis using the model-based  $t$ -value [22]. The data were filtered using a  $P_1(t)$  cutoff of 0.9999 and  $\pm 1.5$ -fold change to identify the most robust changes in gene expression and to obtain an initial subset of differentially regulated genes for further investigation and data interpretation. Subsequent analysis included agglomerative hierarchical and  $k$ -means clustering using the standard correlation distance metric implemented in GeneSpring 6.0 (Silicon Genetics, Redwood City, CA). Functional categorization of differentially regulated genes were mined and statistically analyzed from Gene Ontology [23] using GOMiner [24].

## ***QUANTITATIVE REAL-TIME PCR ANALYSIS***

For each sample, 1.0  $\mu\text{g}$  of total RNA was reverse transcribed by SuperScript II using an anchored oligo-dT primer as described by the manufacturer (Invitrogen). The cDNA (1.0  $\mu\text{L}$ ) was used as a template in a 30  $\mu\text{L}$  PCR reaction containing 0.1  $\mu\text{M}$  of forward and reverse gene-specific primers designed using Primer3 [25], 3 mM  $\text{MgCl}_2$ , 1.0 mM dNTPs, 0.025 IU AmpliTaq Gold, and 1  $\times$  SYBR Green PCR buffer (Applied Biosystems, Foster City, CA). PCR amplification was conducted in MicroAmp Optical 96-well reaction plates (Applied Biosystems) on an Applied Biosystems PRISM 7000 Sequence Detection System under the following conditions: initial denaturation and enzyme activation for 10 min at 95°C, followed by 40 cycles of 95°C for 15 s and 60°C for 1 min. A dissociation protocol was performed to assess the

specificity of the primers and the uniformity of the PCR-generated products. Each plate contained duplicate standards of purified PCR products of known template concentration covering 7 orders of magnitude to interpolate relative template concentrations of the samples from the standard curves of log copy number versus threshold cycle (Ct). No template controls (NTC) were also included on each plate. Samples with a Ct value within 2 standard deviations of the mean Ct values for the NTCs were considered below the limits of detection. The copy number of each unknown sample for each gene was standardized to the geometric mean of three house-keeping genes ( $\beta$ -actin, Gapd and Hprt) to control for differences in RNA loading, quality, and cDNA synthesis. For graphing purposes, the relative expression levels were scaled such that the expression level of the time-matched control group was equal to 1. Statistical analysis was performed with SAS 8.02 (SAS Institute, Cary, NC). Data were analyzed by analysis of variance (ANOVA) followed by Tukey's post hoc test. Differences between treatment groups were considered significant when  $p < 0.05$ . Official gene names and symbols, RefSeq and Entrez Gene IDs, forward and reverse primer sequences, and amplicon sizes are listed in Table 1.

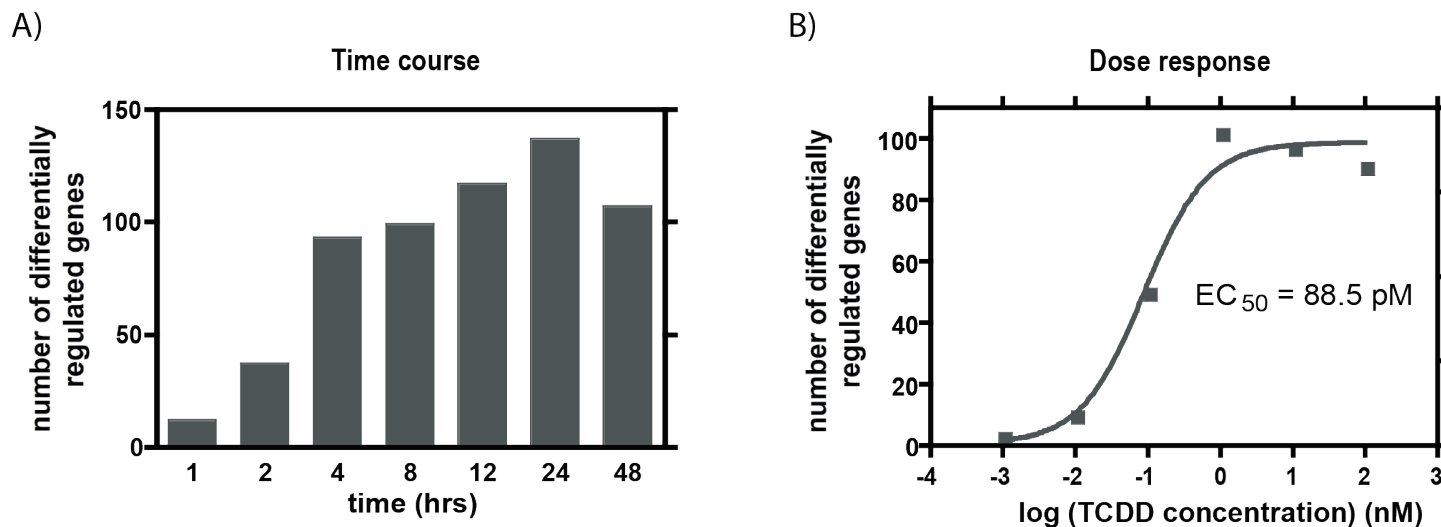
## RESULTS

### ***IN VITRO* MICROARRAY DATA ANALYSIS**

Temporal gene expression profiles were assessed in Hepa1c1c7 wild type cells following treatment with 10 nM TCDD using cDNA microarrays with 13,362 spotted features. Empirical Bayes analysis of the *in vitro* time-course data identified 331 features representing 285 unique genes with a  $P1(t)$  value greater than 0.9999 at one or more time points, and differential expression greater than  $\pm 1.5$  fold relative to time-matched vehicle controls. The number of differentially regulated genes gradually increased from 1 to 24 hrs, followed by a slight decrease at 48 hrs (Figure 2A). *In vitro* dose-response data performed at 12 hrs with TCDD covering 6

**Table 1. Gene names and primer sequences for QRTPCR.**

| RefSeq    | Gene name                                             | Gene symbol | Entrez Gene ID | Forward Primer       | Reverse Primer       | Product Size (bp) |
|-----------|-------------------------------------------------------|-------------|----------------|----------------------|----------------------|-------------------|
| NM_007393 | actin, beta, cytoplasmic                              | Actb        | 11461          | GCTACAGCTTCACCACCACA | TCTCCAGGGAGGAAGAGGAT | 123               |
| NM_009992 | cytochrome P450, family 1, subfamily a, polypeptide 1 | Cyp1a1      | 13076          | AAGTGCAGATGCGGTCTTCT | AAAGTAGGAGGCAGGCACAA | 140               |
| NM_010634 | fatty acid binding protein 5, epidermal               | Fabp5       | 16592          | TGTCATGAACAATGCCACCT | CTGGCAGCTAACTCCTGTCC | 87                |
| NM_008084 | glyceraldehyde-3-phosphate dehydrogenase              | Gapd        | 2597           | GTGGACCTCATGGCCTACAT | TGTGAGGGAGATGCTCAGTG | 125               |
| NM_013556 | hypoxanthine phosphoribosyl transferase               | Hprt        | 24465          | AAGCCTAAGATGAGCGCAAG | TTACTAGGCAGATGGCCACA | 104               |
| NM_010849 | myelocytomatosis oncogene                             | Myc         | 17869          | CTGTGGAGAAGAGGCAAACC | TTGTGCTGGTGAGTGGAGAC | 127               |
| NM_011723 | xanthine dehydrogenase                                | Xdh         | 22436          | GTCGAGGAGATCGAGAATGC | GGTTGTTTCCACTTCCTCCA | 124               |



**Figure 2. Number of genes differentially regulated ( $P(t) > 0.9999$  and  $|\text{fold change}| > 1.5\text{-fold}$ ) as measured by microarray analysis for the (A) time-course and (B) and dose-response studies in mouse hepatoma Hepa1c1c7 cells.**

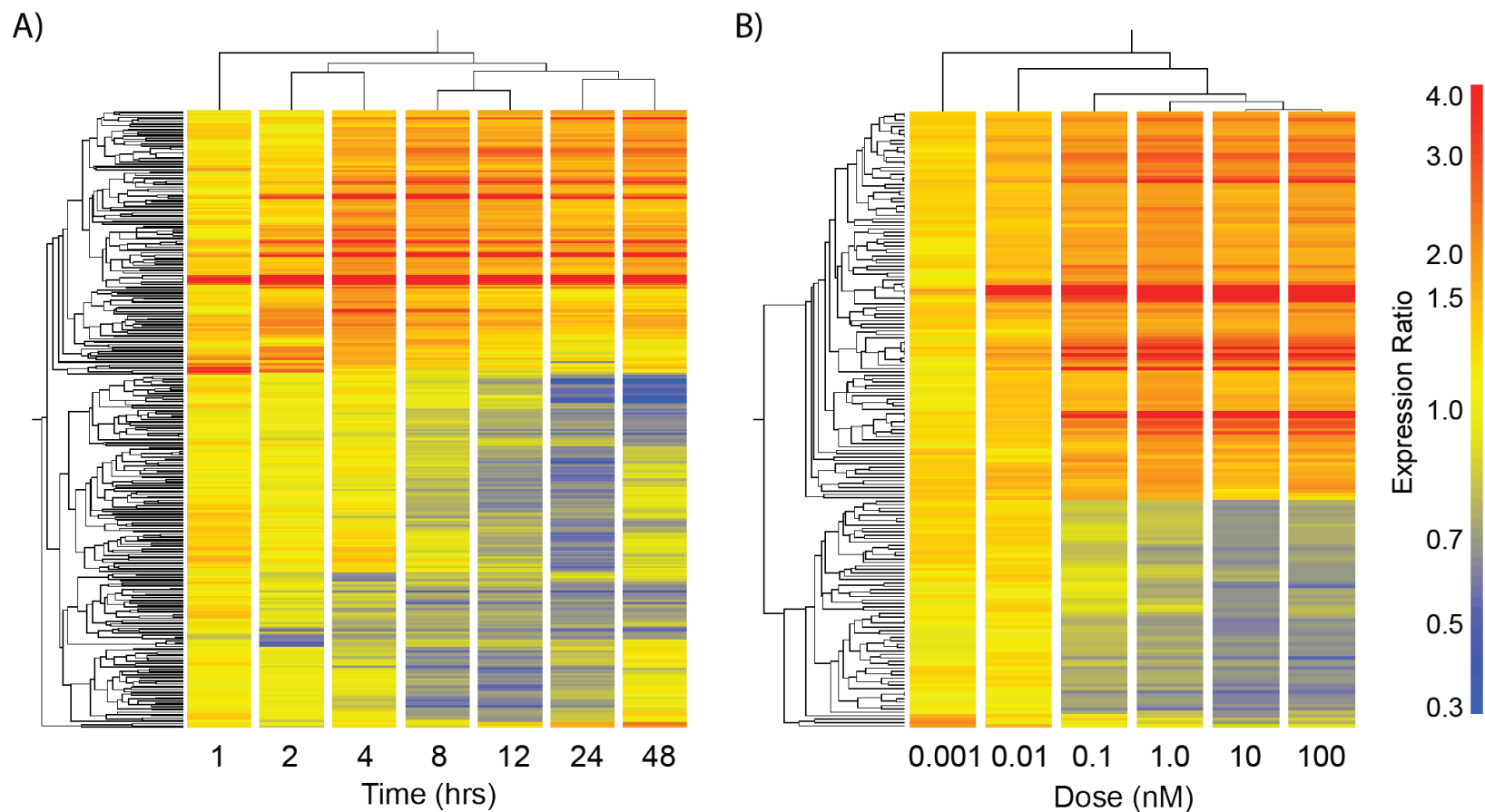
For the time-course study, cells were treated with 10 nM TCDD and harvested at 1, 2, 4, 8, 12, 24 or 48 hrs after treatment. Cells for the 12 hr dose-response study were treated with 0.001, 0.01, 0.1, 1.0, 10 and 100 nM of TCDD.

different concentrations (0.001, 0.01, 0.1, 1.0, 10 and 100 nM), identified 181 features representing 155 unique genes ( $P(t) > 0.9999$  and an absolute-fold change  $> 1.5$  at one or more doses; Figure 2B). Complete *in vitro* time-course and dose-response data are available in Additional file 3 and 4, respectively.

As a control, the gene expression effects elicited by 10 nM TCDD in ARNT-deficient c4 Hepa1c1c7 mutants [26] were examined at 1 and 24 hrs (data not shown). Only ATPase, H<sup>+</sup> transporting, V1 subunit E-like 2 isoform 2 (Atp6v1e2) and SUMO/sentrin specific peptidase 6 (Senp6) exhibited a significant change in expression using the same criteria ( $P(t) > 0.9999$  and an absolute-fold change  $> 1.5$ ). Neither Atp6v1e2 nor Senp6 were among the active genes in wild-type Hepa1c1c7 cells or in C57BL/6 liver samples [18]. These results provide further evidence that the AhR/ARNT signaling pathway mediates TCDD-elicited gene expression responses, which are consistent with *in vivo* microarray results with AhR knockout mice [27].

Hierarchical clustering of the genes expressed in Hepa1c1c7 time-course assays indicate that 2 and 4 hrs were most similar, as were 8 and 12 hrs, and 24 and 48 hrs, while the 1 hr time point was segregated (Figure 3A). A strong dose-response relationship was also evident with clusters sequentially branching out with increasing concentration (Figure 3B). At 12 hrs, 117 genes were differentially expressed with 112 exhibiting a dose-dependent response. Moreover, the fold changes measured in both the time-course and dose-response studies using 10 nM TCDD were comparable. For example, xanthine dehydrogenase (Xdh) and NAD(P)H dehydrogenase, quinone 1 (Nqo1) were induced 2.39- and 4.89-fold respectively in the time-course study, and 2.93- and 4.71-fold in the dose-response study. There is a strong correlation ( $R = 0.97$ ) between the differentially expressed genes at 12 hrs in the time-course with the differentially regulated genes in the dose-response study at 10 nM, demonstrating the





**Figure 3. Hierarchical clustering of the differentially regulated gene lists for A) temporal and B) dose-response microarray studies in mouse hepatoma Hepa1c1c7 cells.**

The results illustrate time- and dose-dependent clustering patterns. From the A) temporal results, the early (2 hr and 4 hr), intermediate (8 hr and 12 hr) and late (24 hr and 48 hr) time points cluster separately, while the 1hr time point clusters alone. Results from the B) dose-response show that the highest doses clustered together, while the remaining doses branched out in a dose-dependent manner.

reproducibility between independent studies and providing further evidence that these genes are regulated by TCDD.

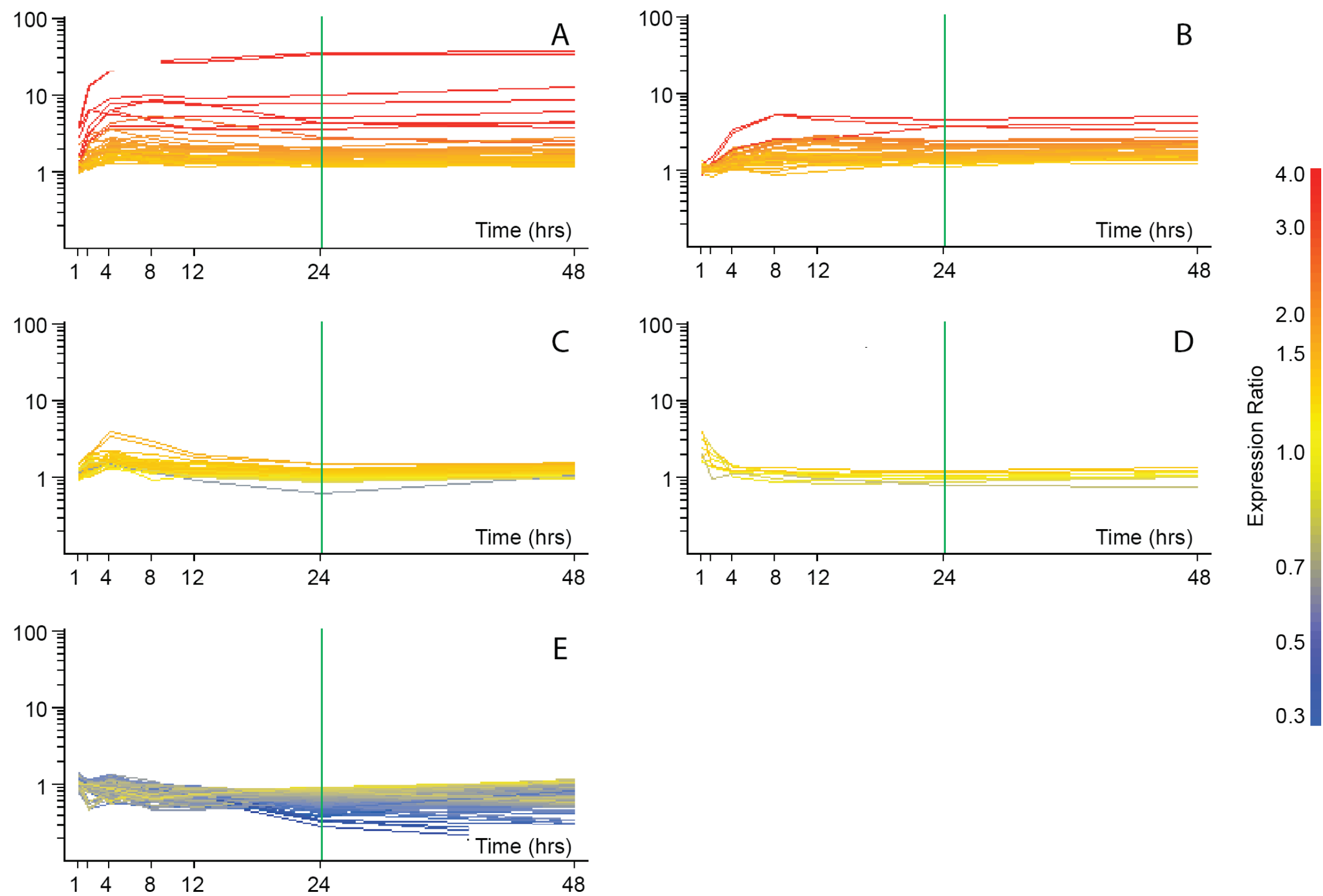
The list of temporally regulated genes was subjected to *k*-means clustering using the standard correlation distance metrics. Five *k*-means clusters best characterized the dataset and identified clusters representing A) up-regulated early and sustained, B) up-regulated intermediate and sustained, C) up-regulated intermediate, D) up-regulated immediate, and E) down-regulated late (Figure 4). These were comparable to the *k*-means clusters identified in hepatic tissue of C57BL/6 mice following treatment with 30 µg/kg TCDD [18]. Although, no discernible functional category is over-represented in any one cluster, the sustained up-regulation of early (Cluster A) and intermediate (Cluster B) responding genes include classic TCDD-responsive genes such as cytochrome P450, family 1, subfamily a, polypeptide 1 (Cyp1a1), Xdh and Nqo1. Many down-regulated late genes were associated with cell cycle regulation such as myelocytomatosis oncogene (Myc). Additionally, targets of Myc, including cyclin D1 and ornithine decarboxylase (Odc1), were also down-regulated suggesting a mechanism for cell cycle arrest [28-30], a common *in vitro* response to TCDD.

#### **CLASSIFICATION OF GENE EXPRESSION RESPONSES FOR COMMON REGULATED GENES**

Using the same filtering criteria ( $P(t) > 0.9999$  and an absolute-fold change  $> 1.5$ ), 678 features representing 619 unique genes were differentially expressed as previously reported in a time-course study conducted in hepatic tissue from C57BL/6 mice orally gavaged with 30 µg/kg TCDD [18]. The number of responsive *in vivo* genes and their temporal expression patterns closely paralleled the results from this *in vitro* study. The fewest number of active genes was observed at 2 hrs, followed by a large increase at 4 hrs, which was sustained to 72 hrs. However, the substantial increase in expressed *in vivo* genes at 168 hrs was attributed to triglyceride

**Figure 4. *K*-means clustering of temporally differentially regulated genes *in vitro*.**

Five *k*-mean clusters corresponding to (A) up-early and sustained, (B) up-intermediate and sustained, (C) up-regulated intermediate, (D) up-regulated immediate, and (E) down-regulated late. Time and expression ratio are indicated on the x- and y-axis respectively. The color of individual gene expression profiles reflects the expression ratio observed at 24 hrs.



accumulation and immune cell infiltration, which were not observed in Hepa1c1c7 cells. This list of 619 of *in vivo* genes served as the basis for subsequent comparisons against TCDD-elicited *in vitro* responses.

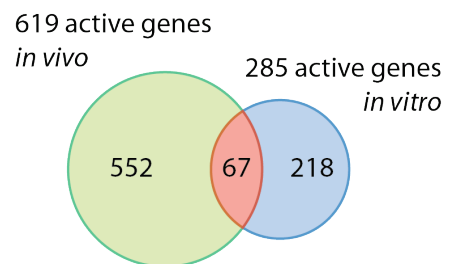
Comparison of *in vitro* and *in vivo* differentially expressed gene lists identified common and model-specific responses (Figure 5A). TCDD treatment resulted in a total of 838 regulated genes in either model, and with 67 common to both. TCDD elicited 218 gene expression changes unique to Hepa1c1c7 cells while 552 genes were specific to C57BL/6 hepatic samples. Although 67 genes were regulated in both models, not all possessed similar temporal patterns of expression. Contingency analysis using a  $2 \times 2$  table and the  $\chi^2$  test resulted in a p-value  $< 0.001$  ( $\alpha = 0.05$ ) that illustrates a statistically significant association between the lists of differentially regulated genes *in vitro* and *in vivo*. Further stratification revealed genes that were either induced in both models (class I), repressed in both models (class II), induced *in vivo* while repressed *in vitro* (class III), or repressed *in vivo* while induced *in vitro* (class IV; Figure 5B). Genes regulated in a similar fashion in both models (classes I and II) accounted for 49 of the 67 common active genes, while the remaining genes exhibited divergent expression profiles (classes III and IV). Hierarchical clustering of the temporal expression values for the 67 overlapping genes identified the same four classes (Figure 5C). The pattern across-model and -time illustrates that the earliest time points (i.e., 1 hr *in vitro* and 2 hr *in vivo* time points) cluster together while the remaining clusters branch into *in vitro* or *in vivo* clusters according to time. These results suggest that potential biomarkers of acute TCDD-mediated responses may best be predicted by the immediate-early *in vitro* gene responses.

*In vitro* and *in vivo* induced genes (class I) include xenobiotic and oxidoreductase enzymes such as abhydrolase domain containing 6 (Abhd6), Cyp1a1, dehydrogenase/reductase

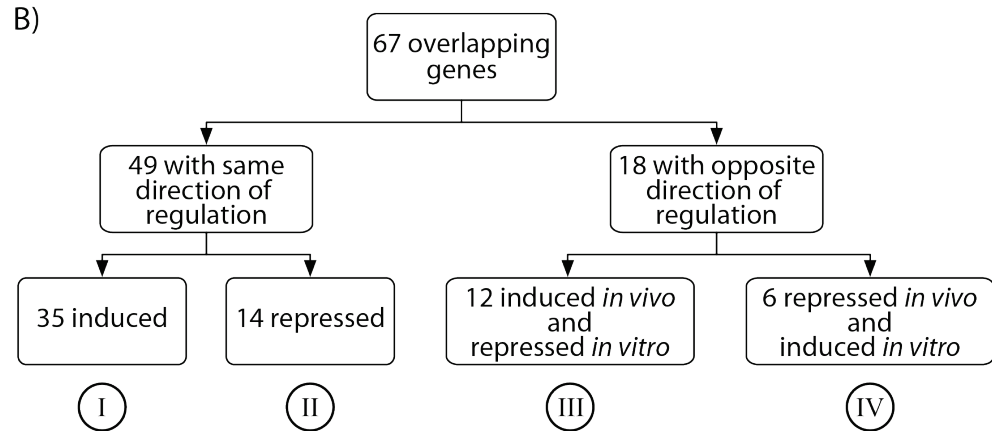
**Figure 5. Comparison of common significant *in vitro* and *in vivo* TCDD-elicited time-dependent gene expression changes.**

A) 285 differentially regulated *in vitro* genes and 619 differentially regulated *in vivo* genes were identified, with 67 genes common to both studies. B) The temporal gene expression profiles from both studies were categorized into (I) induced in both, (II) repressed in both, (III) induced *in vivo* and repressed *in vitro*, and (IV) repressed *in vitro* and induced *in vivo*. C) Hierarchical clustering identified similar classification groups. Clustering across both time and model, separated samples from *in vitro* and *in vivo*, with the exception of the early time points from both studies (1 hr *in vitro* and 2 hr *in vivo*), which clustered together. \* identifies *in vitro* time points.

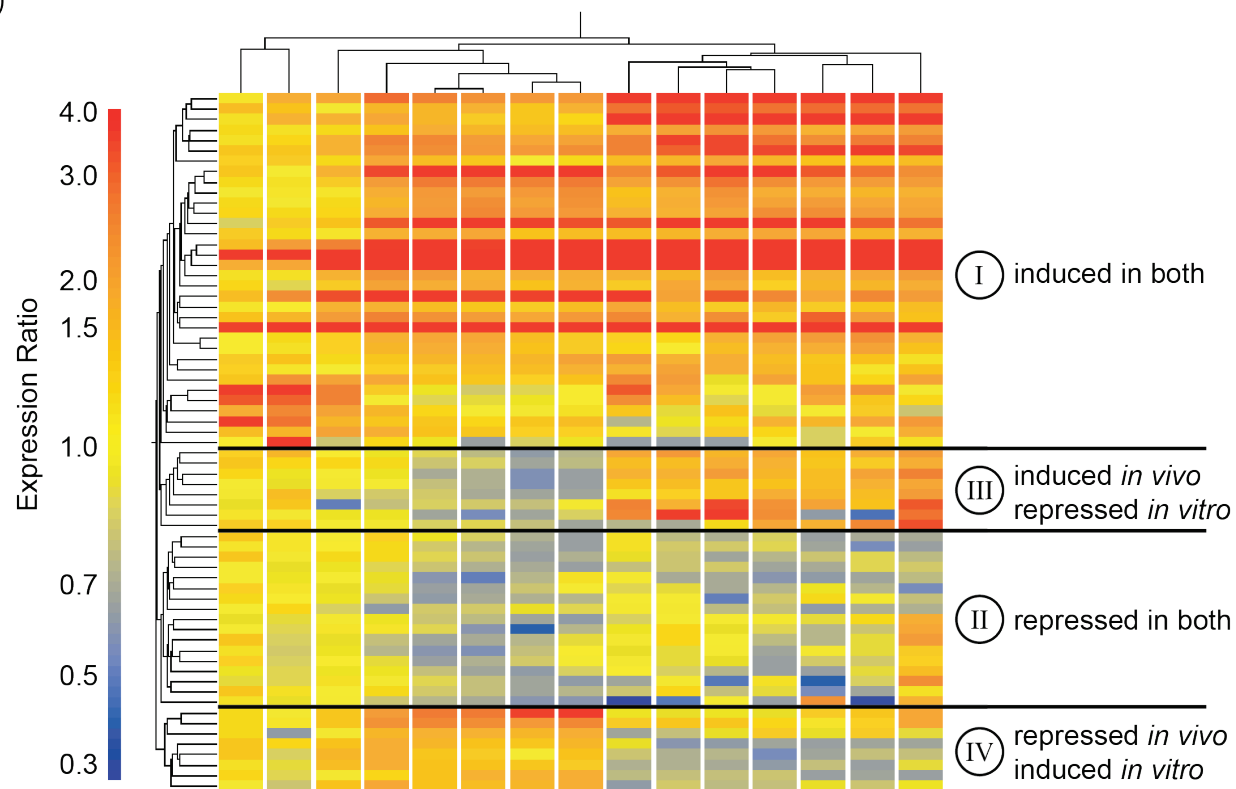
A)



B)



C)



(SDR family) member 3 (Dhrs3), Nqo1, prostaglandin- endoperoxide synthase 1 (Ptgs1), UDP-glucose dehydrogenase (Ugdh), and Xdh (Table 2). These genes have previously been reported to be TCDD-responsive [18, 31], with Cyp1a1 and Nqo1 being members of the “AhR gene battery” [32]. Glutathione S-transferase, alpha 4 (Gsta4) was also induced *in vitro* and *in vivo*, 1.7- and 2.0-fold respectively, consistent with TCDD-mediated induction of phase I and II metabolizing enzymes. Of the 35 genes responding similarly in both models, approximately 71% of were similarly up-regulated (class I) while the remaining genes were repressed across both models (class II). Repressed class II genes include minichromosome maintenance deficient 6 (Mcm6), glycerol kinase (Gyk) and ficolin A (Fcna) (repressed 1.6-, 1.6- and 1.7-fold *in vitro*, respectively). Overall, repressed genes did not share any common discernible biological function.

Forty-two of the 67 common differentially expressed genes were dose-responsive at 12 and 24 hrs *in vitro* and *in vivo*, respectively, further suggesting the role of the AhR in mediating these responses. Microarray-based EC<sub>50</sub> values spanned at least 3 orders of magnitude ranging from 0.05 µg/kg to >150 µg/kg *in vivo*, and 0.0012 nM to 2.4 nM *in vitro* (Table 2). Cyp1a1, the prototypical marker of TCDD exposure, had EC<sub>50</sub> values of 0.05 µg/kg and 0.014 nM, *in vivo* and *in vitro*, respectively, and was induced 38-fold in both time-course studies. Complete data sets for the *in vivo* time-course and dose-responses experiments are available in Additional file 5 and 6.

Of the 67 overlapping genes, 18 exhibited divergent temporal profiles (classes III and IV). Class III contains 12 genes induced *in vivo* but repressed *in vitro*, while 6 were repressed *in vivo* and induced *in vitro* (class IV). Examples of these genes include Myc (class III) and B-cell translocation gene 2 (Btg2, class IV), which are both involved in regulating cell cycle



**Table 2. Classification of common differentially regulated temporal gene expression responses to TCDD in both *in vitro* and *in vivo* models.**

| Accession                                                       | Gene name                                                                 | Gene symbol   | Entrez Gene ID | In vivo                  |                                            |                                         | In vitro                 |                                     |                                         |
|-----------------------------------------------------------------|---------------------------------------------------------------------------|---------------|----------------|--------------------------|--------------------------------------------|-----------------------------------------|--------------------------|-------------------------------------|-----------------------------------------|
|                                                                 |                                                                           |               |                | Fold change <sup>a</sup> | Time points <sup>b</sup>                   | EC <sub>50</sub> <sup>c,d</sup> (µg/kg) | Fold change <sup>a</sup> | Time points <sup>b</sup>            | EC <sub>50</sub> <sup>c,d</sup> (µg/kg) |
| I) Induced both <i>in vivo</i> and <i>in vitro</i> <sup>e</sup> |                                                                           |               |                |                          |                                            |                                         |                          |                                     |                                         |
| BE689910                                                        | RIKEN cDNA 2310001H12 gene                                                | 2310001H12Rik | 69504          | 2.7                      | 2 <sup>f</sup> , 168                       | 48.02                                   | 3.9                      | 1 <sup>f</sup>                      | ND                                      |
| BF226070                                                        | RIKEN cDNA 2600005C20 gene                                                | 2600005C20Rik | 72462          | 2.1                      | 4, 12, 18, 24 <sup>f</sup> , 72, 168       | 2.18                                    | 2.3                      | 4, 8, 12 <sup>f</sup> , 24, 48      | 265.50                                  |
| AI043124                                                        | RIKEN cDNA 2810003C17 gene                                                | 2810003C17Rik | 108897         | 1.6                      | 12 <sup>f</sup>                            | 37.02                                   | 1.7                      | 4 <sup>f</sup>                      | ND                                      |
| AW537038                                                        | expressed sequence AA959742                                               | AA959742      | 98238          | 7.2                      | 4, 8, 12 <sup>f</sup> , 18, 24, 72, 168    | 1.71                                    | 5.2                      | 4, 8 <sup>f</sup> , 12, 24, 48      | 67.79                                   |
| W34507                                                          | abhydrolase domain containing 6                                           | Abhd6         | 66082          | 1.7                      | 4, 8 <sup>f</sup> , 12, 18, 24, 72, 168    | 154.30                                  | 1.5                      | 48 <sup>f</sup>                     | 138.50                                  |
| NM_026410                                                       | cell division cycle associated 5                                          | Cdca5         | 67849          | 8.8                      | 4, 8, 12, 18, 24, 72 <sup>f</sup> , 168    | ND                                      | 1.7                      | 4 <sup>f</sup>                      | ND                                      |
| BG063743                                                        | craniofacial development protein 1                                        | Cfdp1         | 23837          | 3.6                      | 4, 8, 12 <sup>f</sup> , 18, 24, 72, 168    | 14.27                                   | 2.3                      | 4 <sup>f</sup> , 8, 12, 24, 48      | 42.64                                   |
| AA073604                                                        | procollagen, type I, alpha 1                                              | Col1a1        | 12842          | 1.7                      | 18, 24, 72 <sup>f</sup>                    | 0.65                                    | 1.6                      | 4, 8, 12 <sup>f</sup>               | 17.25                                   |
| NM_009992                                                       | cytochrome P450, family 1, subfamily a, polypeptide 1                     | Cyp1a1        | 13076          | 38.4                     | 2, 4, 8, 12, 18, 24 <sup>f</sup> , 72, 168 | 0.05                                    | 37.7                     | 1, 2, 4, 8, 12, 24, 48 <sup>f</sup> | 14.06                                   |
| BE457542                                                        | dehydrogenase/reductase (SDR family) member 3                             | Dhrs3         | 20148          | 2.0                      | 4, 8, 12 <sup>f</sup> , 18, 72, 168        | 0.67                                    | 1.5                      | 8 <sup>f</sup>                      | 2.43                                    |
| AW552715                                                        | DnaJ (Hsp40) homolog, subfamily B, member 11                              | Dnajb11       | 67838          | 1.7                      | 12, 18, 24 <sup>f</sup> , 168              | 3.95                                    | 1.6                      | 8, 12 <sup>f</sup>                  | 9.85                                    |
| AK015223                                                        | dermatan sulphate proteoglycan 3                                          | Dspg3         | 13516          | 6.2                      | 4, 8, 12, 18, 24 <sup>f</sup> , 72, 168    | 0.13                                    | 8.4                      | 2, 4, 8, 12, 24, 48 <sup>f</sup>    | 16.34                                   |
| NM_008655                                                       | growth arrest and DNA-damage-inducible 45 beta                            | Gadd45b       | 17873          | 4.6                      | 2 <sup>f</sup> , 4, 72                     | 133.30                                  | 3.7                      | 1 <sup>f</sup> , 2                  | 1440.00                                 |
| W54349                                                          | glutathione S-transferase, alpha 4                                        | Gsta4         | 14860          | 2.0                      | 18, 24, 72 <sup>f</sup>                    | 0.48                                    | 1.7                      | 8 <sup>f</sup> , 12                 | 56.38                                   |
| BG067127                                                        | interferon regulatory factor 1                                            | Irf1          | 16362          | 1.5                      | 168 <sup>f</sup>                           | ND                                      | 1.7                      | 2 <sup>f</sup> , 4                  | ND                                      |
| AA015278                                                        | integrin beta 1 (fibronectin receptor beta)                               | Itgb1         | 16412          | 1.6                      | 4, 18, 24, 168 <sup>f</sup>                | 97.23                                   | 4.2                      | 4, 8, 12, 24, 48 <sup>f</sup>       | 72.92                                   |
| AA041752                                                        | Jun proto-oncogene related gene d1                                        | Jund1         | 16478          | 2.0                      | 12 <sup>f</sup> , 18, 24                   | 0.99                                    | 2.1                      | 4, 8, 12, 24, 48 <sup>f</sup>       | 50.34                                   |
| BF538945                                                        | lectin, mannose-binding, 1                                                | Lman1         | 70361          | 1.9                      | 12, 72, 168 <sup>f</sup>                   | 13.49                                   | 2.0                      | 4 <sup>f</sup> , 8, 24, 48          | 40.72                                   |
| BG066626                                                        | lipin 2                                                                   | Lpin2         | 64898          | 3.0                      | 4, 12, 24 <sup>f</sup> , 72                | 3.13                                    | 2.3                      | 2, 4 <sup>f</sup> , 8, 12, 24, 48   | 23.83                                   |
| BI440950                                                        | leucine rich repeat containing 39                                         | Lrrc39        | 109245         | 2.9                      | 2 <sup>f</sup> , 4                         | 49.71                                   | 3.1                      | 1 <sup>f</sup> , 2                  | 68.57                                   |
| AW413953                                                        | mitochondrial ribosomal protein L37                                       | Mrpl37        | 56280          | 8.3                      | 2, 4, 8 <sup>f</sup> , 12, 18, 24, 72, 168 | 8.77                                    | 2.7                      | 2, 4 <sup>f</sup> , 8, 12, 24, 48   | 49.59                                   |
| BE623489                                                        | NAD(P)H dehydrogenase, quinone 1                                          | Nqo1          | 18104          | 4.6                      | 4, 8, 12 <sup>f</sup> , 18, 24, 72, 168    | 1.00                                    | 5.2                      | 4, 8 <sup>f</sup> , 12, 24, 48      | 33.74                                   |
| NM_026550                                                       | PAK1 interacting protein 1                                                | Pak1ip1       | 68083          | 3.8                      | 4, 8, 12, 18, 24, 72 <sup>f</sup> , 168    | 0.26                                    | 2.2                      | 4 <sup>f</sup> , 8, 12, 24, 48      | 7.00                                    |
| AA152754                                                        | prostaglandin-endoperoxide synthase 1                                     | Ptgs1         | 19224          | 1.6                      | 168 <sup>f</sup>                           | 1.11                                    | 2.3                      | 4, 8 <sup>f</sup> , 12, 24, 48      | 37.96                                   |
| BG063583                                                        | solute carrier family 20, member 1                                        | Slc20a1       | 20515          | 2.2                      | 2, 4 <sup>f</sup> , 8                      | ND                                      | 1.8                      | 2 <sup>f</sup> , 4                  | ND                                      |
| AJ223958                                                        | solute carrier family 27 (fatty acid transporter), member 2               | Slc27a2       | 26458          | 1.9                      | 12 <sup>f</sup> , 18, 24, 72, 168          | 2.88                                    | 2.1                      | 8, 12 <sup>f</sup> , 24, 48         | 17.42                                   |
| BG066820                                                        | solute carrier family 6 (neurotransmitter transporter, taurine), member 6 | Slc6a6        | 21366          | 1.8                      | 4 <sup>f</sup> , 12                        | 2.48                                    | 1.7                      | 48 <sup>f</sup>                     | 3.06                                    |
| AI592773                                                        | suppression of tumorigenicity 5                                           | St5           | 76954          | 1.6                      | 8, 12 <sup>f</sup>                         | 28.85                                   | 1.7                      | 4 <sup>f</sup> , 8, 12              | 14.69                                   |

Table 2. (cont'd).

| Accession                                                                    | Gene name                                                              | Gene symbol   | Entrez<br>Gene ID | <i>In vivo</i>              |                                           |                                            | <i>In vitro</i>             |                                      |                                            |
|------------------------------------------------------------------------------|------------------------------------------------------------------------|---------------|-------------------|-----------------------------|-------------------------------------------|--------------------------------------------|-----------------------------|--------------------------------------|--------------------------------------------|
|                                                                              |                                                                        |               |                   | Fold<br>change <sup>a</sup> | Time points <sup>b</sup>                  | EC <sub>50</sub> <sup>c,d</sup><br>(µg/kg) | Fold<br>change <sup>a</sup> | Time points <sup>b</sup>             | EC <sub>50</sub> <sup>c,d</sup><br>(µg/kg) |
| BG067168                                                                     | TCDD-inducible poly(ADP-ribose)<br>polymerase                          | Tiparp        | 99929             | 10.3                        | 2, 4 <sup>f</sup> , 12, 18, 24, 72, 168   | 36.49                                      | 6.4                         | 1, 2 <sup>f</sup> , 4, 8, 12, 24, 48 | 18.03                                      |
| BG065761                                                                     | tumor necrosis factor, alpha-induced<br>protein 2                      | Tnfaip2       | 21928             | 5.5                         | 2, 4 <sup>f</sup> , 12, 18, 72            | 36.41                                      | 6.3                         | 2, 4 <sup>f</sup> , 8, 12, 24, 48    | 41.15                                      |
| AA067191                                                                     | UDP-glucose dehydrogenase                                              | Ugdh          | 22235             | 3.1                         | 4, 8, 12 <sup>f</sup> , 18, 24, 72, 168   | 0.79                                       | 1.5                         | 2, 4, 8, 12 <sup>f</sup> , 48        | 4.33                                       |
| NM_011709                                                                    | whey acidic protein                                                    | Wap           | 22373             | 5.9                         | 2, 4, 8, 12, 18, 24, 72, 168 <sup>f</sup> | 0.12                                       | 4.2                         | 2, 4, 8, 12, 24, 48 <sup>f</sup>     | 17.44                                      |
| BG075778                                                                     | xanthine dehydrogenase                                                 | Xdh           | 22436             | 2.7                         | 4, 8, 12 <sup>f</sup> , 18, 24, 72, 168   | 1.24                                       | 2.6                         | 4, 8, 12, 24, 48 <sup>f</sup>        | 34.92                                      |
| BG073881                                                                     | zinc finger protein 36, C3H type-like 1                                | Zfp36l1       | 12192             | 2.2                         | 2 <sup>f</sup>                            | ND                                         | 1.7                         | 1, 2 <sup>f</sup>                    | 2427.00                                    |
| AA031146                                                                     | zinc finger protein 672                                                | Zfp672        | 319475            | 1.6                         | 4 <sup>f</sup>                            | 3.09                                       | 1.5                         | 2 <sup>f</sup>                       | ND                                         |
| <b>II) Repressed both <i>in vivo</i> and <i>in vitro</i><sup>e</sup></b>     |                                                                        |               |                   |                             |                                           |                                            |                             |                                      |                                            |
| BG146493                                                                     | RIKEN cDNA 6330406L22 gene                                             | 6330406L22Rik | 70719             | -1.5                        | 18 <sup>f</sup>                           | 0.51                                       | -1.8                        | 8, 12 <sup>f</sup>                   | 25.67                                      |
| AA140059                                                                     | DNA methyltransferase (cytosine-5) 1                                   | Dnmt1         | 13433             | -1.9                        | 168 <sup>f</sup>                          | ND                                         | -1.6                        | 8 <sup>f</sup> , 12                  | ND                                         |
| AI327022                                                                     | ficolin A                                                              | Fcna          | 14133             | -1.6                        | 18, 24 <sup>f</sup>                       | ND                                         | -1.7                        | 12, 24 <sup>f</sup>                  | ND                                         |
| AA288963                                                                     | fibrinogen-like protein 1                                              | Fgl1          | 234199            | -1.9                        | 24 <sup>f</sup>                           | ND                                         | -1.5                        | 24 <sup>f</sup> , 48                 | ND                                         |
| BE626913                                                                     | GTP binding protein 6 (putative)                                       | Gtpbp6        | 107999            | -3.4                        | 24, 72 <sup>f</sup>                       | ND                                         | -1.7                        | 24, 48 <sup>f</sup>                  | 116.40                                     |
| AA275564                                                                     | glycerol kinase                                                        | Gyk           | 14933             | -1.5                        | 12 <sup>f</sup>                           | 10.2                                       | -1.6                        | 24 <sup>f</sup>                      | ND                                         |
| BG070106                                                                     | lipocalin 2                                                            | Lcn2          | 16819             | -2.8                        | 24 <sup>f</sup>                           | ND                                         | -1.5                        | 24, 48 <sup>f</sup>                  | ND                                         |
| AW049427                                                                     | leucine zipper domain protein                                          | Lzf           | 66049             | -1.6                        | 24 <sup>f</sup>                           | ND                                         | -1.6                        | 48 <sup>f</sup>                      | 78.29                                      |
| AA016759                                                                     | minichromosome maintenance deficient 6<br>mitochondrial methionyl-tRNA | Mcm6          | 17219             | -1.6                        | 18 <sup>f</sup>                           | 3.34                                       | -1.6                        | 8 <sup>f</sup>                       | 58.04                                      |
| BF011268                                                                     | formyltransferase                                                      | Mtfmt         | 69606             | -1.8                        | 24, 72 <sup>f</sup> , 168                 | ND                                         | -1.6                        | 24, 48 <sup>f</sup>                  | ND                                         |
| AA683699                                                                     | RNA (guanine-7-) methyltransferase                                     | Rnmt          | 67897             | -2.0                        | 12 <sup>f</sup>                           | ND                                         | -1.6                        | 8 <sup>f</sup>                       | ND                                         |
|                                                                              | syntrophin, gamma 1                                                    | Sntg1         | 71096             | -1.6                        | 24 <sup>f</sup>                           | 15.27                                      | -1.7                        | 4 <sup>f</sup>                       | 66.61                                      |
| AA199550                                                                     | syntaxin 12                                                            | Stx12         | 100226            | -1.5                        | 18 <sup>f</sup>                           | ND                                         | -1.6                        | 48 <sup>f</sup>                      | ND                                         |
| AA047942                                                                     | thymidine kinase 1                                                     | Tk1           | 21877             | -1.7                        | 18 <sup>f</sup> , 24, 72                  | 0.34                                       | -2.0                        | 8, 12 <sup>f</sup>                   | 153.90                                     |
| <b>III) Induced <i>in vivo</i> and repressed <i>in vitro</i><sup>e</sup></b> |                                                                        |               |                   |                             |                                           |                                            |                             |                                      |                                            |
| AA122925                                                                     | carbonic anhydrase 2                                                   | Car2          | 12349             | 2.4                         | 12, 72, 168 <sup>f</sup>                  | 2.00                                       | -1.8                        | 24 <sup>f</sup> , 48                 | 55.96                                      |
| AI327078                                                                     | coactosin-like 1                                                       | Cotl1         | 72042             | 1.6                         | 168 <sup>f</sup>                          | ND                                         | -1.7                        | 24, 48 <sup>f</sup>                  | 25.19                                      |
| NM_007935                                                                    | enhancer of polycomb homolog 1                                         | Epc1          | 13831             | 1.6                         | 168 <sup>f</sup>                          | 1.16                                       | -2.7                        | 12, 24 <sup>f</sup>                  | 75.21                                      |
| BC002008                                                                     | fatty acid binding protein 5, epidermal                                | Fabp5         | 16592             | 3.9                         | 8, 12 <sup>f</sup>                        | 2.43                                       | -1.9                        | 8, 12 <sup>f</sup> , 24              | 54.14                                      |

Table 2. (cont'd).

| Accession                                                                   | Gene name                                                              | Gene symbol | Entrez<br>Gene ID | <i>In vivo</i>              |                                      |                                            | <i>In vitro</i>             |                          |                                            |
|-----------------------------------------------------------------------------|------------------------------------------------------------------------|-------------|-------------------|-----------------------------|--------------------------------------|--------------------------------------------|-----------------------------|--------------------------|--------------------------------------------|
|                                                                             |                                                                        |             |                   | Fold<br>change <sup>a</sup> | Time points <sup>b</sup>             | EC <sub>50</sub> <sup>c,d</sup><br>(µg/kg) | Fold<br>change <sup>a</sup> | Time points <sup>b</sup> | EC <sub>50</sub> <sup>c,d</sup><br>(µg/kg) |
| NM_026320                                                                   | growth arrest and DNA-damage-inducible,<br>gamma interacting protein 1 | Gadd45gip1  | 102060            | 1.8                         | 168 <sup>f</sup>                     | 4.67                                       | -1.5                        | 8 <sup>f</sup>           | 40.49                                      |
| W11419                                                                      | inhibitor of DNA binding 3                                             | Id3         | 15903             | 1.8                         | 168 <sup>f</sup>                     | 0.34                                       | -1.5                        | 24, 48 <sup>f</sup>      | 88.83                                      |
| AA009268                                                                    | myelocytomatosis oncogene                                              | Myc         | 17869             | 3.7                         | 4, 12 <sup>f</sup> , 168             | 5.59                                       | -2.2                        | 2 <sup>f</sup>           | 148.40                                     |
| NM_011033                                                                   | poly A binding protein, cytoplasmic 2                                  | Pabpc2      | 18459             | 7.0                         | 2 <sup>f</sup>                       | ND                                         | -1.6                        | 12 <sup>f</sup>          | ND                                         |
|                                                                             | REST corepressor 1                                                     | Rcor1       | 217864            | 1.9                         | 4, 8, 18, 72 <sup>f</sup> , 168      | 3.70                                       | -1.6                        | 24 <sup>f</sup>          | 116.50                                     |
| BE980584                                                                    | secretory granule neuroendocrine protein<br>1, 7B2 protein             | Sgne1       | 20394             | 3.3                         | 168 <sup>f</sup>                     | 0.74                                       | -1.5                        | 48 <sup>f</sup>          | 175.00                                     |
| AA462951                                                                    | transcription factor 4                                                 | Tcf4        | 21413             | 1.6                         | 12 <sup>f</sup> , 168                | 5.77                                       | -1.5                        | 24 <sup>f</sup>          | 74.44                                      |
| AA003942                                                                    | tenascin C                                                             | Tnc         | 21923             | 1.6                         | 168 <sup>f</sup>                     | 0.37                                       | -1.8                        | 24 <sup>f</sup> , 48     | 59.34                                      |
| <b>IV) Repressed <i>in vivo</i> and induced <i>in vitro</i><sup>e</sup></b> |                                                                        |             |                   |                             |                                      |                                            |                             |                          |                                            |
| W36712                                                                      | B-cell translocation gene 2, anti-<br>proliferative                    | Btg2        | 12227             | -1.8                        | 18 <sup>f</sup> , 24                 | ND                                         | 1.5                         | 4 <sup>f</sup>           | ND                                         |
| AA174215                                                                    | cathepsin L                                                            | Ctsl        | 13039             | -1.6                        | 24 <sup>f</sup> , 72, 168            | ND                                         | 1.6                         | 8, 48 <sup>f</sup>       | ND                                         |
| AA419858                                                                    | cysteine rich protein 61                                               | Cyr61       | 16007             | -1.6                        | 2 <sup>f</sup>                       | 0.07                                       | 1.6                         | 8, 48 <sup>f</sup>       | ND                                         |
| AW488956                                                                    | polo-like kinase 3                                                     | Plk3        | 12795             | -1.6                        | 4 <sup>f</sup>                       | ND                                         | 1.6                         | 4, 48 <sup>f</sup>       | ND                                         |
| BG068288                                                                    | solute carrier organic anion transporter<br>family, member 1b2         | Slco1b2     | 28253             | -1.7                        | 8 <sup>f</sup> , 12, 18, 24, 72, 168 | ND                                         | 1.6                         | 4 <sup>f</sup>           | 1.18                                       |
| NM_011470                                                                   | small proline-rich protein 2D                                          | Sprr2d      | 20758             | -1.6                        | 18 <sup>f</sup> , 72                 | 1.97                                       | 1.6                         | 4 <sup>f</sup>           | ND                                         |

<sup>a</sup> Maximum absolute-fold change determined by microarray analysis

<sup>b</sup> Time point where genes are differentially regulated with P1(t) > 0.9999 and |fold change| > 1.5

<sup>c</sup> EC<sub>50</sub> valued determined from microarray results

<sup>d</sup> ND = not determined from microarray results

<sup>e</sup> Classification groups as defined in Figure 5B

<sup>f</sup> Time point representing the maximum |fold change|

progression [30, 33-37]. Myc was induced 3.7-fold *in vivo* and repressed 2.2-fold *in vitro*, while Btg2 was repressed 1.8-fold *in vivo* and induced 1.5-fold *in vitro*.

In addition to the regulated genes common to both models, 218 *in vitro*- and 559 *in vivo*-specific genes were identified. Many of the unique *in vitro* responses are involved in cell cycle regulation, including cyclins D1 and B2 (Table 3). Cyclin D1, which complexes with cyclin-dependent kinase 4 (Cdk4) to regulate the progression from G<sub>1</sub> to S phase [38, 39], was down-regulated early and repressed 1.7-fold to 48 hrs. Furthermore, cyclin B2 and cell division cycle 2 homolog A (Cdc2a), which interact to form an active kinase required for G<sub>2</sub> promotion, were down-regulated, 1.8-fold and 1.5-fold, respectively. In addition to cell cycle related genes, UDP glucuronosyltransferase 1 family, polypeptide A2 (Ugt1a2), a phase II metabolizing enzyme, was induced 2.8-fold *in vitro*, but not significantly regulated *in vivo*.

Analysis of the C57BL/6 hepatic time-course identified 552 unique genes that were solely regulated *in vivo*. These included TCDD-induced transcripts for microsomal epoxide hydrolase 1 (Ephx1) and carbonyl reductase 3 (Cbr3), which both function as xenobiotic metabolizing enzymes. Notch gene homolog 1 (Notch1) and growth arrest specific 1 (Gas1), which are both associated with development and differentiation but serve undetermined roles in the liver, were also induced by TCDD (Table 4). Genes related to immune cell accumulation were also specific to the *in vivo* study, coincident with immune cell accumulation at 168 hr as determined by histopathological examination [18].

#### **COMPARISON OF BASAL GENE EXPRESSION LEVELS IN HEPA1C1C7 CELLS AND HEPATIC TISSUE**

In order to further investigate differences in gene expression levels, Hepa1c1c7 cells and C57BL/6 liver samples were directly compared by competitive hybridization on the same array,

**Table 3. Examples of TCDD-elicited gene expression responses unique to Hepa1c1c7 cells.**

| Accession | Gene name                                                 | Gene symbol | Entrez Gene ID | Fold change <sup>a</sup> | Time points <sup>b</sup> (hrs) |
|-----------|-----------------------------------------------------------|-------------|----------------|--------------------------|--------------------------------|
| AA111722  | cyclin D1                                                 | Ccnd1       | 12443          | -1.7                     | 4, 8, 12, 24 <sup>c</sup> , 48 |
| AA914666  | cyclin-dependent kinase inhibitor 2B (p15, inhibits CDK4) | Cdkn2b      | 12579          | 2.4                      | 4 <sup>c</sup> , 8, 48         |
| BC008247  | cyclin B2                                                 | Ccnb2       | 12442          | -1.8                     | 24 <sup>c</sup>                |
| BG064846  | cell division cycle 2 homolog A                           | Cdc2a       | 12534          | -1.5                     | 12, 24 <sup>c</sup>            |
| AA011839  | minichromosome maintenance deficient 2 mitotin            | Mcm2        | 17216          | -1.8                     | 8 <sup>c</sup> , 12            |
| BG074721  | minichromosome maintenance deficient 7                    | Mcm7        | 17220          | -1.7                     | 8, 12 <sup>c</sup> , 24        |
| AA003042  | myeloblastosis oncogene-like 2                            | Mybl2       | 17865          | -2.2                     | 8 <sup>c</sup> , 12, 24        |
| L27122    | UDP glucuronosyltransferase 1 family, polypeptide A2      | Ugt1a2      | 22236          | 2.8                      | 4, 8, 12 <sup>c</sup> , 24, 48 |

<sup>a</sup> Maximum absolute-fold change determined by microarray analysis

<sup>b</sup> Differentially regulated genes with P1(t) > 0.9999 and |fold change| > 1.5

<sup>c</sup> Time point representing the maximum |fold change|

**Table 4. Examples of TCDD-elicited gene expression responses unique to C57BL/6 hepatic tissue.**

| Accession | Gene name                                                   | Gene symbol | Entrez Gene ID | Fold change <sup>a</sup> | Time points <sup>b</sup> (hrs)             |
|-----------|-------------------------------------------------------------|-------------|----------------|--------------------------|--------------------------------------------|
| AA170585  | carbonic anhydrase 3                                        | Car3        | 12350          | -3.5                     | 12 <sup>c</sup> , 18, 24, 168              |
| AK003232  | carbonyl reductase 3                                        | Cbr3        | 109857         | 2.2                      | 12, 18 <sup>c</sup>                        |
| AA571998  | CD3 antigen, delta polypeptide                              | Cd3d        | 12500          | -2.4                     | 12, 18 <sup>c</sup> , 24, 72, 168          |
| BG072496  | ELOVL family member 5, elongation of long chain fatty acids | Elovl5      | 68801          | 2.0                      | 8, 12 <sup>c</sup> , 18, 24, 72, 168       |
| BG072453  | epoxide hydrolase 1, microsomal                             | Ephx1       | 13222          | 1.9                      | 8, 12, 18, 24 <sup>c</sup>                 |
| W84211    | growth arrest specific 1                                    | Gas1        | 14451          | -1.9                     | 4, 8, 18, 24, 72, 168 <sup>c</sup>         |
| W41175    | glycerol phosphate dehydrogenase 2, mitochondrial           | Gpd2        | 14571          | -2.3                     | 8, 12, 18, 24, 72 <sup>c</sup> , 168       |
| W29265    | glutathione S-transferase, alpha 2 (Yc2)                    | Gsta2       | 14858          | 7.2                      | 12, 18, 24, 72 <sup>c</sup> , 168          |
| AA145865  | lymphocyte antigen 6 complex, locus A                       | Ly6a        | 110454         | 2.5                      | 72, 168 <sup>c</sup>                       |
| W98998    | Notch gene homolog 1                                        | Notch1      | 18128          | 3.3                      | 2, 4 <sup>c</sup> , 8, 12, 18, 24, 72, 168 |

<sup>a</sup> Maximum absolute-fold change determined by microarray analysis

<sup>b</sup> Differentially regulated genes with P1(t) > 0.9999 and |fold change| > 1.5

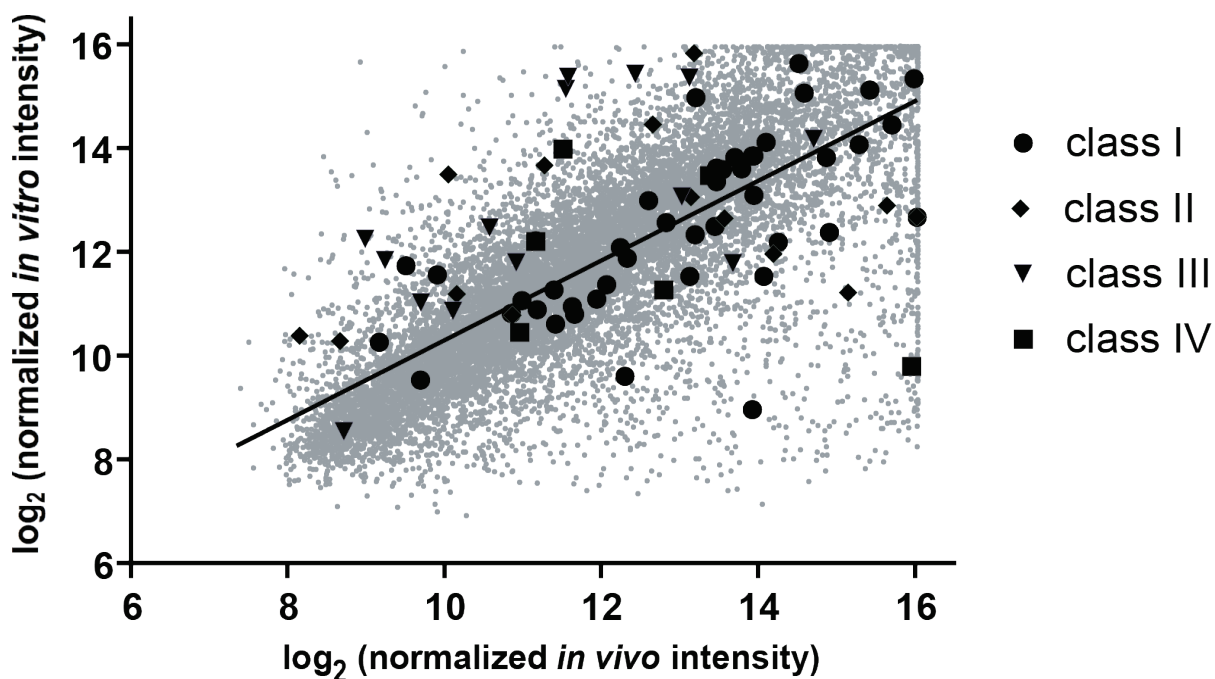
<sup>c</sup> Time point representing the maximum |fold change|

to identify basal gene expression level differences. Subsequent linear regression analysis of the mean normalized signal intensities from the untreated samples resulted in a correlation value of  $R = 0.75$  (Figure 6), which is consistent with basal gene expression comparisons of various *in vitro* rat hepatic systems against whole livers, where correlation values decreased between liver slices ( $R = 0.97$ ), primary cells ( $R = 0.85$ ), BRL3A ( $R = 0.3$ ), and NRL clone 9 ( $R = 0.32$ ) rat liver cell lines [10]. Overall, the correlation illustrates reasonable concordance in basal gene expression levels between the two models. However, data points that deviated from the fitted line indicate differences in the basal expression of individual genes between the Hepa1c1c7 cells and hepatic tissue from C57BL/6 mice. Although there are differences, they may be negligible if the TCDD-elicited responses are conserved *in vitro* and *in vivo*. Complete microarray data for the untreated comparisons are available in Additional file 7.

The relative basal expression of the 67 common active features was further investigated (Figure 6). In general, class I (i.e., induced in both models) genes fell close to the regression line, indicating that the basal expression levels of induced genes were comparable as were their *in vitro* and *in vivo* responses to TCDD. In contrast, basal expression levels of class III genes (i.e., induced *in vivo* while repressed *in vitro*) were generally higher in the Hepa1c1c7 cells, while levels in class II and IV (i.e., repressed in both models and repressed *in vivo* while induced *in vitro*, respectively) genes were scattered around the fitted linear line in Figure 6.

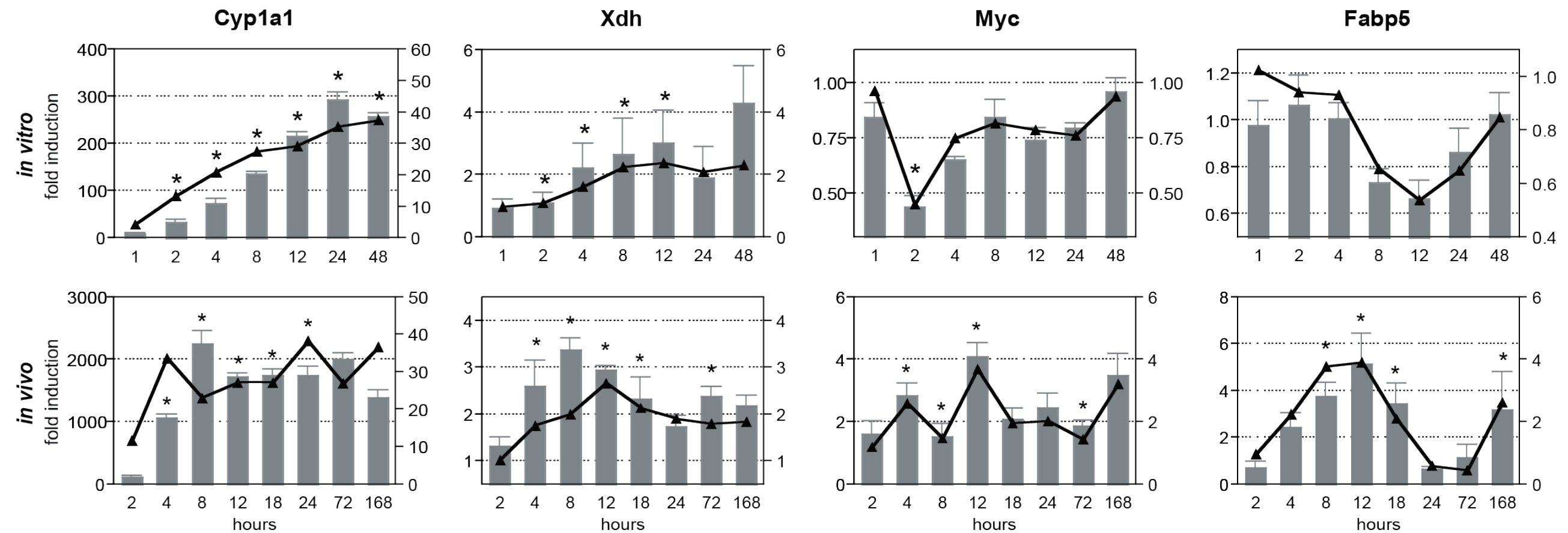
#### **QUANTITATIVE REAL-TIME PCR VERIFICATION OF MICROARRAY RESPONSES**

In total, 14 *in vitro* and 24 *in vivo* responsive genes representing common and model-specific genes were verified by quantitative real-time PCR (QRT-PCR) (see Additional file 8). Of the selected genes regulated in both models, all displayed temporal patterns comparable to the microarray data (Figure 7). For example, Xdh, Myc, and fatty acid binding protein (Fabp5)



**Figure 6. Comparison of Hepa1c1c7 cell and C57BL/6 hepatic tissue basal gene expression.** Untreated samples from Hepa1c1c7 cells and hepatic tissue from immature ovariectomized C57BL/6 mice taken at 0 hrs were competitively hybridized to the 13,362 feature cDNA microarray.  $\log_2$  normalized signal intensities were plotted for *in vitro* versus *in vivo* data to generate the correlation coefficient. The linear correlation coefficient R was 0.75 between *in vitro* and *in vivo* models.





**Figure 7. Quantitative real-time PCR verification of *in vitro* and *in vivo* microarray results.**

All fold changes were calculated relative to time-matched vehicle controls. Bars (left axis) and line (right axis) represent data obtained by QRT-PCR and cDNA microarrays, respectively. Genes are indicated by official gene symbols, and results are the average of four biological replicates. Classes refer to the respective classification categories as illustrated in Figure 5B. Error bars represent the standard error of measurement for the average fold change. \* $p < 0.05$  for QRT-PCR.

exhibited good agreement in fold change and temporal expression pattern when comparing microarray and QRT-PCR data. However, significant data compression were evident when comparing *in vitro* and *in vivo* Cyp1a1 induction by QRT-PCR, although *in vitro* and *in vivo* microarray induction levels were comparable. Previous studies suggest that this is likely due to the limited fluorescence intensity range (0 – 65,535) of microarrays resulting in signal saturation and compression of the true magnitude of induction of transcript levels [40, 41]. Cross-hybridization of homologous probes to a given target sequence on the microarray may also be a contributing factor, especially in comparison to other, more gene-specific measurement techniques [42].

## DISCUSSION

Microarrays have become an invaluable tool in toxicogenomics for comprehensively characterizing gene expression responses following treatment with an environmental contaminant, commercial chemical, natural product or drug, as well as for investigating complex mixtures relevant to human and wildlife exposures. An emerging consensus suggests that toxicogenomics will accelerate drug development and significantly improve quantitative risk assessments [43, 44]. In addition, toxicogenomics support the development and refinement of predictive *in vitro* high-throughput toxicity screening assays that can be used as alternatives to traditional *in vivo* testing. Ideally, *in vitro* high-throughput toxicity screens can be used to rank and prioritize drug candidates, environmental contaminants, and commercial chemicals, which warrant further development or testing. Although *in vitro* responses are assumed to reflect a subset of comparable *in vivo* responses, few studies have completed a comprehensive and systematic comparison. This study closely examined two well-established models, and

comprehensively compared the TCDD-elicited gene expression to assess the predictive value of *in vitro* systems.

Comparative analysis of Hepa1c1c7 cell and hepatic C57BL/6 microarray data identified 67 differentially expressed genes co-regulated by TCDD. Four classes based on their temporal expression patterns were identified (Figure 5B and 5C), with 42 of the 67 common regulated genes exhibiting dose-response characteristics in both models. *In vitro* EC<sub>50</sub> values ranged from 0.001182 nM to 2.4 nM, while *in vivo* the values ranged from 0.05 µg/kg to >150 µg/kg. The wide range of EC<sub>50</sub> values illustrate the varying sensitivity of regulated genes to TCDD in both models.

Hepa1c1c7 cells and hepatic tissue from C57BL/6 mice are the prototypical models used to investigate the mechanisms of action of TCDD and other related compounds, and both exhibited the classic induction of phase I and II metabolizing enzymes including Cyp1a1 and Nqo1 [45, 46]. Gsta4 and Xdh were also up-regulated in both models further demonstrating Hepa1c1c7 cells as a suitable model for investigating TCDD-regulated induction of xenobiotic metabolizing genes. In addition to these genes, the responses of Nqo1, Ugdh and Tnfaip2 were also conserved across models and were categorized as class I genes (similarly induced in both models; Figure 5B and 5C). However, Gsta2 was induced *in vivo*, while no significant effect was detected in Hepa1c1c7 cells, and Ugt1a2 was induced *in vitro*, but not differentially expressed in C57BL/6 hepatic tissue. Although many phase I and II metabolizing enzyme responses were conserved, differences exist that may limit Hepa1c1c7 cells from accurately modeling the full spectrum of *in vivo* hepatic responses elicited by TCDD.

A direct comparison of untreated Hepa1c1c7 cells and C57BL/6 hepatic tissue was performed to further investigate innate differences between the two models. Comparison of the normalized signal intensities revealed a good correlation ( $R = 0.75$ ) between *in vitro* and *in vivo* basal expression levels (Figure 6). This illustrates that many genes are basally expressed to similar levels in both models as illustrated by the cluster of class I (similarly induced genes) closely surrounding the fitted line. Although a correlation exists, there are still differences in basal expression which may be associated with the origins of the models (i.e., normal hepatic tissue versus hepatoma derived Hepa1c1c7 cells), as well as the inability of *in vitro* systems to effectively model complex interactions between different cell types (e.g., Kupffer and stellate cells). For example, *Myc*, a  $G_1$  to S phase cell cycle regulator [30, 33-36], was repressed *in vitro* while being induced *in vivo*, and the model-specific responses may be related to difference in basal expression levels between the two models (Table 2). The levels of *Myc* transcripts in untreated Hepa1c1c7 cells were higher relative to untreated C57BL/6 hepatic tissue, consistent with the proliferative state of the *in vitro* system (data not shown). Examination of other class III genes suggests that they are more highly expressed *in vitro* when compared to *in vivo* (Figure 6). Consequently, differences in basal expression may be a factor contributing to divergent *in vitro* – *in vivo* responses. Another possible source for the model-specific responses may be related to DNA methylation status of the promoter region of TCDD-responsive genes in either model. DNA methylation results in gene silencing [47, 48] and a previous study with Hepa1c1c7 has shown that TCDD-elicited gene expression responses are influenced by DNA methylation status [49]. The differing methylation states between the *in vitro* and *in vivo* systems may further contribute to the model-specific gene expression responses.

Many *in vitro* specific gene expression responses elicited by TCDD were associated with

cell cycle progression and cell cycle arrest. Myc and its downstream target, cyclin D1, which form a kinase complex with Cdk4 [50], were both repressed by TCDD. In contrast, Cdkn1a, an inhibitor of cyclin-dependent kinase 2 (Cdk2)-cyclin E complex kinase activity [51], was induced. Inactivation of the Cdk2-cyclin E complex prevents the phosphorylation of pRb resulting in cell cycle arrest during G<sub>1</sub>. Additionally, the *in vitro* induction of Btg2 suggests an alternative mechanism for cell cycle arrest during the G<sub>2</sub> phase. Constitutively active BTG2 in human leukemia U937 cells induces G<sub>2</sub>/M cell cycle arrest by inhibiting the formation of the cyclin B1 and Cdc2 complex, thereby inhibiting the active kinase function of the complex [37]. Collectively, these results corroborate and extend previous *in vitro* TCDD-mediated cell cycle arrest studies [52-55].

TCDD treatment resulted in a number of divergent gene responses across both models as represented by classes III and IV (Figures 5B and 5C). Genes related to immune cell accumulation, including major histocompatibility complex (MHC) molecules were only observed *in vivo*, and are likely a response to hepatic damage mediated by ROS or fatty accumulation, and therefore independent of direct AhR action [18]. This is characteristic of the complex interaction between different cell types responding to liver injury that cannot be modeled in homogenous cultures of cells.

Pharmacokinetics may also contribute to response differences between the two models. Hepa1c1c7 cells were directly treated, whereas *in vivo*, TCDD must first be delivered to the liver and targeted cells prior to eliciting its effects. Additionally, C57BL/6 studies were able to be carried out to 168 hrs following TCDD treatment, while *in vitro* studies were limited to 48 hrs to minimize potentially confounding effects due to cell confluency. However, early responses associated with classes I and II (induced or repressed in both models; Figure 5B and 5C) are well

conserved and exhibit comparable levels of induction or repression in both models. Hierarchical clustering of the common active genes (Figure 5C) illustrates gene induction occurs early while gene repression occurs later in both models. Clustering across both time and model revealed that gene expression profiles at 1 hr *in vitro* and 2 hr *in vivo* were most similar. This clustering pattern implies that early *in vitro* responses may accurately model early *in vivo* gene expression effects.

Comparative analysis of global gene expression from Hepa1c1c7 cells and hepatic tissue from C57BL/6 mice identified several model-specific responses to TCDD that should be considered when extrapolating *in vitro* results to potential *in vivo* effects. Despite these differences, immortalized cells as well as other emerging *in vitro* systems (e.g., primary cells, stem cells and 3-D culture systems) provide valuable mechanistic information that support the further development of high-throughput toxicity screening assays. However, the relevance of *in vitro* responses requires complementary *in vivo* verification. Furthermore, comparative studies exploiting other *in vitro* and *in vivo* systems, different structurally diverse ligands, and other relevant model species will not only corroborate the relevance of the mechanisms, but will also support more appropriate extrapolations between rodent studies and potential effects in humans and ecologically-relevant species.

## REFERENCES

## REFERENCES

1. Battershill J: **Toxicogenomics: regulatory perspective on current position.** *Hum Exp Toxicol* 2005, **24**:35-40.
2. Bishop W, Clarke D, Travis C: **The genomic revolution: what does it mean for risk assessment?** *Risk Anal* 2001, **21**:983-987.
3. Olden K, Wilson S: **Environmental health and genomics: visions and implications.** *Nat Rev Genet* 2000, **1**:149-153.
4. Suk W, Olden K, Yang R: **Chemical mixtures research: significance and future perspectives.** *Environ Health Perspect* 2002, **110 Suppl 6**:891-892.
5. Tennant R: **The National Center for Toxicogenomics: using new technologies to inform mechanistic toxicology.** *Environ Health Perspect* 2002, **110**:A8-10.
6. Boverhof DR, Zacharewski TR: **Toxicogenomics in risk assessment: applications and needs.** *Toxicol Sci* 2006, **89**:352-360.
7. Lord PG: **Progress in applying genomics in drug development.** *Toxicol Lett* 2004, **149**:371-375.
8. Yang Y, Blomme EAG, Waring JF: **Toxicogenomics in drug discovery: from preclinical studies to clinical trials.** *Chemico-Biological Interactions* 2004, **150**:71-85.
9. Abbott A: **Animal testing: more than a cosmetic change.** *Nature* 2005, **438**:144-146.
10. Boess F, Kamber M, Romer S, Gasser R, Muller D, Albertini S, Suter L: **Gene expression in two hepatic cell lines, cultured primary hepatocytes, and liver slices compared to the in vivo liver gene expression in rats: possible implications for toxicogenomics use of in vitro systems.** *Toxicol Sci* 2003, **73**:386-402.
11. Burczynski M, McMillian M, Ciervo J, Li L, Parker J, Dunn R, Hicken S, Farr S, Johnson M: **Toxicogenomics-based discrimination of toxic mechanism in HepG2 human hepatoma cells.** *Toxicol Sci* 2000, **58**:399-415.
12. Harris A, Dial S, Casciano D: **Comparison of basal gene expression profiles and effects of hepatocarcinogens on gene expression in cultured primary human hepatocytes and HepG2 cells.** *Mutat Res* 2004, **549**:79-99.
13. Hartung T, Gribaldo L: **New hepatocytes for toxicology?** *Trends Biotechnol* 2004, **22**:613-615; discussion 615-616.



14. Luhe A, Hildebrand H, Bach U, Dingermann T, Ahr H: **A new approach to studying ochratoxin A (OTA)-induced nephrotoxicity: expression profiling in vivo and in vitro employing cDNA microarrays.** *Toxicol Sci* 2003, **73**:315-328.
15. Poland A, Knutson JC: **2,3,7,8-tetrachlorodibenzo-p-dioxin and related halogenated aromatic hydrocarbons: examination of the mechanism of toxicity.** *Annu Rev Pharmacol Toxicol* 1982, **22**:517-554.
16. Denison MS, Heath-Pagliuso S: **The Ah receptor: a regulator of the biochemical and toxicological actions of structurally diverse chemicals.** *Bulletin of environmental contamination and toxicology* 1998, **61**:557-568.
17. Hankinson O: **The aryl hydrocarbon receptor complex.** *Annu Rev Pharmacol Toxicol* 1995, **35**:307-340.
18. Boverhof DR, Burgoon LD, Tashiro C, Chittim B, Harkema JR, Jump DB, Zacharewski TR: **Temporal and dose-dependent hepatic gene expression patterns in mice provide new insights into TCDD-Mediated hepatotoxicity.** *Toxicol Sci* 2005, **85**:1048-1063.
19. Yang YH, Speed T: **Design issues for cDNA microarray experiments.** *Nat Rev Genet* 2002, **3**:579-588.
20. Burgoon L, Eckel-Passow J, Gennings C, Boverhof D, Burt J, Fong C, Zacharewski T: **Protocols for the assurance of microarray data quality and process control.** *Nucleic Acids Res* 2005, **33**:e172.
21. Eckel J, Gennings C, Therneau T, Burgoon L, Boverhof D, Zacharewski T: **Normalization of two-channel microarray experiments: a semiparametric approach.** *Bioinformatics* 2005, **21**:1078-1083.
22. Eckel J, Gennings C, Chinchilli V, Burgoon L, Zacharewski T: **Empirical bayes gene screening tool for time-course or dose-response microarray data.** *J Biopharm Stat* 2004, **14**:647-670.
23. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al: **Gene ontology: tool for the unification of biology. The Gene Ontology Consortium.** *Nat Genet* 2000, **25**:25-29.
24. Zeeberg B, Feng W, Wang G, Wang M, Fojo A, Sunshine M, Narasimhan S, Kane D, Reinhold W, Lababidi S, et al: **GoMiner: a resource for biological interpretation of genomic and proteomic data.** *Genome Biol* 2003, **4**:R28.
25. Rozen S, Skaletsky H: **Primer3 on the WWW for general users and for biologist programmers.** *Methods Mol Biol* 2000, **132**:365-386.
26. Seidel S, Denison M: **Differential gene expression in wild-type and arnt-defective mouse hepatoma (Hepa1c1c7) cells.** *Toxicol Sci* 1999, **52**:217-225.

27. Tijet N, Boutros PC, Moffat ID, Okey AB, Tuomisto J, Pohjanvirta R: **Aryl hydrocarbon receptor regulates distinct dioxin-dependent and dioxin-independent gene batteries.** *Mol Pharmacol* 2006, **69**:140-153.
28. Dang CV: **c-Myc target genes involved in cell growth, apoptosis, and metabolism.** *Molecular and Cellular Biology* 1999, **19**:1-11.
29. Felsher D, Zetterberg A, Zhu J, Tlsty T, Bishop J: **Overexpression of MYC causes p53-dependent G2 arrest of normal fibroblasts.** *Proc Natl Acad Sci U S A* 2000, **97**:10544-10548.
30. Nilsson J, Keller U, Baudino T, Yang C, Norton S, Old J, Nilsson L, Neale G, Kramer D, Porter C, Cleveland J: **Targeting ornithine decarboxylase in Myc-induced lymphomagenesis prevents tumor formation.** *Cancer Cell* 2005, **7**:433-444.
31. Wolfle D, Marotzki S, Dartsch D, Schafer W, Marquardt H: **Induction of cyclooxygenase expression and enhancement of malignant cell transformation by 2,3,7,8-tetrachlorodibenzo-p-dioxin.** *Carcinogenesis* 2000, **21**:15-21.
32. Nebert D, Petersen D, Fornace A: **Cellular responses to oxidative stress: the [Ah] gene battery as a paradigm.** *Environ Health Perspect* 1990, **88**:13-25.
33. Gartel A, Shchors K: **Mechanisms of c-myc-mediated transcriptional repression of growth arrest genes.** *Exp Cell Res* 2003, **283**:17-21.
34. Gartel A, Radhakrishnan S: **Lost in transcription: p21 repression, mechanisms, and consequences.** *Cancer Res* 2005, **65**:3980-3985.
35. Lee T, Li L, Philipson L, Ziff E: **Myc represses transcription of the growth arrest gene gas1.** *Proc Natl Acad Sci U S A* 1997, **94**:12886-12891.
36. Packham G, Cleveland J: **Induction of ornithine decarboxylase by IL-3 is mediated by sequential c-Myc-independent and c-Myc-dependent pathways.** *Oncogene* 1997, **15**:1219-1232.
37. Ryu M, Lee M, Hong J, Hahn T, Moon E, Lim I: **TIS21/BTG2/PC3 is expressed through PKC-delta pathway and inhibits binding of cyclin B1-Cdc2 and its activity, independent of p53 expression.** *Exp Cell Res* 2004, **299**:159-170.
38. Barnes-Ellerbe S, Knudsen K, Puga A: **2,3,7,8-Tetrachlorodibenzo-p-dioxin blocks androgen-dependent cell proliferation of LNCaP cells through modulation of pRB phosphorylation.** *Mol Pharmacol* 2004, **66**:502-511.
39. Wang W, Smith R, Safe S: **Aryl hydrocarbon receptor-mediated antiestrogenicity in MCF-7 cells: modulation of hormone-induced cell cycle enzymes.** *Arch Biochem Biophys* 1998, **356**:239-248.

40. Dodd L, Korn E, McShane L, Chandramouli G, Chuang E: **Correcting log ratios for signal saturation in cDNA microarrays.** *Bioinformatics* 2004, **20**:2685-2693.
41. Hsiao L, Jensen R, Yoshida T, Clark K, Blumenstock J, Gullans S: **Correcting for signal saturation errors in the analysis of microarray data.** *Biotechniques* 2002, **32**:330-332, 334, 336.
42. Yuen T, Wurmbach E, Pfeffer R, Ebersole B, Sealfon S: **Accuracy and calibration of commercial oligonucleotide and custom cDNA microarrays.** *Nucleic Acids Res* 2002, **30**:e48.
43. Luhe A, Suter L, Ruepp S, Singer T, Weiser T, Albertini S: **Toxicogenomics in the pharmaceutical industry: hollow promises or real benefit?** *Mutat Res* 2005, **575**:102-115.
44. Pennie W, Tugwood J, Oliver G, Kimber I: **The principles and practice of toxigenomics: applications and opportunities.** *Toxicol Sci* 2000, **54**:277-283.
45. Bernhard H, Darlington G, Ruddle F: **Expression of liver phenotypes in cultured mouse hepatoma cells: synthesis and secretion of serum albumin.** *Dev Biol* 1973, **35**:83-96.
46. Darlington G, Bernhard H, Miller R, Ruddle F: **Expression of liver phenotypes in cultured mouse hepatoma cells.** *J Natl Cancer Inst* 1980, **64**:809-819.
47. Clark S, Melki J: **DNA methylation and gene silencing in cancer: which is the guilty party?** *Oncogene* 2002, **21**:5380-5387.
48. Ng H, Bird A: **DNA methylation and chromatin modification.** *Curr Opin Genet Dev* 1999, **9**:158-163.
49. Jin B, Kim G, Park DW, Ryu D-Y: **Microarray analysis of gene regulation in the Hepa1c1c7 cell line following exposure to the DNA methylation inhibitor 5-aza-2'-deoxycytidine and 2,3,7,8-tetrachlorodibenzo-p-dioxin.** *Toxicol In Vitro* 2004, **18**:659-664.
50. Adhikary S, Eilers M: **Transcriptional regulation and transformation by Myc proteins.** *Nat Rev Mol Cell Biol* 2005, **6**:635-645.
51. Sherr C, Roberts J: **CDK inhibitors: positive and negative regulators of G1-phase progression.** *Genes Dev* 1999, **13**:1501-1512.
52. Ge N, Elferink C: **A direct interaction between the aryl hydrocarbon receptor and retinoblastoma protein. Linking dioxin signaling to the cell cycle.** *J Biol Chem* 1998, **273**:22708-22713.
53. Huang G, Elferink C: **Multiple mechanisms are involved in Ah receptor-mediated cell cycle arrest.** *Mol Pharmacol* 2005, **67**:88-96.

54. Levine-Fridman A, Chen L, Elferink C: **Cytochrome P4501A1 promotes G1 phase cell cycle progression by controlling aryl hydrocarbon receptor activity.** *Mol Pharmacol* 2004, **65**:461-469.
55. Marlowe J, Knudsen E, Schwemberger S, Puga A: **The aryl hydrocarbon receptor displaces p300 from E2F-dependent promoters and represses S phase-specific gene expression.** *J Biol Chem* 2004, **279**:29013-29022.

## CHAPTER 4

Dere E, Lee AW, Burgoon LD, Zacharewski TR: **Comparison of TCDD-Elicited Gene Expression in Human HepG2, Mouse Hepa1c1c7 and Rat H4IIE Hepatoma Cells.** *BMC Genomics* (submitted July 19, 2010)

## CHAPTER 4

# COMPARISON OF TCDD-ELICITED GENE EXPRESSION IN HUMAN HEPG2, MOUSE HEPA1C1C7 AND RAT H4IIE HEPATOMA CELLS

### ABSTRACT

2,3,7,8-Tetrachlorodibenzo-*p*-dioxin (TCDD) is an environmental contaminant that elicits a broad spectrum of toxic effects in a species-specific manner. Current risk assessment practices routinely extrapolate results from *in vivo* and *in vitro* rodent models to assess human risk. In order to further investigate the species-specific responses elicited by TCDD, temporal gene expression responses in human HepG2, mouse Hepa1c1c7, and rat H4IIE cells were compared. Microarray analysis identified a core set of conserved gene expression responses across species consistent with the role of AhR in mediating adaptive metabolic responses. However, significant species-specific, as well as species-divergent responses were identified. Computational analyses of the regulatory regions of species-specific and -divergent responses suggest that dioxin response elements (DREs) are involved. These results are consistent with *in vivo* rat-mouse comparative studies, and more comprehensive comparative DRE searches. Comparative analysis of responses is consistent with *in vivo* rat-mouse comparative gene expression studies, and more comprehensive comparative DRE searches, suggesting that AhR-mediated gene expression is species-specific.

### INTRODUCTION

2,3,7,8-Tetrachlorodibenzo-*p*-dioxin (TCDD) is a ubiquitous environmental contaminant that elicits a broad spectrum of biochemical and physiological effects in a species-specific

manner [1]. These effects include lethality, cancer, developmental abnormalities, immunotoxicity, skin lesions, hepatotoxicity, and xenobiotic enzyme metabolism induction. They result from altered gene expression mediated by the aryl hydrocarbon receptor (AhR), a ligand-activated transcription factor [1, 2]. Briefly, TCDD binds to the cytoplasmic AhR causing its nuclear translocation and heterodimerization with the AhR nuclear translocator (ARNT). The activated heterodimer then binds to specific DNA elements, termed dioxin response elements (DREs), within the regulatory regions of targeted genes to modulate expression, resulting in the downstream physiological responses [3]. Although the structure and function of the AhR are highly conserved [4], the sensitivity to and the responses elicited by TCDD vary widely across species, suggesting that TCDD and related compounds may elicit species-specific AhR-mediated gene expression networks.

Risk assessment assumes that there is a conserved mode of action and comparable toxic response between species. However, there are inherent differences between species that may compromise the extrapolation of rodent data to estimate potential human risks. Moreover, there is a discord between results from preclinical animal testing and human clinical trials regarding toxicity [5]. Similarly, species differ widely in response to TCDD exposure. For example, the LD<sub>50</sub> values range from 1 µg/kg in the guinea pig [6] to > 1000 µg/kg in the hamster, and there are other responses that also exhibit species-specific sensitivities and toxicities. These species-specific responses and sensitivities are not attributed to differences in binding affinity or AhR complex stability [7-9]. Collectively, these data suggest that although the AhR is conserved between species, subsequent differential gene expression effects are species-specific.

To further investigate the differences in TCDD-elicited responses following AhR binding, global gene expression profiles were assessed in human HepG2, mouse Hepa1c1c7, and

rat H4IIE hepatoma cells, following treatment with TCDD. Comparative analysis indicates that TCDD elicits species-specific gene expression profiles, suggesting that AhR-mediated gene expression may not be conserved.

## **MATERIALS AND METHODS**

### ***CELL CULTURE AND TREATMENT***

HepG2 (Dr. Trevor Archer, NIEHS, Research Triangle Park, NC), Hepa1c1c7 (Dr. Oliver Hankinson, University of California, Los Angeles, CA), and H4IIE (Dr. Niels Bols, University of Waterloo, Waterloo, Canada) cells were cultured and treated with TCDD (S. Safe, Texas A&M University, College Station, TX) as previously described [10]. Briefly, cells were treated with either 10 nM TCDD or DMSO vehicle control for 1, 2, 4, 8, 12, 24 or 48 hrs for the time-course studies. For co-treatment studies, cells were pretreated with 10 mg/ml cycloheximide (CHX; Sigma) for 1 hr and then treated for an additional 4 or 12 hrs with 10 nM TCDD or DMSO vehicle (Supplementary Figure 1).

### ***RNA ISOLATION***

Cells were harvested with 2.0 mL of TRIzol® Reagent (Invitrogen) and total RNA isolated according to the manufacturer's protocol followed by an acid phenol:chloroform extraction. Isolated RNA was resuspended in the RNA Storage Solution (Ambion Inc., Austin, TX), quantified ( $A_{260}$ ), and assessed for purity by measuring the  $A_{260}/A_{280}$  ratio and by visual inspection of 1.0 µg on a denaturing gel.

### ***MICROARRAY EXPERIMENTAL DESIGN AND ANALYSIS***

Gene expression changes were analyzed using custom human, mouse and rat cDNA microarrays as previously described [10-12]. Responses to CHX and TCDD co-treatment were



also assayed with cDNA microarrays using a 2×2 factorial design (Supplemental Figure 2) [13]. Additionally, 4×44k whole genome oligonucleotide microarrays from Agilent Technologies (Santa Clara, CA) were used to profile the responses elicited by TCDD 24 hrs post-treatment, according to the manufacturer's Two-Color Microarray-Based Gene Expression Analysis protocol Version 5.0.1, including dye swap labelings. Each microarray was performed in triplicate. Microarray data were normalized using a semiparametric method [14], and statistically analyzed using an empirical Bayes methods [15]. Functional annotation clustering of Gene Ontology (GO) terms for differentially expressed genes was performed using DAVID (Database for Annotation, Visualization, and Integrated Discovery) [16]. Annotation clusters with an enrichment score  $\geq 1.3$  were considered significantly enriched.

### **QUANTITATIVE REAL-TIME PCR**

The same total RNA samples isolated for microarray studies were used for QRT-PCR as previously described [10]. The copy number of each unknown sample was standardized to the geometric mean of three house-keeping genes ( $\beta$ -actin, Gapd, Hprt or Rpl13). Official gene names and symbols, RefSeq and Entrez Gene IDs, forward and reverse primer sequences, and amplicon sizes are provided in Supplementary Table 1. Data were analyzed by analysis of variance (ANOVA) followed by Tukey's *post hoc* test using SAS 9.1 (SAS Institute, Cary, NC). Differences between treatment groups were considered significant when  $p < 0.05$ .

## **RESULTS**

### **TEMPORALLY CONSERVED GENE EXPRESSION RESPONSES ELICITED BY TCDD**

Species-specific cDNA microarrays were used to profile the temporal gene expression elicited by TCDD in human HepG2, mouse Hepa1c1c7, and rat H4IIE cells. The microarrays queried 6995, 8478 and 5169 unique human, mouse and rat genes, respectively (Table 5).

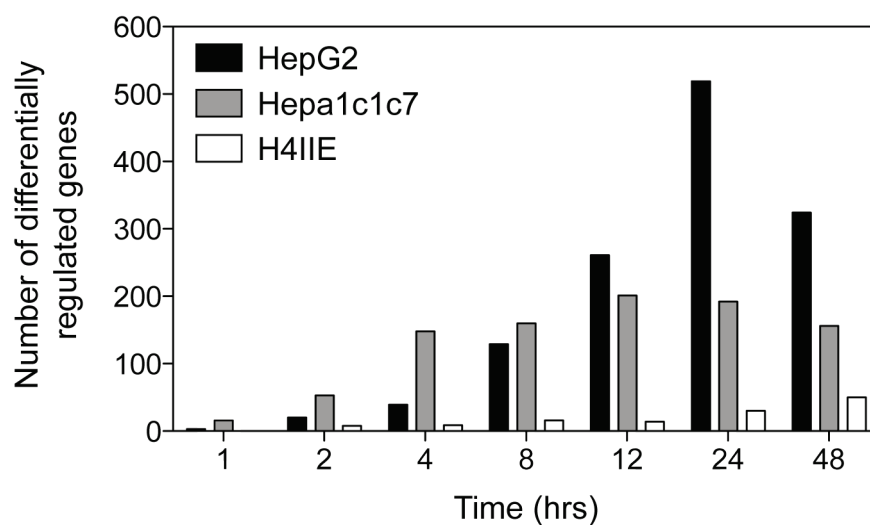
**Table 5. Gene coverage of species-specific cDNA microarray platforms and number of differentially regulated genes.**

|                           | Human HepG2 |                                       | Mouse Hepa1c1c7 |                                       | Rat H4IIE |                                       |
|---------------------------|-------------|---------------------------------------|-----------------|---------------------------------------|-----------|---------------------------------------|
|                           | Total       | Differentially Regulated <sup>a</sup> | Total           | Differentially Regulated <sup>a</sup> | Total     | Differentially Regulated <sup>a</sup> |
| Unique Genes <sup>b</sup> | 6,995       | 624                                   | 8,478           | 438                                   | 5,169     | 56                                    |
| Orthologs <sup>c</sup>    | 6,825       | 616                                   | 8,233           | 432                                   | 4,871     | 52                                    |

<sup>a</sup> |fold change| > 1.4 and P1(t) > 0.999

<sup>b</sup> based on Entrez GeneID

<sup>c</sup> based on HomoloGeneID

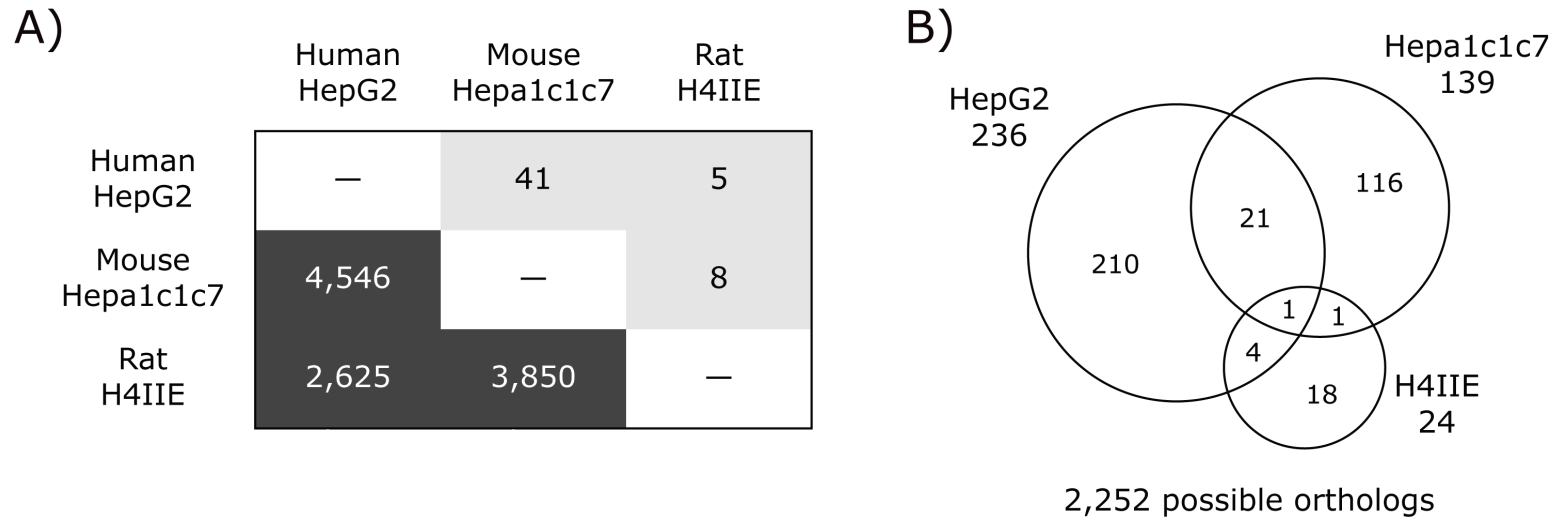


**Figure 8. Number of TCDD-elicited differentially expressed genes in human HepG2, mouse Hepa1c1c7, and rat H4IIE.**

Global gene expression changes were detected using cDNA microarray analysis of cells treated with 10 nM TCDD for 1, 2, 4, 8, 12, 24 and 48 hrs. Differentially expressed genes are defined as having  $P1(t) > 0.999$  and  $|\text{fold change}| > 1.4$ -fold at one or more time points.

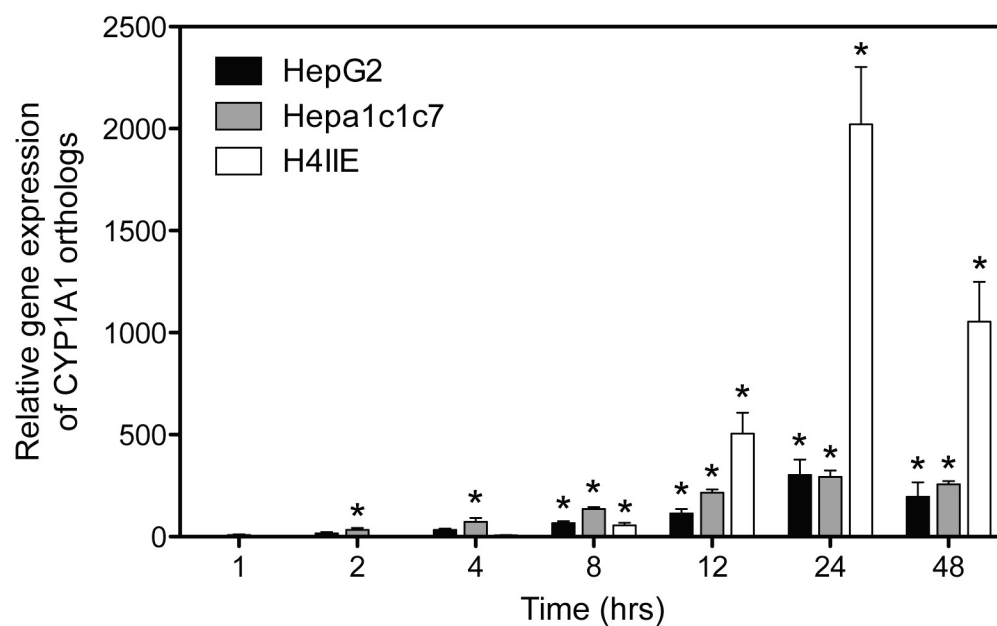
Empirical Bayes analysis identified 691, 439 and 57 differentially expressed genes ( $P(t) > 0.999$  and  $|\text{fold change}| > 1.4$ ) in HepG2, Hepa1c1c7, and H4IIE cells, respectively. HepG2 cells were the most responsive as indicated by both the overall number of differentially expressed genes as well as the number of responsive genes at each time point (Figure 8). H4IIE cells exhibited significantly less differentially expressed genes, partially explained by the smaller microarray and the immaturity of rat genome annotation compared to either the human or mouse. Complete cDNA microarray data sets are provided in Supplementary Tables 2-4.

Pair-wise comparisons of differentially regulated genes were conducted using HomoloGene (build 35) defined orthologs (Figure 9A). Human and mouse cDNA microarrays shared 4546 orthologous genes, however only 0.9% (41 orthologs; Supplementary Table 5) were differentially regulated by TCDD in HepG2 and Hepa1c1c7 cells. Comparison of the rodent platforms identified 3850 orthologs with only 0.2% (8 orthologs; Supplementary Table 6) responding in both Hepa1c1c7 and H4IIE cells. The lack of conserved ortholog expression in Hepa1c1c7 and H4IIE cells is consistent with differences in differential expression observed *in vivo* between mice and rats [17-19]. Time-dependent profiling of HepG2 and H4IIE gene expression identified only 5 conserved responses out of 2625 possible orthologs, representing only 0.2% (Supplementary Table 7). Comparative analysis across all three species with 2252 shared orthologous probes identified only one ortholog that was differentially regulated in all three cell lines (immediate early response 3, IER3; HomoloGene ID 2894). Note that other members of the AhR gene battery, namely CYP1A1, ALDH3A1 and NQO1, were not present across all of the cDNA microarray platforms. However, their responses were conserved across all three cell lines when measured using QRT-PCR. The results for CYP1A1 are shown in Figure 10.



**Figure 9. Cross-species comparison of TCDD-elicited temporal gene expression responses using cDNA microarrays.**

(A) The lower half (dark boxes) of the matrix show the number of orthologs represented between pairs of species-specific cDNA microarray platforms. The upper half (light grey boxes) of the matrix provides the number of TCDD elicited differentially expressed orthologs between pairs of cell lines. (B) The Venn diagram shows the overlap of all differentially expressed genes ( $P(t) > 0.999$  and  $|\text{fold change}| > 1.4$ ) from 2,252 possible orthologs represent across all three microarray platforms. Supplementary Tables 2, 3 and 4 lists the shared responses between pairs of species.



**Figure 10. QRT-PCR verification of the conserved induction of CYP1A1 across the human HepG2, mouse Hepa1c1c7, and rat H4IIE cell lines.**

The same RNA used for microarray analysis was examined by QRT-PCR. Fold changes were calculated relative to time-matched vehicle controls. Error bars represent the standard error of measurement for the average fold change and the asterisk (\*) indicates  $p < 0.05$ .

### ***IDENTIFICATION OF PUTATIVE PRIMARY GENE EXPRESSION RESPONSES***

In order to further investigate AhR-mediated responses, TCDD-elicited differential gene expression was examined in the presence of cycloheximide (CHX), a protein translation inhibitor. Putative primary responses were defined in this study as those where CHX co-treatment either maintained or enhanced the response elicited by TCDD, while responses that were attenuated or blocked by CHX co-treatment were classified as secondary responses based on their assumed dependence on additional protein translation. cDNA microarray analysis confirmed the superinduction of CYP1A1 mRNA in CHX co-treated Hepa1c1c7 cells [20, 21], consistent with the superinduction in human MCF10A cells treated with TCDD [22]. Additionally, Hepa1c1c7 ARNT-deficient c4 mutants treated with TCDD did not exhibit induction of the prototypical AhR battery genes including CYP1A1 [10]. Collectively, these results indicate that CYP1A1 is a primary gene expression response consistent with the direct interaction of the AhR with DREs within the promoter region.

For each species, differentially expressed orthologs were classified as primary or secondary responses based on CHX co-treatment studies at both 4 and 12 hrs. Overall 61, 38 and 2 human, mouse and rat orthologs, respectively, were considered primary responses (Supplementary Tables 8-10). Furthermore, 45, 12 and 10 orthologous genes in the HepG2, Hepa1c1c7 and H4IIE cells were classified as secondary AhR responses (Supplementary Tables 11-13). Comparative examination of the CHX co-treatment data suggested that each cell line had its own unique set of primary responsive orthologs.

### ***WHOLE-GENOME ANALYSIS OF CONSERVED TCDD-ELICITED GENE EXPRESSION RESPONSES***

The lack of whole genome coverage on the human, mouse and rat cDNA microarrays limited the number of orthologs that could be investigated. Therefore, whole genome expression

**Table 6. Gene coverage of species-specific Agilent microarray platforms and number of differentially regulated genes.**

|                           | Human HepG2 |                                       | Mouse Hepa1c1c7 |                                       | Rat H4IIE |                                       |
|---------------------------|-------------|---------------------------------------|-----------------|---------------------------------------|-----------|---------------------------------------|
|                           | Total       | Differentially Regulated <sup>a</sup> | Total           | Differentially Regulated <sup>a</sup> | Total     | Differentially Regulated <sup>a</sup> |
| Unique Genes <sup>b</sup> | 18,499      | 865                                   | 20,929          | 508                                   | 18,244    | 129                                   |
| Orthologs <sup>c</sup>    | 16,781      | 828                                   | 17,543          | 477                                   | 15,705    | 116                                   |

<sup>a</sup> |fold change| > 1.4 and P1(t) > 0.999

<sup>b</sup> based on Entrez GeneID

<sup>c</sup> based on HomoloGeneID



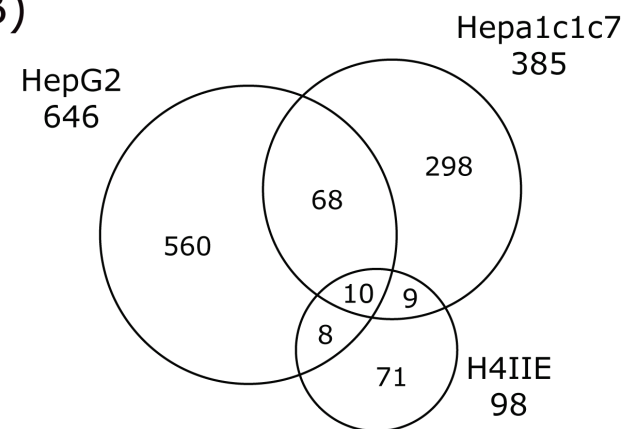
analysis was performed at 24 hrs, one of the most active time points in terms of the number of differentially expressed genes (Figure 8), using the 4×44k Agilent oligonucleotide microarrays. Each microarray contained more than 41,000 individual probes, representing more than 18,000 unique genes (Table 6). Despite the increased coverage, the number of TCDD elicited differentially expressed genes were surprisingly small relative to the cDNA microarray results. For example, the human Agilent microarray consisted of 18,499 known genes, representing a 2.6-fold increase in coverage compared to the human cDNA microarray. However, only 865 unique genes were differentially expressed, a modest increase from the 691 genes identified using cDNA microarrays. Similarly, only 508 and 128 genes were responsive in the Hepa1c1c7 and H4IIE cells, respectively. Complete Agilent microarray data sets are provided in Supplementary Tables 14-16.

The use of whole genome microarrays also increased the number of orthologs that could be examined (Figure 11A). As seen with the cDNA microarray dataset pair-wise comparisons, HepG2 and Hepa1c1c7 cells shared the greatest number of TCDD responsive orthologs (Supplementary Tables 17-19). Ortholog coverage between all three species increased from 2252 on the cDNA microarrays to 12,388 across the Agilent platforms. Comparative analysis ( $P(t) > 0.999$  and  $|\text{fold change}| > 1.4$ ) identified only 10 orthologs that were differentially expressed by TCDD at 24 hrs (Figure 11B; Table 7). Whole genome expression profiling identified the species-conserved induction of CYP1A1, TIPARP and UGT1A6 in all three cell lines. Despite this increased coverage, the number of differentially expressed orthologs across all three cell lines remained small, consistent with our cDNA microarray results.

A)

|                    | Human<br>HepG2 | Mouse<br>Hepa1c1c7 | Rat<br>H4IIE |
|--------------------|----------------|--------------------|--------------|
| Human<br>HepG2     | —              | 86                 | 21           |
| Mouse<br>Hepa1c1c7 | 14,724         | —                  | 19           |
| Rat<br>H4IIE       | 12,999         | 14,227             | —            |

B)



12,388 possible orthologs

**Figure 11. Cross-species comparison of TCDD-elicited gene expression responses at 24 hrs using 44×4K Agilent microarrays.**

(A) The lower half (dark boxes) of the matrix show the number of orthologous genes represented between pairs of species-specific microarrays. The upper half (light grey boxes) provides the number of TCDD-elicited differentially expressed orthologs between pairs of cell lines. (B) The 3-way Venn diagram shows the overlap of all differentially expressed genes ( $P(t) > 0.999$  and  $|\text{fold change}| > 1.4$ ) from 12,388 possible orthologous genes represented across all three platforms. The 10 genes differentially expressed in all three cell lines are listed in Table 7. Supplementary Table 6, 7 and 8 lists the shared responses between pairs of species.

**Table 7. List of common genes identified as differentially expressed by TCDD treatment from whole genome Agilent microarray analysis.**

| Gene Symbol | Gene Name                                                              | Homologene ID | Fold change <sup>a</sup> |   |        |   |        |   |
|-------------|------------------------------------------------------------------------|---------------|--------------------------|---|--------|---|--------|---|
|             |                                                                        |               | Human                    |   | Mouse  |   | Rat    |   |
| NQO1        | NAD(P)H dehydrogenase, quinone 1                                       | 695           | 3.22                     | ▲ | 4.88   | ▲ | 4.82   | ▲ |
| CCND1       | cyclin D1                                                              | 1334          | 1.57                     | ▲ | -1.56  | ▼ | -1.47  | ▼ |
| ID3         | inhibitor of DNA binding 3, dominant negative helix-loop-helix protein | 1633          | 1.74                     | ▲ | -1.83  | ▼ | -1.68  | ▼ |
| TIPARP      | TCDD-inducible poly(ADP-ribose) polymerase                             | 9167          | 1.47                     | ▲ | 7.05   | ▲ | 2.91   | ▲ |
| SPC24       | SPC24, NDC80 kinetochore complex component, homolog                    | 12166         | -1.43                    | ▼ | -1.49  | ▼ | -1.50  | ▼ |
| WDR51A      | WD repeat domain 51A                                                   | 51460         | -1.64                    | ▼ | -1.67  | ▼ | 1.57   | ▲ |
| CYP1A1      | cytochrome P450, family 1, subfamily A, polypeptide 1                  | 68062         | 13.64                    | ▲ | 153.52 | ▲ | 134.16 | ▲ |
| CCNB1       | cyclin B1                                                              | 68982         | -1.39                    | ▼ | -2.12  | ▼ | 1.42   | ▲ |
| GSTA5       | glutathione S-transferase A5                                           | 74378         | -2.47                    | ▼ | 1.47   | ▲ | 5.85   | ▲ |
| UGT1A6      | UDP glucuronosyltransferase 1 family, polypeptide A6                   | 85959         | 2.09                     | ▲ | 2.24   | ▲ | 2.06   | ▲ |

<sup>a</sup> Maximum absolute fold change determined by microarray analysis

<sup>b</sup> Differentially regulated genes with P1(t) > 0.999 and |fold change| > 1.4

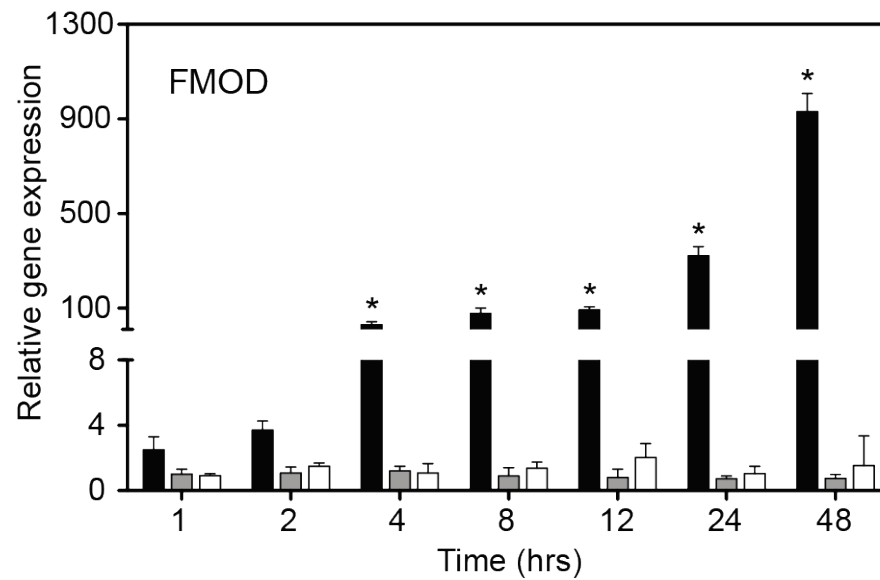
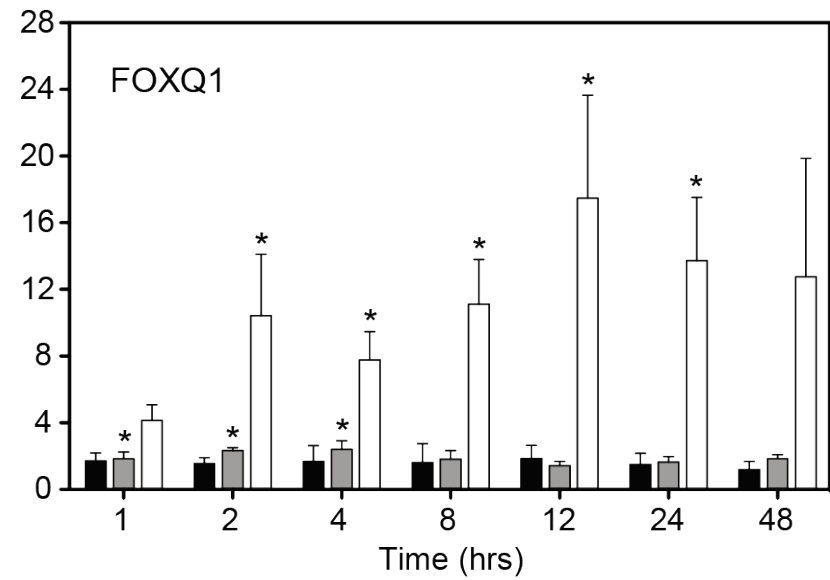
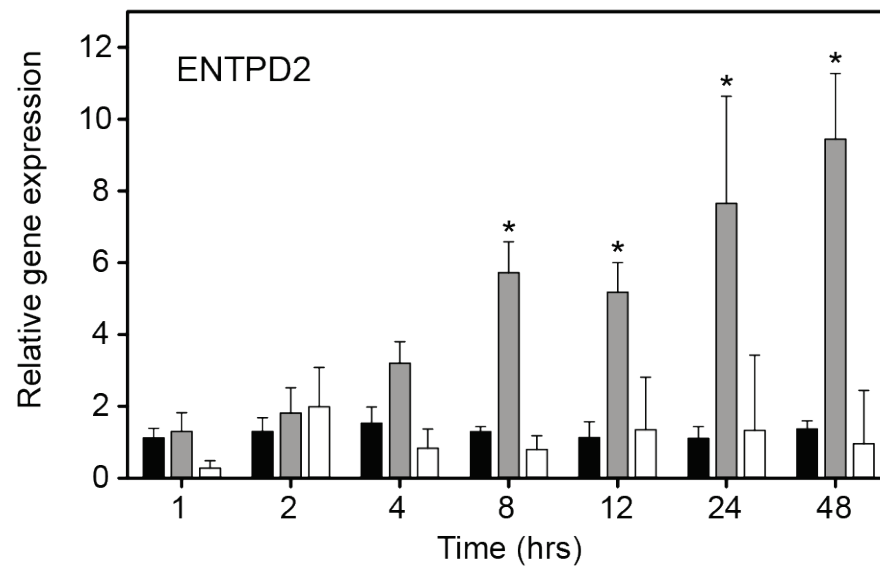
### ***SPECIES-SPECIFIC & SPECIES-DIVERGENT GENE EXPRESSION RESPONSES***

Whole genome comparative analysis of HepG2, Hepa1c1c7 and H4IIE responses identified genes that were species-specific, i.e., differentially regulated in only a single species. For example, microarray analysis at 24 hrs found that fibromodulin (FMOD, HomoloGene ID 1530) was significantly up-regulated 17.2-fold in the HepG2 cells while no significant change in expression was detected in the Hepa1c1c7 and H4IIE cells. Other examples of mouse and rat specific responses include forkhead box Q1 (FOXQ1, HomoloGene ID 7359) and ectonucleoside triphosphate diphosphohydrolase 2 (ENTPD2, HomoloGene ID 20333). FOXQ1 was up-regulated 5.9-fold in the Hepa1c1c7 cells while ENTPD2 was up-regulated 3.2-fold in the H4IIE cells. For both of these genes, the corresponding ortholog in the other two species did not exhibit any significant changes in expression. The species-specific responses of FMOD, FOXQ1 and ENTPD were further verified using QRTPCR (Figure 12). These responses are in concordance with previous reports of species-specific TCDD elicited hepatic gene expression characterized in mice and rats [17, 18].

Comparative analysis of the orthologous gene expression responses in HepG2, Hepa1c1c7 and H4IIE datasets identified 10 orthologs that were differentially expressed by TCDD across the three models. Further analysis indicated that not all responses were directionally conserved, i.e., pattern of expression was not the same in all species. For example, IER3 was induced 1.4-fold in Hepa1c1c7 cells at 4 hrs and 1.5-fold in H4IIE cells at 12 hrs, but repressed -1.9-fold in HepG2 cells at 24 hrs. Likewise, glutathione S-transferase alpha 5 (GSTA5, HomoloGene ID 74378) was induced 1.5-fold in Hepa1c1c7 cells and 4.9-fold in H4IIE cells, but down-regulated 2.5-fold in HepG2 cells. QRTPCR confirmed the divergent regulation of GSTA5 in the rat and human cell lines, but found Hepa1c1c7 cells were relatively

**Figure 12. QRTPCR verification of examples of species-specific orthologous gene expression responses identified from whole-genome microarray analysis at 24 hrs in the human HepG2, mouse Hepa1c1c7 and rat H4IIE cell lines.**

The same RNA used for microarray analysis was examined by QRTPCR. Representative orthologs that exhibited significant expression in only one species were verified; FMOD, ENTPD2 and FOXQ1 and were differentially expressed in only the HepG2, Hepa1c1c7 and H4IIE cells respectively. Fold changes were calculated relative to time-matched vehicle controls. Error bars represent the standard error of measurement for the average fold change and the asterisk (\*) indicates  $p < 0.05$ .



■ HepG2  
 ■ Hepa1c1c7  
 □ H4IIE

non-responsive (Figure 13A). Two other orthologs, cyclin D1 (CCND, HomoloGene ID 1334) and inhibitor of DNA binding 3 (ID3, HomoloGene ID 1633), also exhibited divergent responses between species, where the orthologs were down-regulated in the rodent cell lines but induced in HepG2 cells (Table 7). These species-specific and species-divergent responses are consistent with TCDD elicited hepatic differential gene expression reported in mice and rats [17, 18], where similar results were observed *in vivo* between the rodent models.

### **FUNCTIONAL ENRICHMENT ANALYSIS**

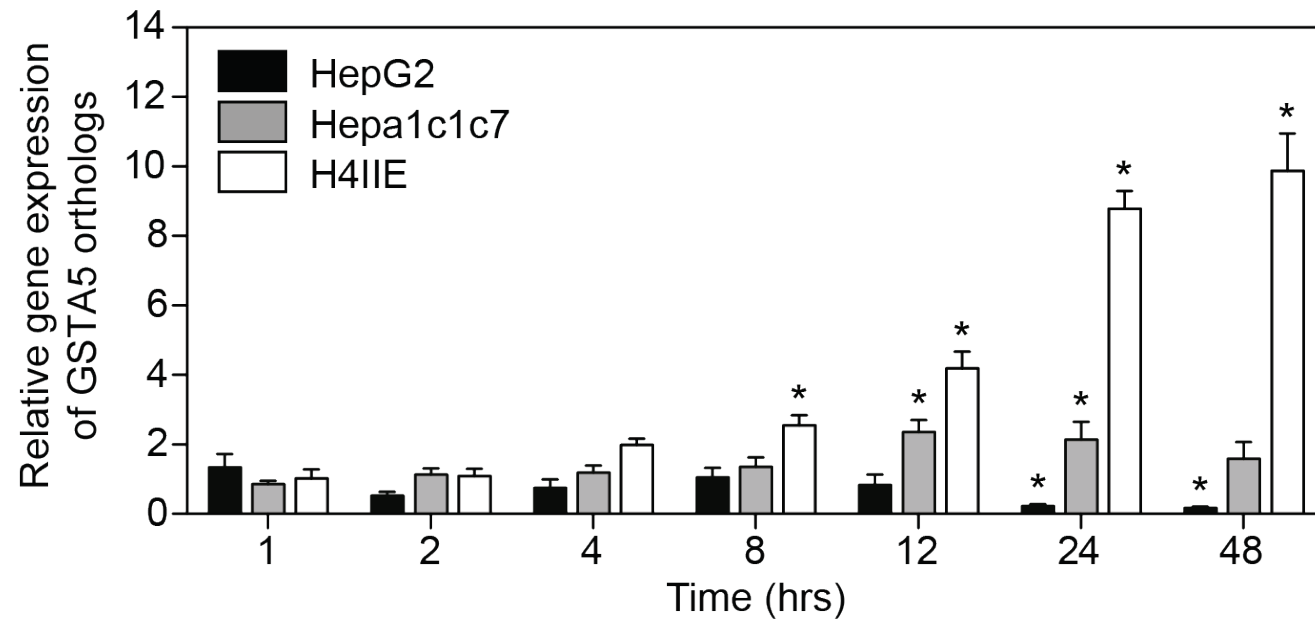
Microarray analysis identified a small subset of TCDD responsive genes in the entire genome of each species. Clustering the list of differentially responsive genes from each species based on GO terms identified functionally conserved molecular and biological processes. Among the enriched categories common to all three species included GO terms related to xenobiotic response (GO:0009410) and metabolism (GO:0006805), and glutathione transferase activity (GO:0004364). The conservation of these functional categories was expected since TCDD induces the expression of many xenobiotic metabolizing enzymes. However, other GO terms were not enriched among all three species. Despite the nature of these continuous cell lines and their dysregulation of genes related to cell cycle control and regulation, GO terms associated with these functions were not enriched by TCDD treatment in all the cell lines. Analysis of both the human HepG2 and mouse Hepa1c1c7 cell lines identified significantly enriched functional clusters associated with cell division (GO:0051301) and cell cycle (GO:0007049), while none were observed in the rat H4IIE cells. Species-differences in functional clustering of TCDD-elicited differentially expressed genes further corroborate species-specific responses to TCDD.

**Figure 13. Comparative analysis of GSTA5 orthologs.**

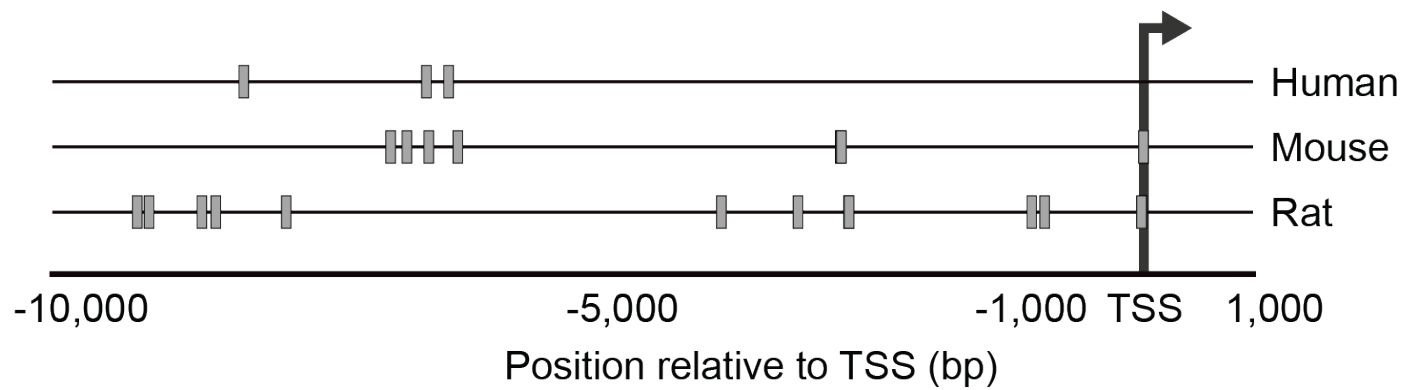
(A) QRT-PCR verification of the divergent expression of GSTA5 orthologs in human HepG2, mouse Hepalclc7 and rat H4IIE cell lines. Fold changes were calculated relative to time-matched vehicle controls. Error bars represent the standard error of measurement for the average fold change and  $*p < 0.05$ . (B) Gene regulatory regions (-10,000 to 1,000 bp relative to the transcription start site [TSS]) were computationally searched for the 5 bp DRE core (GCGTC) and indicated by the grey boxes. Each core was extended by the flanking 7 bp, and the resulting 19 bp sequence was compared to a consensus DRE sequence using a position weight matrix developed from *bona fide* functional DREs.



A)



B)



## **DIOXIN RESPONSE ELEMENT ANALYSIS**

Putative functional DREs are not equally distributed between species with more DREs associated with known human genes [23] (Dere *et. al*, manuscript in submission). Within the human genome, there are 213,355 DRE cores in the proximal promoter regions of known genes (10kb upstream to 1kb downstream of a transcriptional start site [TSS]), with 139,289 and 100,614 DRE cores, in the mouse and rat genomes, respectively. To further investigate the divergent expression of CCND1, ID3 and GSTA5, their proximal promoter regions were searched for DRE cores. Human orthologs of CCND1 and ID3 contained 12 and 14 DRE cores, respectively, greater than the number found in either the mouse or rat genomes. In contrast, mouse and rat orthologs of GSTA5 each had 7 and 12 DRE cores, respectively, while the human only had 3 (Figure 13B). In each of these divergently regulated orthologs, the species with the greatest number of DRE cores within the regulatory region had the highest fold induction suggesting that species-specific regulons have important roles in regulating gene expression.

Each DRE core in the GSTA5 orthologs was extended by 7 bp on either end and assessed for sequence similarity by measuring the Euclidean distance between sequence pairs. Only the mouse and rat GSTA5 contained highly similar DRE sequences (Euclidian distance of  $\leq 3.0$ ), and no human DRE sequences with high sequence similarity. Furthermore, the position of the conserved DRE sequence 5 bp upstream of the TSS in the mouse appears to be positionally conserved with the DRE sequence in rat the ortholog located 22 bp upstream of the TSS. Collectively, the disproportionate number of DREs between species and lack of sequence and spatially conserved DREs may account for the divergent regulation of human, mouse and rat GSTA5 orthologs.

## DISCUSSION

This study comprehensively and systematically compares the gene expression responses elicited by TCDD across human, mouse and rat cells. Incorporating both custom cDNA microarrays to profile the temporal responses and more comprehensive commercial oligonucleotide microarrays, identified a limited number of conserved responses between species. In addition, divergent and a large number of species-specific responses were identified that may contribute to species-specific differences in sensitivity and toxicity. These results are consistent with the poor response correlations of orthologous genes between C57BL/6 mice, Sprague Dawley and Long-Evans rats [17, 18]. Despite the conservation of the AhR [4], there are significant differences in TCDD elicited gene expression between species.

Previous *in vivo* and *in vitro* studies examining TCDD-elicited global gene expression demonstrated that the AhR targets a limited portion of the genome [17-19, 24-28]. In addition, PWM-based computational searches identified a low percentage of orthologs with conserved putative functional DREs within their regulatory regions (10kb upstream of the TSS and the 5'UTR) [23] (Dere *et. al*, manuscript in submission). Our comparative *in vitro* microarray results corroborate these findings. Temporal analysis using custom cDNA microarrays found that TCDD elicited a response in 9.9% of the represented genes in the HepG2 cells, 5.2% in the Hep1c1c7 cells and only 1.1% in the H4IIE cells. Similar results were obtained using whole-genome microarrays where 4.7%, 2.4% and 0.7% of the genes exhibiting differential expression in HepG2, Hep1c1c7 and H4IIE cells, respectively, at 24 hrs.

All three cell lines differentially expressed a core set of conserved gene responses that included the induction of CYP1A1, NQO1 and UGT1A6, members of the AhR gene battery [29]. However, a significant number of responses were specific to a single species (Figures 9B

and 11B), as reported in *in vivo* studies [17, 18]. These studies not only identified species-specific responses, but also orthologs with divergent responses (i.e., same gene up-regulated in one species and down-regulated in other). Comparisons of C57BL/6 mouse and Sprague Dawley rat responses found that 29% of the commonly regulated orthologs exhibited divergent regulation [18]. Similarly, GSTA5, CCND1 and ID3, exhibited divergent regulation across HepG2, Hep1c1c7 and H4IIE cells (Table 7). Each of these genes exhibited the same pattern with Hep1c1c7 and H4IIE cells having comparable profiles while HepG2 cells exhibited the divergent response. For example, GSTA5 was up-regulated in Hep1c1c7 and H4IIE cells, but down-regulated in HepG2 cells (Table 7). QRTPCR verified the divergent response of GSTA5 in the H4IIE and HepG2 cells, but did not detect a significant induction in the Hep1c1c7 cells (Figure 13A).

Computational analysis of GST5A found a disproportionate number of DRE cores within the regulatory region sequence of each species (Figure 13B). Sequence analysis also found DRE sequences with high similarity in the mouse and rat orthologs, but not in the human GSTA5. This identification of species-specific DREs is consistent with the divergent regulation of GSTA5 between these cell lines. Overall, differences in ortholog expression may contribute to species-specific differences in TCDD sensitivity and toxicity.

Although a core set of conserved gene responses was identified, consistent with the role of AhR in mediating the adaptive metabolic responses, further evidence of species-specific genome-wide gene expression profiles (i.e., species-specific regulons) is also provided. This is consistent with species-specific differences in TCDD sensitivity and toxicity [6, 8, 30, 31], as well as the reported *in vivo* species-specific gene expression profiles [17, 18], and the lack of conserved putative DREs [23]. Undoubtedly, the limited number of conserved responses and bias

towards humans and mice is due to the immaturity of the annotation for the rat genome, limits the overall interspecies comparison [32]. Differences in AhR levels, co-activator availability, and protocols used in their isolation of these hepatoma cells may also have introduced biases. Nevertheless, the identification of numerous species-specific responses, evidence of divergent gene expression responses between species, and the discovery of distinct putative primary response sets in each cell line provides further compelling evidence that the effects of TCDD post-AhR binding are not conserved between species.

## REFERENCES

## REFERENCES

1. Poland A, Knutson JC: **2,3,7,8-tetrachlorodibenzo-p-dioxin and related halogenated aromatic hydrocarbons: examination of the mechanism of toxicity.** *Annu Rev Pharmacol Toxicol* 1982, **22**:517-554.
2. Denison MS, Heath-Pagliuso S: **The Ah receptor: a regulator of the biochemical and toxicological actions of structurally diverse chemicals.** *Bulletin of environmental contamination and toxicology* 1998, **61**:557-568.
3. Hankinson O: **The aryl hydrocarbon receptor complex.** *Annu Rev Pharmacol Toxicol* 1995, **35**:307-340.
4. Schmidt JV, Bradfield CA: **Ah receptor signaling pathways.** *Annu Rev Cell Dev Biol* 1996, **12**:55-89.
5. Olson H, Betton G, Robinson D, Thomas K, Monro A, Kolaja G, Lilly P, Sanders J, Sipes G, Bracken W, et al: **Concordance of the toxicity of pharmaceuticals in humans and in animals.** *Regul Toxicol Pharmacol* 2000, **32**:56-67.
6. Schwetz BA, Norris JM, Sparschu GL, Rowe UK, Gehring PJ, Emerson JL, Gerbig CG: **Toxicology of chlorinated dibenzo-p-dioxins.** *Environ Health Perspect* 1973, **5**:87-99.
7. Denison MS, Vella LM, Okey AB: **Structure and function of the Ah receptor for 2,3,7,8-tetrachlorodibenzo-p-dioxin. Species difference in molecular properties of the receptors from mouse and rat hepatic cytosols.** *J Biol Chem* 1986, **261**:3987-3995.
8. Olson JR, Holscher MA, Neal RA: **Toxicity of 2,3,7,8-tetrachlorodibenzo-p-dioxin in the golden Syrian hamster.** *Toxicol Appl Pharmacol* 1980, **55**:67-78.
9. Poland A, Glover E, Kende AS: **Stereospecific, high affinity binding of 2,3,7,8-tetrachlorodibenzo-p-dioxin by hepatic cytosol. Evidence that the binding species is receptor for induction of aryl hydrocarbon hydroxylase.** *J Biol Chem* 1976, **251**:4936-4946.
10. Dere E, Boverhof DR, Burgoon LD, Zacharewski TR: **In vivo-in vitro toxicogenomic comparison of TCDD-elicited gene expression in Hepa1c1c7 mouse hepatoma cells and C57BL/6 hepatic tissue.** *BMC Genomics* 2006, **7**:80.
11. Kim S, Dere E, Burgoon LD, Chang C-C, Zacharewski TR: **Comparative analysis of AhR-mediated TCDD-elicited gene expression in human liver adult stem cells.** *Toxicol Sci* 2009, **112**:229-244.
12. Kwekel JC, Forgacs AL, Burgoon LD, Williams KJ, Zacharewski TR: **Tamoxifen-elicited uterotrophy: cross-species and cross-ligand analysis of the gene expression program.** *BMC medical genomics* 2009, **2**:19.

13. Yang YH, Speed T: **Design issues for cDNA microarray experiments.** *Nat Rev Genet* 2002, **3**:579-588.
14. Eckel J, Gennings C, Therneau T, Burgoon L, Boverhof D, Zacharewski T: **Normalization of two-channel microarray experiments: a semiparametric approach.** *Bioinformatics* 2005, **21**:1078-1083.
15. Eckel J, Gennings C, Chinchilli V, Burgoon L, Zacharewski T: **Empirical bayes gene screening tool for time-course or dose-response microarray data.** *J Biopharm Stat* 2004, **14**:647-670.
16. Dennis G, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, Lempicki RA: **DAVID: Database for Annotation, Visualization, and Integrated Discovery.** *Genome Biol* 2003, **4**:P3.
17. Boutros PC, Yan R, Moffat ID, Pohjanvirta R, Okey AB: **Transcriptomic responses to 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD) in liver: comparison of rat and mouse.** *BMC Genomics* 2008, **9**:419.
18. Boverhof DR, Burgoon LD, Tashiro C, Sharratt B, Chittim B, Harkema JR, Mendrick DL, Zacharewski TR: **Comparative toxicogenomic analysis of the hepatotoxic effects of TCDD in Sprague Dawley rats and C57BL/6 mice.** *Toxicol Sci* 2006, **94**:398-416.
19. Fletcher N, Wahlström D, Lundberg R, Nilsson CB, Nilsson KC, Stockling K, Hellmold H, Håkansson H: **2,3,7,8-Tetrachlorodibenzo-p-dioxin (TCDD) alters the mRNA expression of critical genes associated with cholesterol metabolism, bile acid biosynthesis, and bile transport in rat liver: a microarray study.** *Toxicol Appl Pharmacol* 2005, **207**:1-24.
20. Lusska A, Wu L, Whitlock JP: **Superinduction of CYP1A1 transcription by cycloheximide. Role of the DNA binding site for the liganded Ah receptor.** *J Biol Chem* 1992, **267**:15146-15151.
21. Ma Q: **Induction and superinduction of 2,3,7,8-tetrachlorodibenzo-p-dioxin-inducible poly(ADP-ribose) polymerase: role of the aryl hydrocarbon receptor/aryl hydrocarbon receptor nuclear translocator transcription activation domains and a labile transcription repressor.** *Arch Biochem Biophys* 2002, **404**:309-316.
22. Joiakim A, Mathieu PA, Elliott AA, Reiners JJ: **Superinduction of CYP1A1 in MCF10A cultures by cycloheximide, anisomycin, and puromycin: a process independent of effects on protein translation and unrelated to suppression of aryl hydrocarbon receptor proteolysis by the proteasome.** *Molecular Pharmacology* 2004, **66**:936-947.
23. Sun YV, Boverhof DR, Burgoon LD, Fielden MR, Zacharewski TR: **Comparative analysis of dioxin response elements in human, mouse and rat genomic sequences.** *Nucleic Acids Res* 2004, **32**:4512-4523.



24. Jin B, Kim G, Park DW, Ryu D-Y: **Microarray analysis of gene regulation in the Hepa1c1c7 cell line following exposure to the DNA methylation inhibitor 5-aza-2'-deoxycytidine and 2,3,7,8-tetrachlorodibenzo-p-dioxin.** *Toxicol In Vitro* 2004, **18**:659-664.
25. McHale CM, Zhang L, Hubbard AE, Zhao X, Baccarelli A, Pesatori AC, Smith MT, Landi MT: **Microarray analysis of gene expression in peripheral blood mononuclear cells from dioxin-exposed human subjects.** *Toxicology* 2007, **229**:101-113.
26. Puga A, Maier A, Medvedovic M: **The transcriptional signature of dioxin in human hepatoma HepG2 cells.** *Biochem Pharmacol* 2000, **60**:1129-1142.
27. Silkworth JB, Carlson EA, McCulloch C, Illouz K, Goodwin S, Sutter TR: **Toxicogenomic analysis of gender, chemical, and dose effects in livers of TCDD- or aroclor 1254-exposed rats using a multifactor linear model.** *Toxicol Sci* 2008, **102**:291-309.
28. Xiong KM, Peterson RE, Heideman W: **Aryl hydrocarbon receptor-mediated down-regulation of *sox9b* causes jaw malformation in zebrafish embryos.** *Molecular Pharmacology* 2008, **74**:1544-1553.
29. Nebert DW, Roe AL, Dieter MZ, Solis WA, Yang Y, Dalton TP: **Role of the aromatic hydrocarbon receptor and [Ah] gene battery in the oxidative stress response, cell cycle control, and apoptosis.** *Biochem Pharmacol* 2000, **59**:65-85.
30. Bickel M: **Polychlorinated persistent compounds.** *Experientia* 1982, **38**:879-882.
31. Vos JG, Moore JA, Zinkl JG: **Toxicity of 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD) in C57B1/6 mice.** *Toxicol Appl Pharmacol* 1974, **29**:229-241.
32. Boverhof DR, Zacharewski TR: **Toxicogenomics in risk assessment: applications and needs.** *Toxicol Sci* 2006, **89**:352-360.

## CHAPTER 5

Dere E, Forgacs AL, Zacharewski TR and Burgoon LD: **Genome-Wide Computational Analysis of Dioxin Response Element Locations and Distributions in the Human, Mouse and Rat Genomes.** *Chemical Research In Toxicology* (Submitted September 21, 2010)

## CHAPTER 5

# GENOME-WIDE COMPUTATIONAL ANALYSIS OF DIOXIN RESPONSE ELEMENT LOCATION AND DISTRIBUTION IN THE HUMAN, MOUSE AND RAT GENOMES

### ABSTRACT

The aryl hydrocarbon receptor (AhR) mediates responses elicited by 2,3,7,8-tetrachlorodibenzo-*p*-dioxin through the core consensus sequence 5'-GCGTG-3' dioxin response elements (DRE). The human, mouse and rat genomes were computationally searched for all DRE cores. Each core was then extended by 7 bp upstream and downstream and the matrix similarity (MS) scores for the resulting 19 bp DRE sequences were calculated using a revised position weight matrix constructed from bona fide functional DREs. In total, 72,318 human, 70,720 mouse and 88,651 rat high-scoring ( $MS \geq 0.8473$ ) putative DREs were identified. Gene encoding intragenic DNA regions had ~1.6-times more putative DREs than the non-coding intergenic DNA regions. Furthermore, the promoter region spanning  $\pm 1.5$  kb of a TSS had the highest density of putative DREs within the genome. Chromosomal analysis found that densities of chromosomes X and Y were significantly lower than the mean chromosomal density. Interestingly, the 10 kb upstream promoter region on chromosome X of the genomes were significantly less dense than the chromosomal mean, while the same region in chromosome Y was the most dense. In addition to providing a detailed genomic map of all DRE cores in the human, mouse and rat genomes, these data will further aid the elucidation of AhR-mediated signal transduction.

## INTRODUCTION

*Cis*-regulatory elements located in the promoter region of genes are transcription factor binding sites that regulate gene expression and most transcription factors have a preferred response element sequence to which they bind. The identification and location of these elements is important in elucidating transcription factor binding, signal transduction, and ultimately, their gene expression networks. Binding to elements in the proximal promoter region stabilizes the general transcriptional machinery at the transcriptional start site (TSS) to regulate gene expression. However, global location analyses of transcription factor binding using ChIP-chip and ChIP-seq technologies have demonstrated binding at sites distant from the TSS, which also have important roles in regulating gene expression [1-4]. A comprehensive map of transcription factor binding element locations and distribution within a genome provides valuable information in elucidating and modeling the gene expression network of a transcription factor.

The AhR is a ligand activated transcription factor belonging to the basic-helix-loop-helix-PAS (bHLH-PAS) family of proteins that serve as environmental sensors to different stimuli [5]. 2,3,7,8-Tetrachlorodibenzo-*p*-dioxin (TCDD) is the prototypical AhR ligand, a widespread environmental contaminant that elicits diverse species-specific effects, including tumor promotion, teratogenesis, hepatotoxicity, modulation of endocrine systems, immunotoxicity, and enzyme induction [6, 7]. These effects are a result of changes in gene expression mediated by the AhR [8]. The binding of TCDD and related compounds to the cytosolic AhR triggers a conformational change and translocation of the activated AhR complex to the nucleus where it heterodimerizes with the aryl hydrocarbon nuclear translocator (ARNT), another bHLH-PAS family member. The heterodimer then binds to dioxin response elements (DREs) containing the substitution intolerant 5'-GCGTG-3' core, to regulate gene expression [8, 9]. Evidence indicates

that nucleotides adjacent to the core consensus sequence modulate DNA-binding affinity and enhancer function [10-12].

Position weight matrices (PWMs) provide a similarity assessment of a motif or putative response element [13], and when compared to a consensus sequence can be used to rank and prioritize potential transcription factor binding site preferences. However, PWMs suffer from high false positive prediction rates since the probability of any nucleotide at any position within the binding site is assumed to be independent of those occurring at all other positions. Fortunately, the DRE PWM is based on the substitution intolerant 5'-GCGTG-3' core, thus reducing the frequency of false positives [14].

We have previously identified the location and distribution of DREs relative to the TSS for a limited number of genes [14] from prior builds of the human, mouse and rat genome assemblies [14]. Improvements and innovations in sequencing technologies have provided higher quality sequence data with significantly fewer sequence gaps [15-17] in the most recent genome resulting in better gene annotation. The latest mouse and rat genome builds were used to construct a revised PWM based on updated sequence information for 13 *bona fide* functional DREs. In addition, we expanded the scope of our computational search and distributional analysis of DREs to include the entire human, mouse and rat genomes. This includes analyses of the intragenic (10 kb upstream to end of 3' untranslated regions [UTRs]) and intergenic DNA regions, chromosome and gene regions (10 kb upstream of a TSS, 5' and 3' UTRs, and coding sequence [CDS]). Collectively, these results provide a detailed genomic map for all putative DREs in the human, mouse and rat genomes that will serve as an important resource for the further elucidation of AhR gene expression networks.

## METHODS

### ***POSITION WEIGHT MATRIX***

We have previously constructed a PWM using 13 *bona fide* functional DRE sequences from previous assembly builds of the mouse (mm3) and rat (rn2) genomes [14]. These sequences were updated using the sequence information from the current genome assemblies for the mouse (mm9) and rat (rn4) (Table 8, updated sequences are underlined). Additionally, the previously identified sequence for the *bona fide* rat *Aldh3a1* DRE could no longer be found in the rn4 genome build and was replaced with a functional DRE located 6,787 bp upstream of the TSS [18]. Also note that the gene names for *GstYa* and *Ugt1a1* have changed to *Gsta2* and *Ugt1a6*, respectively, in the latest rat assembly. Updated sequences were used to develop a revised *PWM* using the *bona fide* 19 bp DRE-centered sequences (Figure 14). The replaced rat *Aldh3a1* DRE sequence had the lowest matrix similarity (MS) score (0.8473), which was subsequently used as a threshold value to define 19 bp DRE sequences as putative DREs that were functional.

### ***WHOLE-GENOME IDENTIFICATION OF DRES***

Sequences for human (hg19), mouse (mm9), and rat (rn4) genome assemblies and their associated annotation within the refGene and refLink databases, were downloaded from the UCSC Genome Browser [19]. Individual segments of a gene region (i.e., the 10 kb sequence upstream of a TSS, the 5' and 3' UTRs and the CDS) for each mature gene encoding reference sequence (RefSeqs with NM prefixed identifiers), were determined using the genomic coordinates within the refGene databases (Figure 15A). Intragenic DNA regions within the genomes were computationally identified by merging overlapping gene regions (defined in Figure 15A) from both strands of the genome, and the DNA between adjacent intragenic regions are defined as the non-transcribed intergenic DNA regions (Figure 15B). The lengths for each of

**Table 8. *Bona fide* DRE sequences used to construct the revised<sup>a</sup> position weight matrix.**

| Species | Gene Symbol         | RefSeq Identifier | Position Relative to TSS | <i>Bona Fide</i> DRE Sequence <sup>b</sup>              | Matrix Similarity Score |
|---------|---------------------|-------------------|--------------------------|---------------------------------------------------------|-------------------------|
| Mouse   | Cyp1a1              | NM_001136059      | -491                     | caagctc <b>GCGTG</b> agaagcg                            | 0.9466                  |
|         | Cyp1a1              | NM_001136059      | -871                     | cctgtgt <b>GCGTG</b> ccaagca                            | 0.9128                  |
|         | Cyp1a1              | NM_001136059      | -984                     | cggagtt <b>GCGTG</b> agaagag                            | 0.9598                  |
|         | Cyp1a1              | NM_001136059      | -1,059                   | ccagcta <b>GCGTG</b> acagcac                            | 0.9260                  |
|         | Cyp1a1              | NM_001136059      | -1,206                   | cggtgtt <b>GCGTG</b> cgatgct                            | 0.9610                  |
|         | Cyp1b1              | NM_009994         | -872                     | ccccctt <b>GCGTG</b> cggagct                            | 0.9514                  |
| Rat     | Cyp1a1              | NM_012540         | -1,045                   | cggagtt <b>GCGTG</b> agaagag                            | 0.9598                  |
|         | Cyp1a1              | NM_012540         | -1,120                   | <u>ccagcta</u> <b>GCGTG</b> acagcac                     | 0.9260                  |
|         | Aldh3a1             | NM_031972         | -6,787                   | <u>tgccctg</u> <b>GCGTG</b> <u>actttgt</u> <sup>c</sup> | 0.8473 <sup>d</sup>     |
|         | Nqo1                | NM_017000         | -400                     | tcccctt <b>GCGTG</b> caaaggc                            | 0.9332                  |
|         | Sod1                | NM_017050         | -274                     | gaggcct <b>GCGTG</b> cgcgctt                            | 0.8481                  |
|         | Gsta2 <sup>e</sup>  | NM_017013         | -910                     | gcatgtt <b>GCGTG</b> catccct                            | 0.8728                  |
|         | Ugt1a6 <sup>e</sup> | NM_057105         | -3,856                   | agaatgt <b>GCGTG</b> acaaggt                            | 0.8950                  |

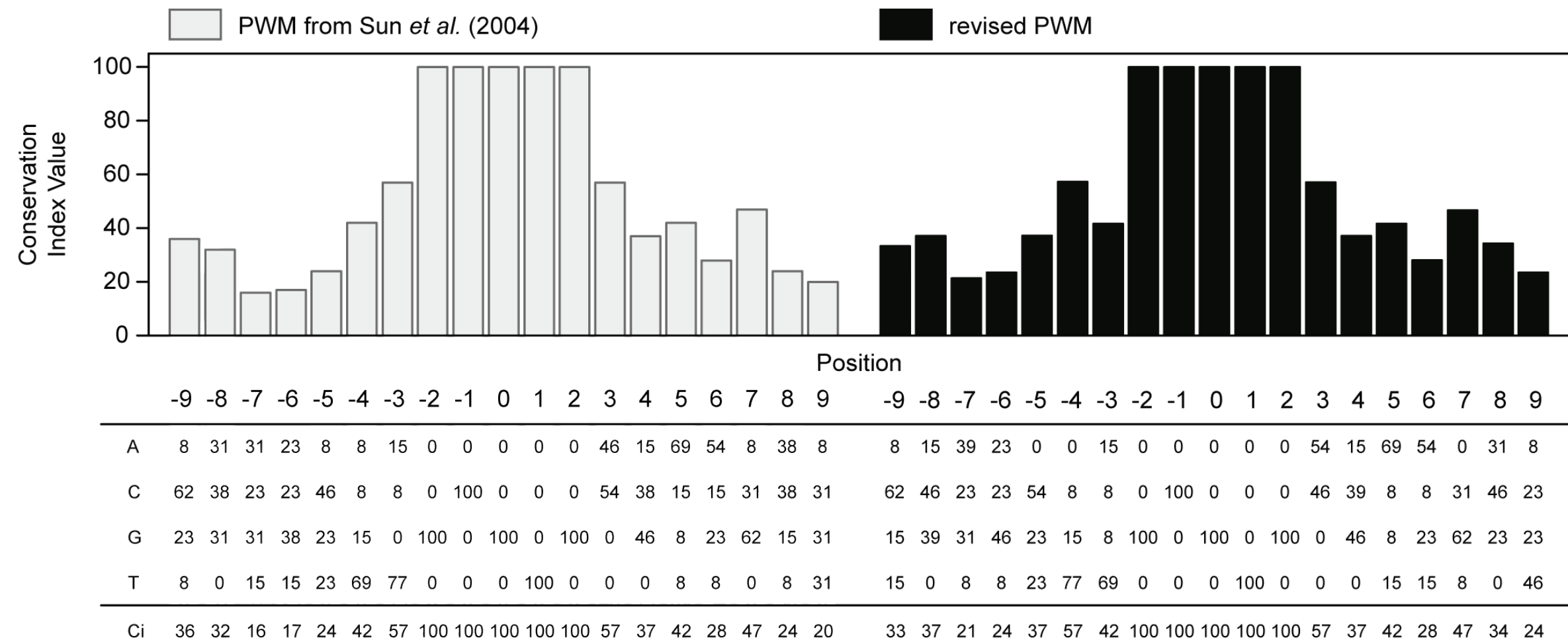
<sup>a</sup> *bona fide* DRE sequences were updated using builds mm9 and rn4 genome builds

<sup>b</sup> sequences used in *Sun et. al.*, 2004 were updated with the mm9 and rn4 genome builds; revised sequences are underlined

<sup>c</sup> replaces previous DRE sequence for rat Aldh3a1

<sup>d</sup> denotes the MS score used as the threshold score

<sup>e</sup> Gsta2 and Ugt1a6 were previously named GstYa and Ugt1a1 respectively, and were renamed within the rn4 genome build



**Figure 14. Comparison of the previously published position weight matrix (PWM) and conservation index ( $C_i$ ) for dioxin response elements (DREs) with the revised PWM.** The matrix and plot are the revised PWM and  $C_i$  using the current mouse (mm9) and rat (rn4) genome assemblies from the UCSC Genome Browser. The matrix (bottom) shows the percentage of occurrence for a specific nucleotide at that given position. For example, positions -2 to 2 define the substitution intolerant 5'-GCGTG-3' core, each nucleotide within the core has a  $C_i$  value of 100. The histogram (top) is a graphical representation of the  $C_i$  values, which are listed below the PWM. The  $C_i$  provides a measure of conservation at each base pair position. If a PWM is 100% conserved at a position, the  $C_i$  value is 100, whereas if the position is truly random (A=25%, C=25%, G=25%, T=25%) then the  $C_i$  value is 0.



these defined regions and the number of RefSeqs on each chromosome are provided for the human, mouse and rat genomes in Supplementary Table 1. In total, 28,906 human, 24,327 mouse, and 15,737 rat mature RefSeqs were searched. Gene annotation associated with each RefSeq sequence was derived from the refLink database in the UCSC Genome Browser.

The sequence of each individual chromosome was computationally searched for the substitution intolerant 5'-GCGTG-3' core sequence using a previously described search algorithm [14]. Each core was then extended by 7 bp upstream and downstream of the core. MS scores for the 19 bp DRE sequences were calculated using the revised PWM. For genomic location analysis, the position of a DRE core is defined as the center base (5'-GCGTG-3') of the 5 bp core sequence (underlined). Putative DRE densities were calculated based on the number of putative DREs occurring in an interrogated region (e.g., intergenic DNA region or 5' UTR) divided by the total sum of the region length. Results from the computational genome-wide DRE search can be downloaded as bedGraph track format (Supplementary file 5-7) and uploaded to the UCSC Genome Browser for visualization (Figure 16).

### ***RANDOM SEQUENCE COMPARISON***

To investigate the random frequency of DRE cores within each genome, 25,000 random sequences of 15 kb in length were computationally generated by randomly selecting A, C, G or T's. These sequences were then searched for DRE cores, and the 19 bp DRE sequence MS score was calculated using the described algorithm [14] with the revised PWM.

### ***MICROARRAY ANALYSIS***

Whole-genome microarray analysis of hepatic tissue from mice orally gavaged with 30 µg/kg TCDD was performed using 4×44k whole genome oligonucleotide arrays from Agilent Technologies (Santa Clara, CA). The same RNA from a previous study was used for the gene

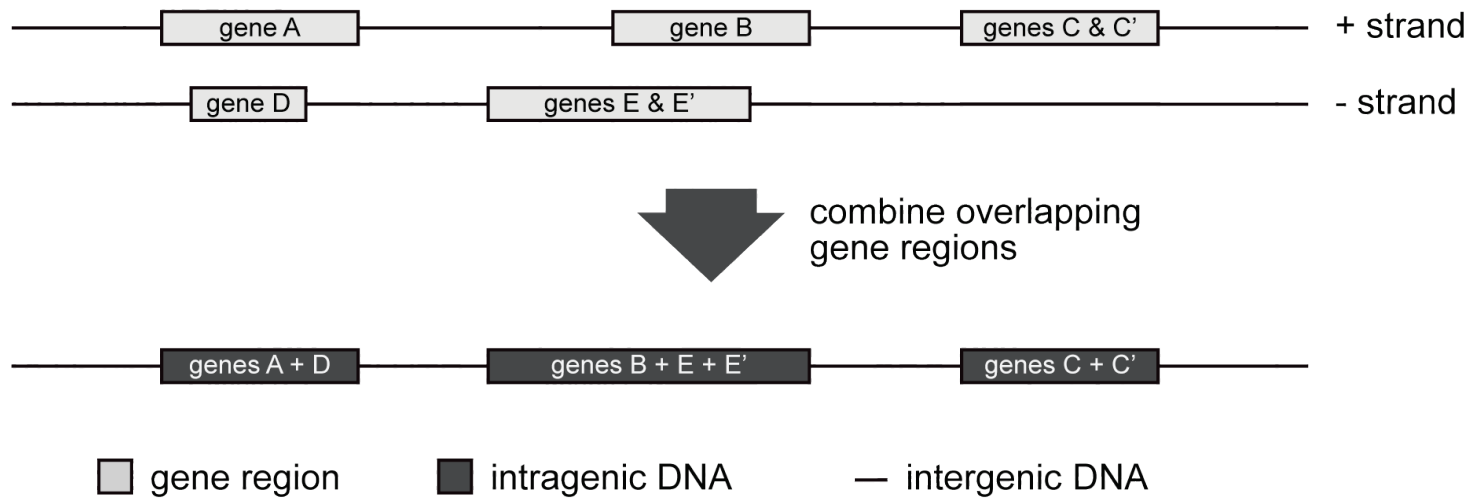
**Figure 15. Defining the various genomic regions used for DRE location analysis.**

A) Genomic locations from the UCSC Genome Browser refGene database were used to obtain sequences for 10 kb region upstream of the TSS, the 5' and 3' UTRs, and the CDS of every known human, mouse and rat RefSeq sequence. A gene region is defined as the sequence spanning the region 10 kb upstream of a TSS through to the end of the 3' UTR. B) Intragenic DNA regions in a genome were determined by combining the non-overlapping gene regions. For example, gene regions of tissue specific isoforms of a gene that have different TSS positions were merged to determine the longest spanning range (genes C & C' and genes E & E'). Additionally, overlapping genes on both strands of the genome were also merged (genes B + E + E'). Non-transcribed DNA segments that span the regions between adjacent intragenic regions are defined as the intergenic DNA regions.

A)



B)



expression profiling [20]. Changes in gene expression due to TCDD treatment were conducted according to the manufacturer's Two-Color Microarray-Based Gene Expression Analysis protocol Version 5.0.1. Microarray data were normalized using a semiparametric method [21], and statistically analyzed using an empirical Bayes method [22]. Differentially expressed genes were determined by both a fold change and a statistical cutoff ( $|\text{fold change}| \geq 1.5$  and  $P(t) \geq 0.99$ ).

## RESULTS

### ***POSITION WEIGHT MATRIX (PWM)***

Our previous PWM used *bona fide* DRE sequence information from earlier drafts of the mouse (mm3) and rat (rn2) genomes (Figure 14). These sequences have since been updated with the most current information available from the mouse (mm9) and rat (rn4) genome assemblies (Table 8). As a result, the sequence of two *bona fide* DREs in the promoter region of the mouse and rat *Cyp1a1* gene have changed (Table 8, see footnote b). Additionally, the previously used DRE for rat *Aldh3a1* could no longer be found in the latest rat genomic sequence, and was replaced with a recently characterized DRE located 6.8 kb upstream of the TSS [18] (Table 8). These updates altered the PWM and the conservation index ( $C_i$ ) vector, which represents the degree of conservation of the individual nucleotide position, primarily in the 7 bp flanking 5' arm of the consensus sequence (Figure 14). Recalculation of MS scores for the *bona fide* DREs identified the rat *Aldh3a1* motif as having the lowest score, 0.8473, which was subsequently used to characterize computationally identified sequences as putative DREs.

### ***GENOME-WIDE DISTRIBUTION OF DRES***

Our previous computational search for the 5'-GCGTG-3' invariant DRE core was limited to sequences 5 kb upstream and 2 kb downstream of a TSS for known RefSeqs in previous

genome builds [14]. This current study extended the search to the entire human, mouse and rat genomes, including the non-transcribed intergenic DNA regions (Figure 15B). Computational searches identified 1.65, 1.04 and 1.07 million DRE cores in the human, mouse and rat genomes, respectively (Table 9). After extending these cores by the 7 bp upstream and downstream flanking sequences, MS scores were calculated using the revised PWM. A total of 72,318 human, 70,720 mouse and 88,651 rat 19 bp DRE sequences had a MS score greater than or equal to 0.8473, and were classified as putatively functional DREs (Table 9). The density of putative DREs with respect to the total length of the genomes were 23.4, 26.6, and 32.6 DREs per million base pairs (Mbp) in the human, mouse and rat, respectively. These values were determined from searching 3.10 billion human, 2.66 billion mouse and 2.72 billion rat base pairs (Table 9).

Although less than half of the genomes are comprised of intragenic DNA (Figure 15B), 53% of all putative DREs in the human and mouse genomes were identified in these regions, while only 38% of all putative DREs mapped to the intragenic DNA in the rat (Table 9). This difference is likely a result of the relatively fewer number of rat RefSeqs (15,737) compared to the human (28,906) and the mouse (24,327). Relative putative DRE densities (i.e., intragenic/intergenic DNA putative density ratio) suggest that intragenic regions have ~1.6-times greater putative DRE density compared to intergenic DNA regions in each genome. For example, the human genome had putative DRE densities per Mbp of 30.2 and 18.7 in the intragenic and intergenic DNA regions, respectively ( $30.2/18.7 = 1.6$ ). This suggests that there is a greater likelihood of putative DREs in the intragenic regions of the genome as opposed to the non-transcribed intergenic DNA regions. Further examination of DRE distribution within defined gene region segments (i.e., 10 kb upstream, 5' and 3' UTRs and CDS; Figure 15A) found that segment-specific putative DREs densities were comparable in human and mouse regions.

**Table 9. Distribution of DRE cores, putative DREs and putative DRE densities across the human, mouse and rat genomes.**

|                                              | Species | Genome    | Intergenic DNA <sup>a</sup> | Intragenic DNA <sup>a</sup> | Gene Region <sup>b</sup>   |                     |                  |                     |
|----------------------------------------------|---------|-----------|-----------------------------|-----------------------------|----------------------------|---------------------|------------------|---------------------|
|                                              |         |           |                             |                             | 10kb upstream <sup>c</sup> | 5' UTR <sup>c</sup> | CDS <sup>c</sup> | 3' UTR <sup>c</sup> |
| Region length (Mbp <sup>d</sup> )            | Human   | 3,096     | 1,836                       | 1,260                       | 295                        | 236                 | 1,599            | 44                  |
|                                              | Mouse   | 2,655     | 1,588                       | 1,067                       | 247                        | 149                 | 1,040            | 31                  |
|                                              | Rat     | 2,719     | 1,973                       | 746                         | 159                        | 74                  | 569              | 18                  |
| Total number of DRE cores                    | Human   | 1,648,651 | 759,030                     | 889,621                     | 303,016                    | 172,235             | 1,054,383        | 26,925              |
|                                              | Mouse   | 1,036,996 | 492,703                     | 544,293                     | 154,921                    | 82,306              | 510,060          | 16,546              |
|                                              | Rat     | 1,070,366 | 676,193                     | 394,173                     | 94,514                     | 41,455              | 292,526          | 9,711               |
| Total number of putative DREs <sup>e</sup>   | Human   | 72,318    | 34,322                      | 37,996                      | 13,272                     | 7,606               | 46,004           | 1,123               |
|                                              | Mouse   | 70,720    | 33,018                      | 37,702                      | 10,176                     | 6,024               | 35,819           | 1,220               |
|                                              | Rat     | 88,651    | 54,888                      | 33,763                      | 7,962                      | 3,515               | 24,954           | 870                 |
| Putative DRE density (per Mbp <sup>d</sup> ) | Human   | 23.4      | 18.7                        | 30.2                        | 45.0                       | 32.2                | 28.8             | 25.6                |
|                                              | Mouse   | 26.6      | 20.8                        | 35.3                        | 41.3                       | 40.5                | 34.5             | 39.5                |
|                                              | Rat     | 32.6      | 27.8                        | 45.3                        | 50.2                       | 47.6                | 43.9             | 49.4                |

<sup>a</sup> intergenic and intragenic DNA region are defined in Figure 15B

<sup>b</sup> gene region is defined as the transcribed gene plus 10 kb upstream of the TSS as depicted in Figure 15A

<sup>c</sup> regions are defined using the genomic locations in the refGene database from the UCSC Genome Browser

<sup>d</sup> Mbp = million basepairs

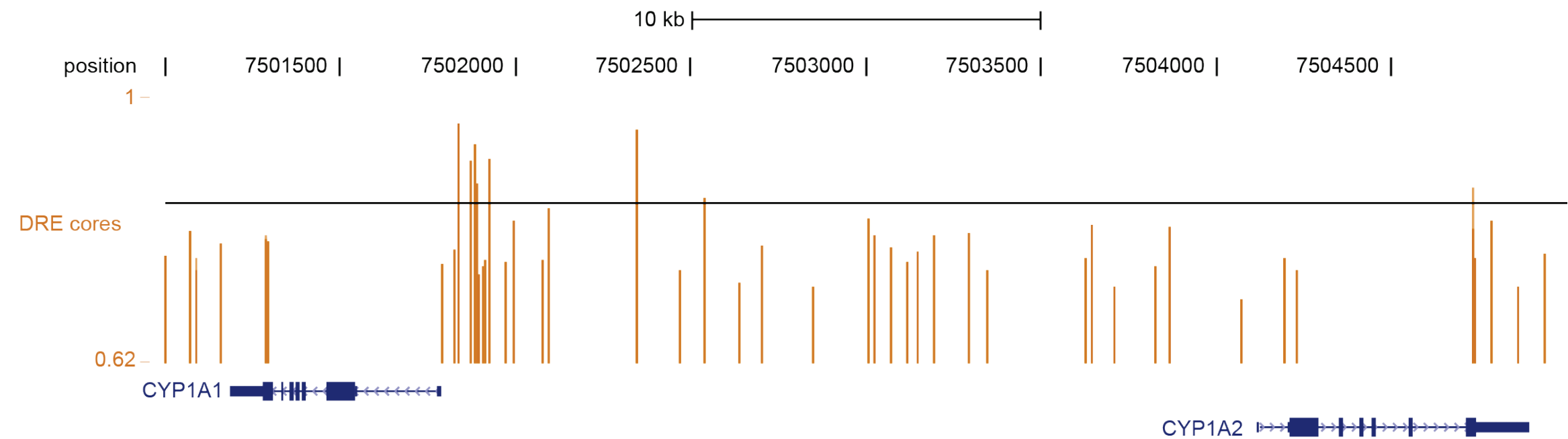
<sup>e</sup> putative DREs defined as the 19 bp DRE centered core containing sequence with a MS score  $\geq 0.8473$

However, the density of putative DREs was generally higher in the rat genome (Table 9), likely due to the relative immaturity of gene annotation associated with the rat genome. The location and MS score for each identified 19 bp DRE sequence has been loaded into the UCSC Genome browser and can be visualized as a bedGraph track (Figure 16).

### ***CHROMOSOME LEVEL ANALYSIS OF PUTATIVE DREs***

In order to further investigate putative DRE distribution across the genomes, chromosomal level analysis was performed (Tables 10 and 11). Examination of individual chromosomes identified examples where the putative DRE density was significantly different than the mean chromosomal value (outside the 99% confidence interval of the mean; Table 10, see footnotes d and e). For example, putative DRE densities for rat chromosome 2 and human chromosome 13 were 26.5 and 16.7 per Mbp, respectively, which were significantly less than the mean value for each genome (34.6 and 24.5 Mbp, for the rat and human, respectively). Furthermore, human chromosomes 16 and 17 had significantly greater putative DREs density than the mean chromosomal density. There are also instances where the putative DRE densities in the intergenic DNA (Table 10), or in a specific gene region segment (i.e., 10 kb upstream region, CDS and UTRs; Table 11), were significantly different than the chromosomal mean for that region. These data suggest that there are chromosome- and segment-specific biases in putative DRE densities across the genome that may have biological relevance in AhR-mediated responses.

Interestingly, putative DRE densities in chromosome X and Y of the human and mouse were significantly lower than the chromosomal average (Tables 10 and 11; there currently is no sequence data available for chromosome Y in the rat). For example, mouse chromosome Y has an intragenic putative DRE density of 16.4 per Mbp, almost half the density of any other mouse



**Figure 16. Visualization of DRE sequence locations in the UCSC Genome Browser for human CYP1A1 and CYP1A2 gene regions and adjacent intergenic regions.** The genomic location and MS score for each identified 19 bp DRE sequence has been loaded into the UCSC Genome Browser as a bedGraph track (see DRE cores track at top). The vertical bars represent the 5 bp DRE core and the height of the bar provides an indication of the MS score for the 19 bp DRE core containing sequence. The horizontal black line within the DRE cores track indicates the threshold MS score (0.8473) to assist with the identification of putative functional DREs.



**Table 10. Chromosomal density of putative DREs<sup>a</sup> (per Mbp<sup>b</sup>) within the intergenic and intragenic DNA regions<sup>c</sup> of the human, mouse and rat genomes.**

| Chromosome | Human             |                   |                   | Mouse             |                   |                   | Rat               |                   |                   |
|------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|
|            | Total             | Intergenic DNA    | Intragenic DNA    | Total             | Intergenic DNA    | Intragenic DNA    | Total             | Intergenic DNA    | Intragenic DNA    |
| 1          | 22.7              | 17.5              | 28.9              | 24.0              | 20.1              | 30.1 <sup>d</sup> | 32.9              | 28.4              | 42.3              |
| 2          | 22.9              | 20.0              | 27.0 <sup>d</sup> | 28.2              | 23.3              | 33.9              | 26.5 <sup>d</sup> | 23.0 <sup>d</sup> | 38.8 <sup>d</sup> |
| 3          | 19.8 <sup>d</sup> | 16.1 <sup>d</sup> | 24.2 <sup>d</sup> | 23.7              | 18.7              | 33.9              | 35.2              | 31.1              | 43.9              |
| 4          | 19.3 <sup>d</sup> | 17.2              | 23.1 <sup>d</sup> | 28.8              | 21.5              | 39.5 <sup>e</sup> | 30.1              | 26.1              | 39.1 <sup>d</sup> |
| 5          | 21.0              | 19.0              | 24.4 <sup>d</sup> | 32.3 <sup>e</sup> | 24.3 <sup>e</sup> | 42.4 <sup>e</sup> | 31.6              | 26.9              | 45.4              |
| 6          | 20.9              | 18.8              | 23.9 <sup>d</sup> | 26.2              | 21.2              | 32.8              | 32.8              | 28.2              | 46.2              |
| 7          | 25.3              | 20.7              | 30.6              | 26.7              | 19.7              | 36.7              | 34.8              | 29.2              | 49.2              |
| 8          | 24.3              | 21.6              | 28.6              | 29.6              | 23.2              | 39.9 <sup>e</sup> | 36.1              | 30.9              | 47.5              |
| 9          | 22.0              | 16.7 <sup>d</sup> | 31.3              | 30.1              | 23.9 <sup>e</sup> | 36.9              | 33.5              | 30.1              | 42.8              |
| 10         | 26.4              | 22.9 <sup>e</sup> | 30.5              | 28.6              | 22.1              | 37.6              | 43.8 <sup>e</sup> | 37.4 <sup>e</sup> | 53.9 <sup>e</sup> |
| 11         | 24.7              | 18.9              | 31.2              | 33.9 <sup>e</sup> | 26.7 <sup>e</sup> | 40.9 <sup>e</sup> | 29.6              | 24.4 <sup>d</sup> | 44.0              |
| 12         | 24.0              | 20.0              | 28.5              | 27.0              | 22.3              | 35.7              | 62.7 <sup>e</sup> | 52.0 <sup>e</sup> | 81.3 <sup>e</sup> |
| 13         | 16.7 <sup>d</sup> | 13.4 <sup>d</sup> | 25.3 <sup>d</sup> | 26.6              | 23.7 <sup>e</sup> | 31.6              | 31.3              | 26.4              | 44.0              |
| 14         | 19.9 <sup>d</sup> | 15.1 <sup>d</sup> | 28.9              | 24.8              | 21.0              | 31.1              | 32.4              | 29.0              | 43.1              |
| 15         | 23.2              | 18.6              | 30.1              | 26.7              | 19.3              | 37.8              | 29.8              | 26.1              | 41.2              |
| 16         | 35.0 <sup>e</sup> | 25.6 <sup>e</sup> | 48.0 <sup>e</sup> | 24.6              | 19.4              | 32.9              | 33.3              | 27.6              | 48.3              |
| 17         | 34.8 <sup>e</sup> | 26.6 <sup>e</sup> | 40.7 <sup>e</sup> | 31.0 <sup>e</sup> | 22.7              | 40.9 <sup>e</sup> | 36.1              | 33.6              | 43.1              |
| 18         | 23.3              | 19.6              | 29.8              | 26.7              | 20.2              | 37.7              | 30.7              | 26.7              | 43.3              |
| 19         | 43.6 <sup>e</sup> | 29.7 <sup>e</sup> | 53.3              | 30.8 <sup>e</sup> | 23.2              | 38.4              | 43.2 <sup>e</sup> | 38.6 <sup>e</sup> | 53.5              |
| 20         | 32.5 <sup>e</sup> | 27.0 <sup>e</sup> | 38.7 <sup>e</sup> |                   |                   |                   | 45.0 <sup>e</sup> | 34.4 <sup>e</sup> | 68.5 <sup>e</sup> |

**Table 10. (cont'd).**

| Chromosome     | Human             |                   |                   | Mouse             |                   |                   | Rat               |                   |                   |
|----------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|
|                | Total             | Intergenic DNA    | Intragenic DNA    | Total             | Intergenic DNA    | Intragenic DNA    | Total             | Intergenic DNA    | Intragenic DNA    |
| 21             | 23.0              | 15.6 <sup>d</sup> | 44.2 <sup>e</sup> |                   |                   |                   |                   |                   |                   |
| 22             | 33.3 <sup>e</sup> | 21.5              | 51.5 <sup>e</sup> |                   |                   |                   |                   |                   |                   |
| X              | 19.4 <sup>d</sup> | 16.9              | 24.7 <sup>d</sup> | 12.6 <sup>d</sup> | 11.1 <sup>d</sup> | 16.4 <sup>d</sup> | 16.0 <sup>d</sup> | 14.5 <sup>d</sup> | 25.5 <sup>d</sup> |
| Y <sup>f</sup> | 9.5 <sup>d</sup>  | 8.1 <sup>d</sup>  | 27.9              | 4.5 <sup>d</sup>  | 3.7 <sup>d</sup>  | 16.4 <sup>d</sup> |                   |                   |                   |
| Mean density   | 24.5              | 19.5              | 32.3              | 26.1              | 20.5              | 34.5              | 34.6              | 29.8              | 46.9              |
| Std. Dev.      | 7.1               | 4.7               | 8.9               | 6.5               | 4.9               | 6.9               | 8.9               | 7.2               | 11.1              |

<sup>a</sup> putative DREs defined as the 19 bp DRE centered core containing sequence with a MS score  $\geq 0.8473$

<sup>b</sup> Mbp = million basepairs

<sup>c</sup> intergenic and intragenic DNA region are defined in Figure 15B

<sup>d</sup> putative DRE density is less than the lower limit of the 99% confidence interval of the mean

<sup>e</sup> putative DRE density is greater than the upper limit of the 99% confidence interval of the mean

<sup>f</sup> no sequence data for chromosome Y is available in rn4 build of the rat genome

chromosome for the same region. In contrast, the putative density in the 5' UTR for chromosome Y was 84.1 per Mbp, nearly double the chromosomal average in the mouse genome. Human chromosome Y was similar with a lower putative DRE density in the intragenic region, but the 5' UTR density was more than 2.6-times greater than the mean chromosomal value. Similar to chromosome Y, the putative densities in the intragenic regions of chromosome X were significantly lower than the mean in each genome. However, unlike chromosome Y, the density in the 5' UTR was also lower than the mean chromosome value. Such region differences in chromosomes X and Y may contribute to sex-specific AhR-mediated responses. Supplementary Tables 2 and 3 provide a complete chromosomal summary of the total number of putative DRE cores in the intergenic and intragenic DNA regions, the UTRs and the CDS for the human, mouse and rat genomes.

### ***RANDOM SEQUENCE ANALYSIS***

To examine the chance occurrence of putative DREs, 25,000 random sequences of 15 kb were generated and searched for DREs. The computational search found 731,636 core sequences and extending these sequences by 7 bp on both ends, identified 108,210 chance occurrences of putative DREs (MS score  $\geq 0.8473$ ). In total, 375 Mbp were searched resulting in 288.6 putative DREs per Mbp. This chance occurrence of putative DRE density is significantly greater than the calculated densities in each genome both at the global and chromosomal level, suggesting that regions with a high density of putative DREs have a greater likelihood of being biologically significant.

### ***PUTATIVE DRE DENSITY PROXIMAL TO THE TSS***

Putative DRE densities across genomes and chromosomes provide a gross estimate of occurrence. Finer analysis of different gene region segments generally found greater putative

Table 11. Chromosomal density of putative DREs<sup>a</sup> (per Mbp<sup>b</sup>) within the 10kb upstream, 5' and 3' UTRs, and CDS regions<sup>c</sup> of RefSeq sequences in the human, mouse and rat genomes.

| Chromosome     | Human             |                   |                   |                   | Mouse             |                   |                   |                   | Rat               |                   |                   |                   |
|----------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|
|                | 10kb upstream     | 5' UTR            | CDS               | 3' UTR            | 10kb upstream     | 5' UTR            | CDS               | 3' UTR            | 10kb upstream     | 5' UTR            | CDS               | 3' UTR            |
| 1              | 35.8 <sup>d</sup> | 27.5              | 27.3              | 22.1 <sup>d</sup> | 39.8              | 30.6 <sup>d</sup> | 29.6              | 32.8              | 43.3 <sup>d</sup> | 52.6              | 41.7              | 34.6 <sup>d</sup> |
| 2              | 62.2 <sup>e</sup> | 40.2              | 27.1              | 26.2              | 38.8              | 41.2              | 34.0              | 36.1              | 49.7              | 34.9 <sup>d</sup> | 37.1 <sup>d</sup> | 36.2 <sup>d</sup> |
| 3              | 51.4              | 18.1 <sup>d</sup> | 22.9 <sup>d</sup> | 24.8              | 41.9              | 42.6              | 31.2              | 40.3              | 48.4              | 33.6 <sup>d</sup> | 44.0              | 61.9              |
| 4              | 57.4 <sup>e</sup> | 29.2              | 28.6              | 27.5              | 49.4 <sup>e</sup> | 46.1              | 37.2              | 43.1              | 42.6 <sup>d</sup> | 36.1 <sup>d</sup> | 38.2              | 42.9 <sup>d</sup> |
| 5              | 54.0              | 43.8              | 28.9              | 30.4              | 49.1 <sup>e</sup> | 58.7 <sup>e</sup> | 38.8 <sup>e</sup> | 34.0              | 53.7              | 53.4              | 43.5              | 50.8              |
| 6              | 41.4              | 28.4              | 26.8              | 23.8              | 35.0 <sup>d</sup> | 37.7              | 32.4              | 38.5              | 49.2              | 54.6              | 44.7              | 39.1 <sup>d</sup> |
| 7              | 49.9              | 32.5              | 27.2              | 30.4              | 36.5              | 41.6              | 39.1 <sup>e</sup> | 44.9              | 60.2 <sup>e</sup> | 49.9              | 49.0              | 51.8              |
| 8              | 52.2              | 27.6              | 27.0 <sup>d</sup> | 29.7              | 48.7 <sup>e</sup> | 33.6 <sup>d</sup> | 39.6 <sup>e</sup> | 50.4 <sup>e</sup> | 50.5              | 49.1              | 45.2              | 50.4              |
| 9              | 46.1              | 28.9              | 30.3              | 35.1 <sup>e</sup> | 40.2              | 42.9              | 36.4              | 44.8              | 47.4              | 51.5              | 41.2              | 66.4 <sup>e</sup> |
| 10             | 50.8              | 29.4              | 23.2              | 24.5              | 40.3              | 33.6 <sup>d</sup> | 34.2              | 40.1              | 54.3              | 66.4 <sup>e</sup> | 51.6              | 48.7              |
| 11             | 34.9 <sup>d</sup> | 37.4              | 27.7              | 24.4              | 48.2 <sup>e</sup> | 52.0 <sup>e</sup> | 39.5 <sup>e</sup> | 47.2 <sup>e</sup> | 48.8              | 33.9 <sup>d</sup> | 44.8              | 77.6 <sup>e</sup> |
| 12             | 42.4              | 27.4              | 26.0 <sup>d</sup> | 16.4 <sup>d</sup> | 42.3              | 38.5              | 36.4              | 50.9 <sup>e</sup> | 77.7              | 69.1 <sup>e</sup> | 84.5 <sup>e</sup> | 61.1              |
| 13             | 84.8 <sup>e</sup> | 47.6 <sup>e</sup> | 31.8              | 48.1 <sup>e</sup> | 34.0 <sup>d</sup> | 36.2              | 30.3              | 37.6              | 42.5 <sup>d</sup> | 54.6              | 42.6              | 55.2              |
| 14             | 56.7 <sup>e</sup> | 32.7              | 26.2 <sup>d</sup> | 30.2              | 40.6              | 40.7              | 31.7              | 40.6              | 42.0 <sup>d</sup> | 55.3              | 42.3              | 67.1 <sup>e</sup> |
| 15             | 47.0              | 27.0 <sup>d</sup> | 34.1              | 27.3              | 48.4 <sup>e</sup> | 41.9              | 39.1 <sup>e</sup> | 41.4              | 46.5              | 42.7 <sup>d</sup> | 39.3              | 73.3 <sup>e</sup> |
| 16             | 51.6              | 41.3              | 49.4 <sup>e</sup> | 47.9 <sup>e</sup> | 39.8              | 40.9              | 29.9              | 41.3              | 65.1              | 63.3 <sup>e</sup> | 43.9              | 65.9 <sup>e</sup> |
| 17             | 33.8 <sup>d</sup> | 60.5 <sup>e</sup> | 31.5              | 18.2 <sup>d</sup> | 46.4 <sup>e</sup> | 46.4              | 41.9 <sup>e</sup> | 39.8              | 46.3              | 53.8              | 41.5              | 37.4 <sup>d</sup> |
| 18             | 76.0 <sup>e</sup> | 30.1              | 35.7              | 35.1 <sup>e</sup> | 40.4              | 38.3              | 37.8              | 37.0              | 53.8              | 48.8              | 39.3              | 25.2 <sup>d</sup> |
| 19             | 25.0 <sup>d</sup> | 50.9 <sup>e</sup> | 52.7 <sup>e</sup> | 27.6              | 43.0              | 51.3 <sup>e</sup> | 37.3              | 38.4              | 55.4              | 48.9              | 52.4              | 58.1              |
| 20             | 49.9              | 33.7              | 37.0 <sup>e</sup> | 20.7 <sup>d</sup> |                   |                   |                   |                   | 64.0 <sup>e</sup> | 58.9 <sup>e</sup> | 73.7 <sup>e</sup> | 78.3 <sup>e</sup> |
| 21             | 41.6              | 19.2 <sup>d</sup> | 32.5              | 21.7 <sup>d</sup> |                   |                   |                   |                   |                   |                   |                   |                   |
| 22             | 42.9              | 51.4 <sup>e</sup> | 41.3 <sup>e</sup> | 28.3              |                   |                   |                   |                   |                   |                   |                   |                   |
| X              | 26.8 <sup>d</sup> | 25.0 <sup>d</sup> | 21.4 <sup>d</sup> | 12.9 <sup>d</sup> | 20.5 <sup>d</sup> | 19.8 <sup>d</sup> | 14.8 <sup>d</sup> | 14.5 <sup>d</sup> | 31.0 <sup>d</sup> | 30.3 <sup>d</sup> | 23.2 <sup>d</sup> | 34.4 <sup>d</sup> |
| Y <sup>f</sup> | 45.2              | 98.9 <sup>e</sup> | 42.6 <sup>e</sup> | 16.7 <sup>d</sup> | 26.7 <sup>d</sup> | 84.1 <sup>e</sup> | 13.3 <sup>d</sup> | 0.0 <sup>d</sup>  |                   |                   |                   |                   |
| Mean density   | 48.3              | 37.0              | 31.6              | 27.1              | 40.5              | 42.8              | 33.5              | 37.8              | 51.1              | 49.6              | 45.9              | 53.2              |
| Std. Dev.      | 13.6              | 16.8              | 8.0               | 8.5               | 7.3               | 12.5              | 7.4               | 11.4              | 9.9               | 10.9              | 12.6              | 15.3              |

<sup>a</sup> putative DREs defined as the 19 bp DRE centered core containing sequence with a MS score  $\geq 0.8473$

<sup>b</sup> Mbp = million basepairs

<sup>c</sup> regions are defined using the genomic locations in the refGene database from the UCSC Genome Browser

<sup>d</sup> putative DRE density is less than the lower limit of the 99% confidence interval of the mean

<sup>e</sup> putative DRE density is greater than the upper limit of the 99% confidence interval of the mean

<sup>f</sup> no sequence data for chromosome Y is available in rn4 build of the rat genome

DRE density in the 10 kb upstream and 5' UTR regions. To further investigate these segments, the number of putative DREs in non-overlapping 100 bp windows spanning the region 10 kb upstream and 5 kb downstream of a TSS were plotted (Figure 17). Putative DREs were not equally distributed within this 15 kb region, with the highest density occurring within  $\pm 1.5$  kb of a TSS. In each species, the density was the greatest at approximately 100 bp directly upstream of the TSS. A sharp 3' drop from the maximum was observed followed by a secondary peak 200-400 bp downstream of the TSS before putative DRE occurrence returned to basal levels.

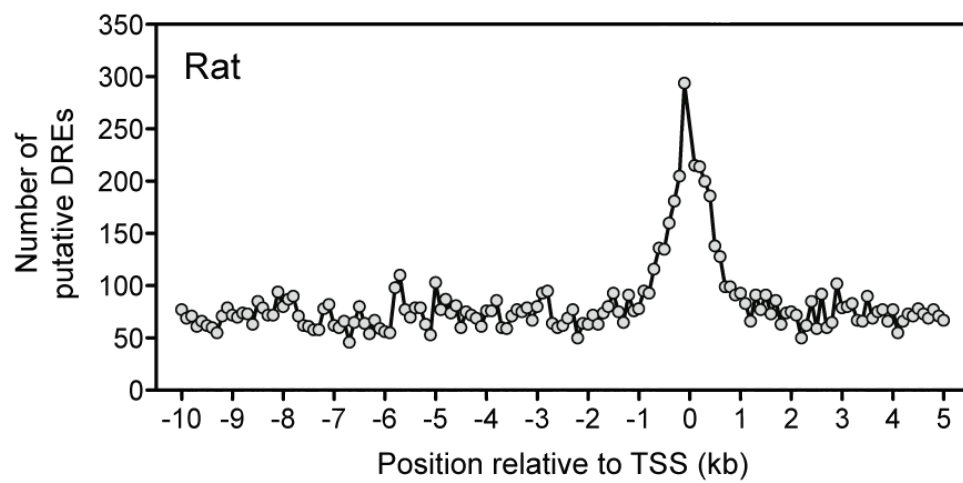
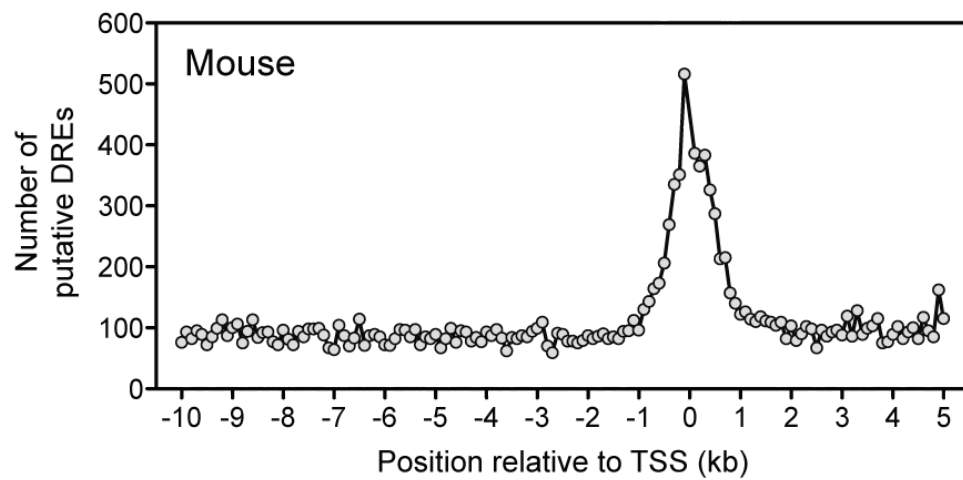
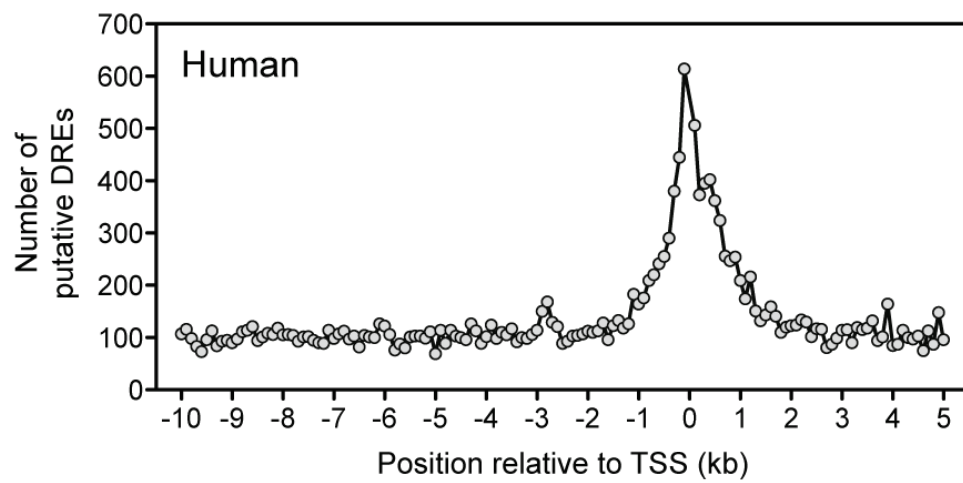
### **GENE LEVEL ANALYSIS OF DRES**

Unique Entrez Gene identifiers for mature gene-encoding RefSeqs (NM prefixed RefSeq identifiers) were obtained from the UCSC Genome Browser refLink database and used to determine the distribution of putative DREs associated with 18,893 human, 20,018 mouse and 15,342 rat annotated genes (Table 12). The majority of all known genes had at least one DRE core present within 10 kb upstream of the TSS and the transcribed gene. However, 55 human, 343 mouse, and 327 rat genes did not have a DRE core within this same region. It is surprising to identify so many genes without a DRE core since the average gene region length (10 kb upstream of a TSS plus the transcribed gene) in the different genomes is 61 kb and the 5'-GCGTG-3' sequence is expected to occur once every 512 bp.

Further restricting this analysis to the 19 bp DRE sequences with a MS score  $\geq 0.8473$  (i.e., putative DREs) identified 69%, 63% and 64% of all human, mouse and rat genes, respectively, had at least one putative DRE (Table 12). Moreover, approximately 60% of all human, mouse and rat genes have 1 to 10 putative DREs (Figure 18). Interestingly, the maximum number of putative DREs was found in human *PTPRN2* with 134, mouse *Wwox* with 73, and rat *Odz2* with 65. Orthologs of these genes also had a high number of putative DREs. For example,

**Figure 17. Distribution of putative DREs in the regions 10 kb upstream to 5 kb downstream of a TSS for all RefSeq sequences.**

The -10 kb to 5 kb region of a TSS were divided into non-overlapping 100 bp windows. The total number of putative DREs (MS score  $\geq 0.8473$ ) were determined for each 100 bp window and graphed. The density of putative DREs was greatest in the 3 kb region centered around the TSS.



**Table 12. Analysis of DRE core and putative DRE containing RefSeq sequences and genes in the human, mouse and rat genomes.**

|                                  | Human   |        | Mouse   |        | Rat     |        |
|----------------------------------|---------|--------|---------|--------|---------|--------|
|                                  | RefSeqs | Genes  | RefSeqs | Genes  | RefSeqs | Genes  |
| Genome <sup>a</sup>              | 28,906  | 18,893 | 24,327  | 20,018 | 15,737  | 15,342 |
| With a DRE core                  | 28,871  | 18,858 | 23,982  | 19,675 | 15,409  | 15,015 |
| With a putative DRE <sup>b</sup> | 20,502  | 13,050 | 15,885  | 12,628 | 10,105  | 9,809  |

<sup>a</sup> based RefSeqs and Entrez Gene IDs stored in the refGene and refLink databases from the UCSC Genome Browser

<sup>b</sup> putative DREs defined as the 19 bp DRE centered core containing sequence with a MS score  $\geq 0.8473$



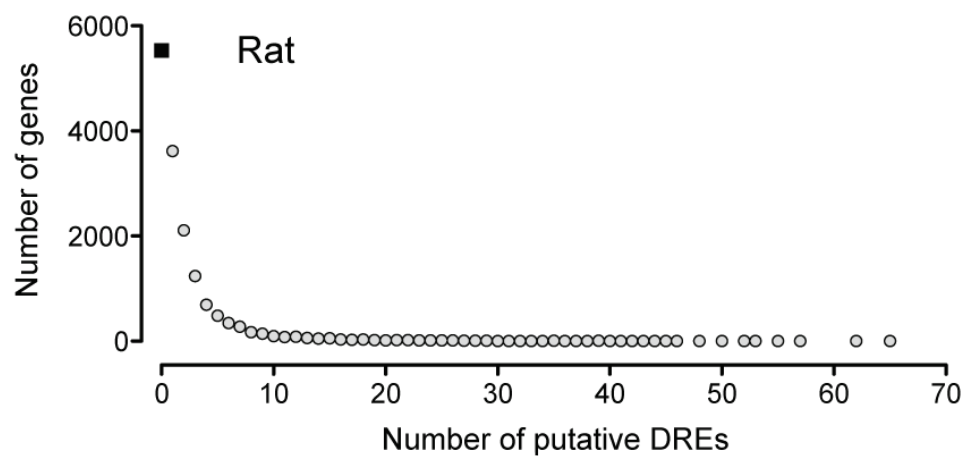
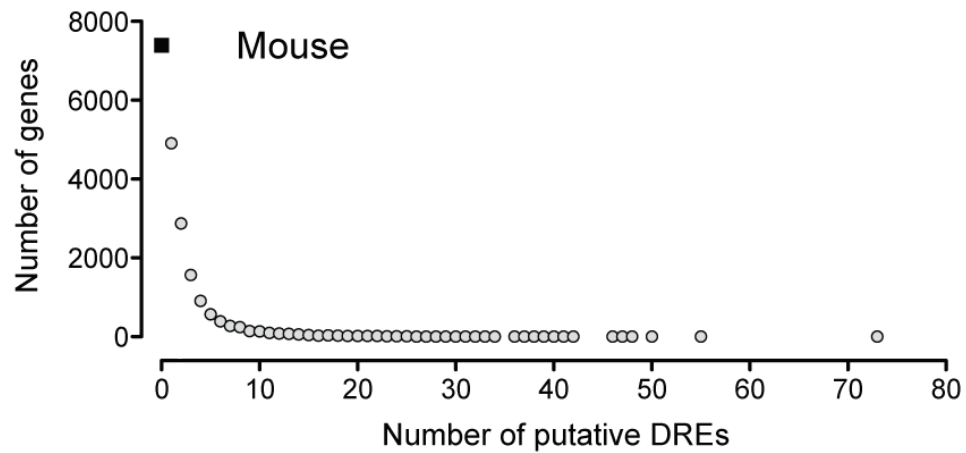
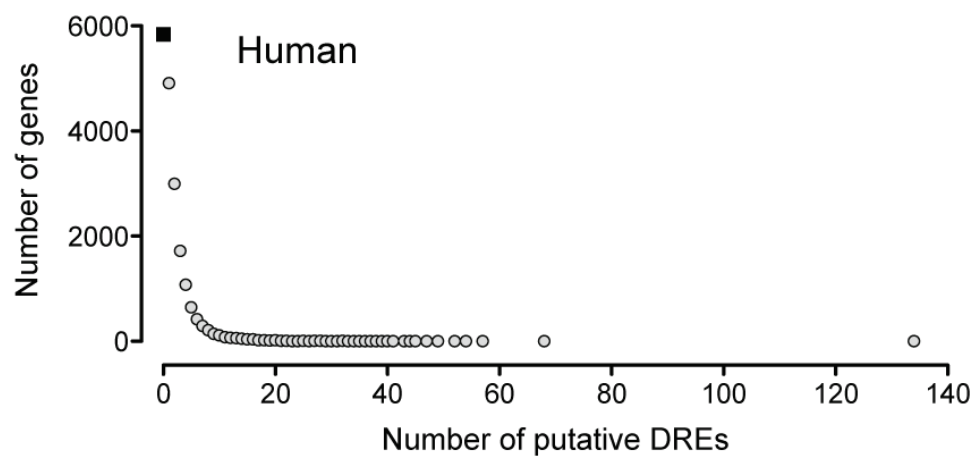
there were 24 and 25 putative DREs in the mouse and rat *PTPRN2*, respectively. However, there is no reported evidence that TCDD regulates the expression of human *PTPRN2*, mouse *Wwox*, rat *Odz2* or their orthologs.

Global hepatic temporal gene expression analysis at 2, 4, 8, 12, 18, 24, 72, and 168 hrs identified 1,337 genes that were differentially expressed ( $|\text{fold change}| \geq 1.5$  and  $P1(t) \geq 0.99$ ) at one or more time points following a single oral dose of 30  $\mu\text{g/kg}$  TCDD at in immature, ovariectomized C57BL/6 mice. Of these, 888 had putative DREs within the 10 kb upstream or transcribed regions (includes 5' and 3' UTR and CDS). Genes that exhibited significant differential expression in the mouse liver included *Fabp12* with 8 putative DREs (23.5-fold induction), and *Cyp1a1* with 7 putative DREs (205-fold induction). The remaining 449 differentially expressed genes, which included unannotated and hypothetical genes, did not have a putative DRE. Examining only well-annotated genes found 373 TCDD responsive genes without a putative DRE within the region 10 kb upstream or transcribed region. The responses of some these genes include the up-regulation of *Chad* (+6.88-fold) and *Olf114* (+9.97-fold), and the repression of *Serpina7* (-7.98-fold). The complete microarray data set is available in Supplementary Table 4. The responses of *Olf114* and *Serpina7* have previously been reported to be AhR-dependent [23-25], however it is unclear if the responses of these genes are directly mediated by the activated AhR complex, or secondary responses.

Differentially regulated genes identified through microarray analysis of TCDD-treated immature, ovariectomized Sprague Dawley rats [26, 27] were also searched for putative DREs. From those studies, 604 genes were responsive ( $|\text{fold change}| \geq 1.5$  and  $P1(t) \geq 0.99$ ) at 2 or more time points and 528 had at least one putative DRE within the 10 kb upstream or transcribed regions. This current mouse microarray study and the previous rat studies covered 5,451

**Figure 18. Frequency of putative DREs within known human, mouse and rat genes.**

For each species, the gene region (10 kb upstream of a TSS through to the end of the 3' UTR) was searched for putative DREs. Approximately 35% of all known genes did not contain a DRE core (black box) while nearly 60% of all genes had between 1 and 10 putative DREs. Approximately 5% of all genes have more than 10 putative DREs.



orthologous genes, and only 52 of those were responsive in both models and possessed at least one putative DRE. These results are consistent with our previous orthologous promoter analysis that demonstrated few human, mouse and rat orthologs had positionally conserved DRE upstream of a TSS [14].

## DISCUSSION

Genome-wide identification of potential *cis*-acting regulatory elements provides important information for elucidating signaling networks. Many computational and traditional *in vitro* approaches have generally focused on relatively few genes and a small segment of a target gene promoter, while neglecting more distal elements, which may also have important roles in gene regulation [14, 28-33]. In order to fully elucidate the signaling transduction of transcription factors, both proximally and distally located response elements need to be identified and characterized.

The structure and function of the AhR as well as its mode of action are highly conserved, with homologs found in nearly all vertebrates. Activation of the AhR by TCDD results in the regulation of target gene expression via the substitution intolerant DRE core sequence, 5'-GCGTG-3'. Previous computational analyses of DRE sequences were limited to the proximal promoter regions (5 kb upstream to 2 kb downstream of a TSS) of known genes in earlier drafts of the human, mouse and rat genomes [14]. This complementary study leverages the availability of higher quality finished human and mouse assemblies [15, 34], as well as the most current build of the rat genome to establish a revised PWM and calculate MS scores for DRE sequences located throughout the human, mouse and rat genomes, including the non-transcribed intragenic DNA regions.

Approximately 60% of the human and mouse genomes consist of stretches of non-transcribed intergenic DNA, while we define the remaining 40% as intragenic regions that include the 10 kb upstream promoter region, the 5' and 3' UTRs and the CDS (Table 9). Despite these length differences, the total number of DRE core sequences and putative DREs were comparable in the intergenic DNA and intragenic regions. The draft assembly of the rat genome consists predominantly of intergenic DNA (73%), reflecting the immaturity of the annotation associated with the rat genome. Consequently, the intergenic DNA bias in the rat resulted in the identification of a greater number of DRE cores and putative DREs in the intergenic DNA regions compared to intragenic DNA. Even within intragenic regions, putative DREs were found in CDS and 3' UTR regions, and not limited to proximal-promoters (Table 11). It has been suggested that the relative location of a bound transcription factor may have different roles in regulating gene expression. For example, the estrogen receptor (ER), p53 and forkhead box protein A1 [1-4], interact with proximal and distal response elements located throughout the genome, including the intergenic DNA. Transcription factor binding at the core promoter is presumed to stabilize the basal transcriptional machinery, while more distal motifs may exert regulation through a looping mechanism or by altering chromatin structure [35-37]. Consequently, a comprehensive map of potential AhR binding sites throughout the genome provides important information for elucidating the AhR gene expression network.

Computational searches identified putative DREs in all genome regions. However, once the size of each region was taken into consideration, the density of putative DREs was found to be highest in the intragenic DNA regions of all three species. Moreover, putative DRE densities varied dramatically across chromosomes with some chromosomes having significantly higher densities (e.g., human chromosome 19, mouse chromosome 5, and rat chromosome 12)

compared to the mean chromosomal density, while others (e.g., human chromosome 13, and chromosome rat 2) were significantly less dense. Interestingly, the sex chromosomes, and especially chromosome Y, the rat genome withstanding, were the least dense in terms of the total putative DREs amongst all the other chromosomes. Putative DRE densities within the 10 kb upstream region, the UTRs and the CDS were also substantially different from the mean chromosomal value for those regions. TCDD elicits sex-specific physiological and gene expression responses in rodents [7, 38, 39]. These differences in sensitivity and physiological responses may be influenced by DREs differentially regulating gene expression on the sex chromosomes. Note that there currently is no sequence information for chromosome Y in the draft of the rat assembly, which will likely be resolved in the next phase of the rat genome sequencing effort [16, 40].

Within human and mouse chromosomes putative DRE densities were highest in the 5' UTR and the region 10 kb upstream of the TSS. In contrast, DRE densities in rat genes were slightly higher in the 3' UTR compared to either the 10kb upstream region or the 5' UTR. However, as already mentioned, the limited annotation of the rat genome may have biased the identification of DREs to the 3' UTR. A more finite analysis of the density around the proximal promoter found the greatest putative DRE density within  $\pm 1.5$  kb of the TSS of known RefSeq sequences for all three species, and the maximum density occurring 100 bp upstream of a TSS. This coincides with 70% of all RNA polymerase II (Pol II) binding [2, 3], suggesting that proximal AhR binding recruits and stabilizes Pol II binding at the TSS. Additionally, due to the GC rich nature of the DRE core sequence, the putative DRE density profile mirrors the CpG island frequency in the proximal promoter region [41]. The methylation status of the putative DRE cores within these CpG islands may have significant roles in mediating the gene expression

responses. However, a recent study investigating the methylation of *bona fide* functional DREs in the human *CYP1A1* promoter did not significantly alter its expression [42].

Searching the region 10 kb upstream of a TSS and the transcribed region for all known genes in the genomes found that approximately 65% of all genes contained at least one putative DRE. However, many responses may be tissue specific and/or secondary, thereby not involving direct interaction with the AhR. This correlates with microarray studies, which report that TCDD elicits the differential expression of only a subset of genes in the entire genome [27, 43-46]. Furthermore, the effects elicited by TCDD are known to be tissue- and cell-specific, therefore, the presence of a putative DRE within the gene region is not sufficient to elicit a transcriptional response. Although our use of a MS score  $\geq 0.8473$  to define a 19 bp DRE sequence as putative is justifiable based on experimental data indicating it is the lowest scoring *bona fide* functional DRE (i.e., rat *Aldh3a1* DRE), recent protein-binding microarray studies indicate that more degenerative transcription factor binding sites may also have important functional roles in regulating gene expression [47, 48].

Additionally, transcription factors can also regulate gene expression through a response element-independent manner by tethering to other proximally bound transcription factors. For example, the progesterone receptor can tether with Sp1, Stat5 and AP1 to regulate genes independently of the progesterone response element sequences [49-51]. It has been demonstrated that the AhR is recruited to estrogen-responsive regions in a gene-specific [52] and DRE-independent manner [53]. Furthermore, studies have shown that the AhR:ARNT heterodimer can regulate target gene expression by interacting with an alternate response element sequence and independent of the DRE core consensus sequence [54, 55]. AhR-mediated gene expression is also dependent on several other factors, which include cell and tissue type, the species and sex of

the organism, and the availability of co-regulators [56]. All of these factors must be taken into context in order to fully comprehend AhR-mediated gene regulation.

Computationally searching the human, mouse and rat genome assemblies has clearly revealed that putative DREs are not randomly distributed. Our detailed genomic map has identified putative DREs in intergenic and intragenic DNA regions. Furthermore, putative DRE distributions vary across specific genome regions. This suggests that the interaction of the AhR with putative DREs in different genomic locations may have differing roles in regulating gene expression. Complementary studies are in progress to investigate AhR complex binding to DREs located in intergenic and intragenic regions.



## REFERENCES

## REFERENCES

1. Carroll JS, Liu XS, Brodsky AS, Li W, Meyer CA, Szary AJ, Eeckhoute J, Shao W, Hestermann EV, Geistlinger TR, et al: **Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1.** *Cell* 2005, **122**:33-43.
2. Carroll JS, Meyer CA, Song J, Li W, Geistlinger TR, Eeckhoute J, Brodsky AS, Keeton EK, Fertuck KC, Hall GF, et al: **Genome-wide analysis of estrogen receptor binding sites.** *Nature Genetics* 2006, **38**:1289-1297.
3. Lin C-Y, Vega VB, Thomsen JS, Zhang T, Kong SL, Xie M, Chiu KP, Lipovich L, Barnett DH, Stossi F, et al: **Whole-genome cartography of estrogen receptor alpha binding sites.** *PLoS Genet* 2007, **3**:e87.
4. Wederell ED, Bilenky M, Cullum R, Thiessen N, Dagpinar M, Delaney A, Varhol R, Zhao Y, Zeng T, Bernier B, et al: **Global analysis of in vivo Foxa2-binding sites in mouse adult liver using massively parallel sequencing.** *Nucleic Acids Res* 2008, **36**:4549-4564.
5. Gu Y, Hogenesch J, Bradfield C: **The PAS superfamily: sensors of environmental and developmental signals.** *Annu Rev Pharmacol Toxicol* 2000, **40**:519-561.
6. Denison MS, Heath-Pagliuso S: **The Ah receptor: a regulator of the biochemical and toxicological actions of structurally diverse chemicals.** *Bulletin of environmental contamination and toxicology* 1998, **61**:557-568.
7. Poland A, Knutson JC: **2,3,7,8-tetrachlorodibenzo-p-dioxin and related halogenated aromatic hydrocarbons: examination of the mechanism of toxicity.** *Annu Rev Pharmacol Toxicol* 1982, **22**:517-554.
8. Hankinson O: **The aryl hydrocarbon receptor complex.** *Annu Rev Pharmacol Toxicol* 1995, **35**:307-340.
9. Swanson H, Chan W, Bradfield C: **DNA binding specificities and pairing rules of the Ah receptor, ARNT, and SIM proteins.** *J Biol Chem* 1995, **270**:26292-26302.
10. Gillesby BE, Stanostefano M, Porter W, Safe S, Wu ZF, Zacharewski TR: **Identification of a motif within the 5' regulatory region of pS2 which is responsible for AP-1 binding and TCDD-mediated suppression.** *Biochemistry* 1997, **36**:6080-6089.
11. Lusska A, Shen E, Whitlock JP: **Protein-DNA interactions at a dioxin-responsive enhancer. Analysis of six bona fide DNA-binding sites for the liganded Ah receptor.** *J Biol Chem* 1993, **268**:6575-6580.

12. Shen ES, Whitlock JP: **Protein-DNA interactions at a dioxin-responsive enhancer. Mutational analysis of the DNA-binding site for the liganded Ah receptor.** *J Biol Chem* 1992, **267**:6815-6819.
13. Quandt K, Frech K, Karas H, Wingender E, Werner T: **MatInd and MatInspector: new fast and versatile tools for detection of consensus matches in nucleotide sequence data.** *Nucleic Acids Res* 1995, **23**:4878-4884.
14. Sun YV, Boverhof DR, Burgoon LD, Fielden MR, Zacharewski TR: **Comparative analysis of dioxin response elements in human, mouse and rat genomic sequences.** *Nucleic Acids Res* 2004, **32**:4512-4523.
15. Church DM, Goodstadt L, Hillier LW, Zody MC, Goldstein S, She X, Bult CJ, Agarwala R, Cherry JL, DiCuccio M, et al: **Lineage-specific biology revealed by a finished genome assembly of the mouse.** *PLoS Biol* 2009, **7**:e1000112.
16. Worley KC, Weinstock GM, Gibbs RA: **Rats in the genomic era.** *Physiol Genomics* 2008, **32**:273-282.
17. Zody MC, Jiang Z, Fung H-C, Antonacci F, Hillier LW, Cardone MF, Graves TA, Kidd JM, Cheng Z, Abouelleil A, et al: **Evolutionary toggling of the MAPT 17q21.31 inversion region.** *Nature Genetics* 2008, **40**:1076-1083.
18. Reisdorph R, Lindahl R: **Constitutive and 3-methylcholanthrene-induced rat ALDH3A1 expression is mediated by multiple xenobiotic response elements.** *Drug Metab Dispos* 2007, **35**:386-393.
19. Rhead B, Karolchik D, Kuhn R, Hinrichs A, Zweig A, Fujita P, Diekhans M, Smith K, Rosenbloom K, Raney B, et al: **The UCSC Genome Browser database: update 2010.** *Nucleic Acids Research* 2010, **38**:D613.
20. Boverhof DR, Burgoon LD, Tashiro C, Chittim B, Harkema JR, Jump DB, Zacharewski TR: **Temporal and dose-dependent hepatic gene expression patterns in mice provide new insights into TCDD-Mediated hepatotoxicity.** *Toxicol Sci* 2005, **85**:1048-1063.
21. Eckel J, Gennings C, Therneau T, Burgoon L, Boverhof D, Zacharewski T: **Normalization of two-channel microarray experiments: a semiparametric approach.** *Bioinformatics* 2005, **21**:1078-1083.
22. Eckel J, Gennings C, Chinchilli V, Burgoon L, Zacharewski T: **Empirical bayes gene screening tool for time-course or dose-response microarray data.** *J Biopharm Stat* 2004, **14**:647-670.
23. Tijet N, Boutros PC, Moffat ID, Okey AB, Tuomisto J, Pohjanvirta R: **Aryl hydrocarbon receptor regulates distinct dioxin-dependent and dioxin-independent gene batteries.** *Mol Pharmacol* 2006, **69**:140-153.

24. Ovando BJ, Vezina CM, McGarrigle BP, Olson JR: **Hepatic gene downregulation following acute and subchronic exposure to 2,3,7,8-tetrachlorodibenzo-p-dioxin.** *Toxicol Sci* 2006, **94**:428-438.
25. Yauk CL, Jackson K, Malowany M, Williams A: **Lack of change in microRNA expression in adult mouse liver following treatment with benzo(a)pyrene despite robust mRNA transcriptional response.** *Mutation research* 2010.
26. Boverhof DR, Burgoon LD, Tashiro C, Sharratt B, Chittim B, Harkema JR, Mendrick DL, Zacharewski TR: **Comparative toxicogenomic analysis of the hepatotoxic effects of TCDD in Sprague Dawley rats and C57BL/6 mice.** *Toxicol Sci* 2006, **94**:398-416.
27. Fletcher N, Wahlström D, Lundberg R, Nilsson CB, Nilsson KC, Stockling K, Hellmold H, Håkansson H: **2,3,7,8-Tetrachlorodibenzo-p-dioxin (TCDD) alters the mRNA expression of critical genes associated with cholesterol metabolism, bile acid biosynthesis, and bile transport in rat liver: a microarray study.** *Toxicol Appl Pharmacol* 2005, **207**:1-24.
28. Bourdeau V, Deschênes J, Métivier R, Nagai Y, Nguyen D, Bretschneider N, Gannon F, White JH, Mader S: **Genome-wide identification of high-affinity estrogen response elements in human and mouse.** *Mol Endocrinol* 2004, **18**:1411-1427.
29. Lemay DG, Hwang DH: **Genome-wide identification of peroxisome proliferator response elements using integrated computational genomics.** *J Lipid Res* 2006, **47**:1583-1587.
30. Menendez D, Inga A, Resnick MA: **Estrogen receptor acting in cis enhances WT and mutant p53 transactivation at canonical and noncanonical p53 target sequences.** *Proc Natl Acad Sci USA* 2010, **107**:1500-1505.
31. Nukaya M, Moran S, Bradfield CA: **The role of the dioxin-responsive element cluster between the Cyp1a1 and Cyp1a2 loci in aryl hydrocarbon receptor biology.** *Proc Natl Acad Sci USA* 2009, **106**:4923-4928.
32. Ortiz-Barahona A, Villar D, Pescador N, Amigo J, Del Peso L: **Genome-wide identification of hypoxia-inducible factor binding sites and target genes by a probabilistic model integrating transcription-profiling data and in silico binding site prediction.** *Nucleic acids research* 2010.
33. van Batenburg MF, Li H, Polman JA, Lachize S, Datson NA, Bussemaker HJ, Meijer OC: **Paired hormone response elements predict caveolin-1 as a glucocorticoid target gene.** *PLoS ONE* 2010, **5**:e8839.
34. Xue Y, Sun D, Daly A, Yang F, Zhou X, Zhao M, Huang N, Zerjal T, Lee C, Carter NP, et al: **Adaptive evolution of UGT2B17 copy-number variation.** *Am J Hum Genet* 2008, **83**:337-346.

35. Farnham PJ: **Insights from genomic profiling of transcription factors.** *Nat Rev Genet* 2009, **10**:605-616.
36. Li Q, Barkess G, Qian H: **Chromatin looping and the probability of transcription.** *Trends Genet* 2006, **22**:197-202.
37. Long X, Miano JM: **Remote control of gene expression.** *J Biol Chem* 2007, **282**:15941-15945.
38. Silkworth JB, Carlson EA, McCulloch C, Illouz K, Goodwin S, Sutter TR: **Toxicogenomic analysis of gender, chemical, and dose effects in livers of TCDD- or aroclor 1254-exposed rats using a multifactor linear model.** *Toxicol Sci* 2008, **102**:291-309.
39. Walker NJ, Wyde ME, Fischer LJ, Nyska A, Bucher JR: **Comparison of chronic toxicity and carcinogenicity of 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD) in 2-year bioassays in female Sprague-Dawley rats.** *Mol Nutr Food Res* 2006, **50**:934-944.
40. Twigger SN, Pruitt KD, Fernández-Suárez XM, Karolchik D, Worley KC, Maglott DR, Brown G, Weinstock G, Gibbs RA, Kent J, et al: **What everybody should know about the rat genome and its online resources.** *Nat Genet* 2008, **40**:523-527.
41. Saxonov S, Berg P, Brutlag DL: **A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters.** *Proc Natl Acad Sci USA* 2006, **103**:1412-1417.
42. Nakajima M, Iwanari M, Yokoi T: **Effects of histone deacetylation and DNA methylation on the constitutive and TCDD-inducible expressions of the human CYP1 family in MCF-7 and HeLa cells.** *Toxicol Lett* 2003, **144**:247-256.
43. Boutros PC, Yan R, Moffat ID, Pohjanvirta R, Okey AB: **Transcriptomic responses to 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD) in liver: comparison of rat and mouse.** *BMC Genomics* 2008, **9**:419.
44. Carlson EA, McCulloch C, Koganti A, Goodwin SB, Sutter TR, Silkworth JB: **Divergent transcriptomic responses to aryl hydrocarbon receptor agonists between rat and human primary hepatocytes.** *Toxicol Sci* 2009, **112**:257-272.
45. Henry EC, Welle SL, Gasiewicz TA: **TCDD and a putative endogenous AhR ligand, ITE, elicit the same immediate changes in gene expression in mouse lung fibroblasts.** *Toxicol Sci* 2010, **114**:90-100.
46. Kim S, Dere E, Burgoon LD, Chang C-C, Zacharewski TR: **Comparative analysis of AhR-mediated TCDD-elicited gene expression in human liver adult stem cells.** *Toxicol Sci* 2009, **112**:229-244.

47. Badis G, Berger MF, Philippakis AA, Talukder S, Gehrke AR, Jaeger SA, Chan ET, Metzler G, Vedenko A, Chen X, et al: **Diversity and complexity in DNA recognition by transcription factors.** *Science* 2009, **324**:1720-1723.
48. Jaeger SA, Chan ET, Berger MF, Stottmann R, Hughes TR, Bulyk ML: **Conservation and regulatory associations of a wide affinity range of mouse transcription factor binding sites.** *Genomics* 2010.
49. Cicatiello L, Addeo R, Sasso A, Altucci L, Petrizzi VB, Borgo R, Cancemi M, Caporali S, Caristi S, Scafoglio C, et al: **Estrogens and progesterone promote persistent CCND1 gene activation during G1 by inducing transcriptional derepression via c-Jun/c-Fos/estrogen receptor (progesterone receptor) complex assembly to a distal regulatory element and recruitment of cyclin D1 to its own gene promoter.** *Mol Cell Biol* 2004, **24**:7260-7274.
50. Owen GI, Richer JK, Tung L, Takimoto G, Horwitz KB: **Progesterone regulates transcription of the p21(WAF1) cyclin- dependent kinase inhibitor gene through Sp1 and CBP/p300.** *J Biol Chem* 1998, **273**:10696-10701.
51. Stoecklin E, Wissler M, Schaetzle D, Pfitzner E, Groner B: **Interactions in the transcriptional regulation exerted by Stat5 and by members of the steroid hormone receptor family.** *J Steroid Biochem Mol Biol* 1999, **69**:195-204.
52. Ahmed S, Valen E, Sandelin A, Matthews J: **Dioxin increases the interaction between aryl hydrocarbon receptor and estrogen receptor alpha at human promoters.** *Toxicol Sci* 2009, **111**:254-266.
53. Beischlag TV, Perdew GH: **ER alpha-AHR-ARNT protein-protein interactions mediate estradiol-dependent transrepression of dioxin-inducible gene transcription.** *J Biol Chem* 2005, **280**:21607-21611.
54. Boutros PC, Moffat ID, Franc MA, Tijet N, Tuomisto J, Pohjanvirta R, Okey AB: **Dioxin-responsive AHRE-II gene battery: identification by phylogenetic footprinting.** *Biochem Biophys Res Commun* 2004, **321**:707-715.
55. Sogawa K, Numayama-Tsuruta K, Takahashi T, Matsushita N, Miura C, Nikawa J-i, Gotoh O, Kikuchi Y, Fujii-Kuriyama Y: **A novel induction mechanism of the rat CYP1A2 gene mediated by Ah receptor-Arnt heterodimer.** *Biochem Biophys Res Commun* 2004, **318**:746-755.
56. Odom DT, Dowell RD, Jacobsen ES, Gordon W, Danford TW, MacIsaac KD, Rolfe PA, Conboy CM, Gifford DK, Fraenkel E: **Tissue-specific transcriptional regulation has diverged significantly between human and mouse.** *Nature Genetics* 2007, **39**:730-732.

## CHAPTER 6

Dere E, Lo R, Celius T, Matthews J, and Zacharewski TR: **Integration of Genome-Wide Computational DRE Search, AhR ChIP-chip and Gene Expression Analyses of TCDD-Elicited Responses in the Mouse Liver.** *BMC Genomics* (Submitted September 21, 2010)

## CHAPTER 6

# INTEGRATION OF GENOME-WIDE COMPUTATIONAL DRE SEARCH, AHR CHIP-CHIP AND GENE EXPRESSION ANALYSES OF TCDD- ELICITED RESPONSES IN THE MOUSE LIVER

### ABSTRACT

The aryl hydrocarbon receptor (AhR) is a ligand-activated transcription factor (TF) that mediates responses to 2,3,7,8-tetrachlorodibenzo-*p*-dioxin (TCDD) via binding to dioxin response elements (DREs). Global genomic chromatin immunoprecipitation (ChIP-chip) and gene expression analyses were performed on hepatic tissue from immature ovariectomized mice orally gavaged with 30µg/kg TCDD. ChIP-chip analysis identified 14,446 and 974 AhR enriched regions (1% false discovery rate) at 2 and 24 hrs, respectively. Enrichment density was greatest in the proximal promoter region, and more specifically, within  $\pm 1.5$  kb of a transcriptional start site (TSS). AhR enrichments also occurred in regions distal to a TSS (e.g., intergenic DNA and 3' UTR), which may impart important regulatory effects. TF binding site analyses identified over-represented DRE sequences within enriched regions. However, 57.8% and 48.5% of all AhR enriched regions at 2 and 24 hrs, respectively, lacked the DRE core. Integration of ChIP-chip, DRE and microarray analyses identified 625 responsive genes ( $|\text{fold change}| \geq 1.5$ ,  $P(t) > 0.999$ ) mediated by AhR interaction at a DRE. Differentially regulated genes with AhR enrichment were functionally associated with fatty acid and lipid metabolism and transport, and xenobiotic metabolism. The data has expanded the AhR regulatory network by including intragenic and intergenic genomic AhR-DNA interactions. Moreover, the AhR can interact with



DNA independent of a DRE core suggesting there are alternative mechanisms of AhR-mediated gene regulation.

## INTRODUCTION

The aryl hydrocarbon receptor (AhR) is a ligand activated transcription factor (TF) belonging to the basic-helix-loop-helix-PAS (bHLH-PAS) family of proteins that serve as environmental sensors [1]. 2,3,7,8-Tetrachlorodibenzo-*p*-dioxin (TCDD) is the prototypical AhR ligand, a ubiquitous environmental contaminant that elicits diverse species-specific effects, including tumor promotion, teratogenesis, hepatotoxicity, modulation of endocrine systems, immunotoxicity and enzyme induction [2, 3]. These effects result from alterations in gene expression mediated by the AhR [4]. Several studies have demonstrated the requirement for the AhR in mediating TCDD-elicited responses. For example, mice carrying low-affinity AhR alleles are less susceptible to the effects elicited by TCDD [5]. Additionally, AhR-null mice fail to induce responses typically observed following treatment with TCDD and related compounds [6].

TCDD binding to the cytosolic AhR results in a conformational change and translocation to the nucleus. The activated AhR complex heterodimerizes with the aryl hydrocarbon nuclear translocator (ARNT), another bHLH-PAS family member, and binds dioxin response elements (DREs) containing the substitution intolerant 5'-GCGTG-3' core sequence to regulate changes in gene expression [4, 7]. Computational searches for all DRE cores in the human, mouse and rat genome identified the highest density of DREs proximal to a transcriptional start site (TSS) (Dere *et al.*, in submission). However, a significant number of DRE cores and putative functional DREs have been identified in distal regions within non-coding intergenic segments of the genome. It has been proposed that enrichments for other TFs on outlying regions may be

functionally relevant through tertiary looping of genomic DNA and/or via protein tethering mechanisms [8].

The role of specific transcriptional regulators has been studied on a gene-by-gene basis, primarily focusing on regions proximal to the TSS. However, the coupling of chromatin immunoprecipitation with either genomic tiling microarrays (ChIP-chip) or next-generation sequencing (ChIP-seq) has facilitated genome-wide analysis of protein-DNA interactions for a variety of receptors [9-15], TFs [16-19] and components of the basal transcriptional machinery [9, 20, 21]. Genome-wide location analyses further suggest that TF binding at *cis*-regulatory enhancers in intergenic DNA regions of the genome may also have functional significance [9, 16, 22, 23].

Several studies have investigated AhR-mediated gene expression responses using various technologies [24-29]. Although AhR-DNA interactions have primarily focused on the regulation of *CYP1A1* [4, 30], recent global ChIP studies have extended our knowledge of AhR-DNA interactions by examining promoter region binding profiles using *in vitro* and *in vivo* models [31-34] (Lo *et al.*, in preparation). Our study provides a comprehensive analysis by examining TCDD-induced AhR binding across the entire mouse genome. In addition, we examined AhR binding within chromosomes, intragenic and intergenic DNA regions, and in specific genic regions (i.e., 10 kb upstream of a TSS, 5' and 3' untranslated regions [UTRs], coding sequence [CDS]). Global AhR enrichment data are also integrated with computational DRE core analysis (Dere *et al.*, in submission), and complementary whole-genome gene expression profiling to provide a more comprehensive evaluation of the hepatic AhR regulatory network elicited by TCDD.

## **MATERIALS AND METHODS**

### ***ANIMAL HANDLING AND TREATMENT***

Hepatic tissue samples from immature female ovariectomized C57BL/6 mice obtained from a previous study [35] were used for both ChIP assays at 2 and 24 hrs, and gene expression analyses across all time points. Briefly, mice were orally gavaged with 30 µg/kg TCDD and sacrificed by cervical dislocation at 2, 4, 8, 12, 18, 24, 72 or 168 hrs post-dose. Tissues were removed, weighed, and multiple samples (~100 mg each) were flash frozen in liquid nitrogen and stored at -80°C until further use.

### ***CHROMATIN IMMUNOPRECIPITATION (CHIP) AND CHIP-CHIP EXPERIMENTS***

ChIP assays were performed as previously described [32] with the following changes. Approximately 100 mg of mouse liver was homogenized in 1% formaldehyde and incubated for 10 min at room temperature. Tissue homogenate was centrifuged at 10,000 RPM for 3 min at 4°C. Pellet was washed in ice-cold PBS, centrifuged, and resuspended in 900 µL of TSEI (20 mM Tris-HCl [pH 8.0], 150 mM NaCl, 2 mM EDTA, 1% Triton X-100, 0.1% sodium dodecyl sulfate) + 1× Protease Inhibitor Cocktail (Sigma, St. Louis, MO). Samples were sonicated 12 times for 10 s each time at 25% amplitude using a Branson 450 sonifier. Supernatant was transferred to fresh microcentrifuge tubes and incubated with rabbit IgG (5 µg; Sigma) and anti-AhR (5 µg; SA-210, Biomol) overnight at 4°C under gentle agitation. ChIP samples were washed and the DNA was isolated as previously described [32]. For ChIP-chip experiments, immunoprecipitated DNA isolated following immunoprecipitation with anti-AhR of liver extracts from TCDD-treated mice was linearly amplified using a whole genome amplification kit according to the manufacturer's instructions (Sigma). Linearly amplified DNA (7.5 µg) was fragmented by limited DNaseI digestion and hybridized to Affymetrix GeneChip® mouse 2.0R

tiling arrays (Affymetrix, Santa Clara, CA) as previously described [32]. The hybridization and washing steps were performed according to the manufacturer's protocol at the Centre for Applied Genomics (Toronto, Canada). Data were normalized and analyzed using CisGenome and mapped against mouse genome version mm9 [36]. Enriched regions with a false discovery rate (FDR) of 1.0% (0.01) were determined by comparing triplicate samples of AhR<sub>TCDD</sub> to triplicate IgG<sub>TCDD</sub> using a moving average (MA) approach with default settings in TileMap v2 [37]. Regions were merged if the gap between them was <300 bp and the number of probes failing to reach the cut-off was <5. Regions were discarded if they were <120 bp or did not contain at least 5 continuous probes above the cut-off. ChIPed DNA was purified using the PCR purification kit from BioBasic Inc. (Markham, ON) and quantified using quantitative real-time PCR (QRT-PCR) (KAPA SYBR Fast qPCR Master Mix; KAPA Biosystems, Toronto, ON) (ChIP-PCR). Fold enrichment values were calculated relative to IgG controls. ChIP-PCR primer sequences are provided in Supplementary Table 1.

### **CHIP-CHIP LOCATION ANALYSIS**

The mouse genomic assembly (mm9) and associated annotation within the refGene and refLink databases were downloaded from the UCSC Genome Browser [38]. Individual segments of a gene region (i.e. the 10 kb sequence upstream of a TSS, the 5' and 3' UTRs and the CDS) for each mature gene encoding reference sequence (RefSeqs with NM prefixed identifiers) were determined using the genomic coordinates within the refGene databases (Supplementary Figure 1). Intragenic DNA regions within the genomes were computationally identified by merging overlapping gene regions (Supplementary Figure 1) from both strands of the genome, and the DNA between adjacent intragenic regions are defined as the non-transcribed intergenic DNA regions (Supplementary Figure 1). AhR enrichment densities were calculated based on the

number of significant enriched regions occurring in an interrogated region (e.g., intergenic DNA region or 5' UTR) divided by the total sum of the region length. Gene annotation associated with each RefSeq sequence was derived from the refLink database in the UCSC Genome Browser.

### **TRANSCRIPTION FACTOR MOTIF ANALYSIS**

The locations of AhR enrichment were compared against the location of the 5'-GCGTG-3' DRE core sequence in the mouse genome (Dere *et al.*, in submission). Analysis of TF motifs over-represented in conjunction with the DRE core were performed using RegionMiner, a program within the Genomatix suite of applications. Identified module families and individual matrices with z-scores > 3 were considered significant. *De novo* motif discovery was performed using the Gibbs motif sampler in CisGenome on AhR regions of enrichment sequences not containing a DRE. Matrices for over-represented motifs were compared to existing TF binding motifs in JASPAR and TRANSFAC [39, 40] using STAMP [41].

### **COMPARISON WITH MICROARRAY GENE EXPRESSION**

Results from the ChIP-chip and DRE analysis were integrated with whole-genome gene expression profiling data from mice orally gavaged with 30 µg/kg TCDD using 4 × 44k whole-genome oligonucleotide arrays from Agilent Technologies (Santa Clara, CA) (Dere *et al.*, in submission). The genomic locations of the differentially responsive genes ( $|\text{fold change}| \geq 1.5$  and  $P1(t) > 0.999$ ) were obtained for each RefSeq sequence associated with the gene from the refGene database in the UCSC Genome Browser. Circos plots [42] were generated to visualize the locations of DRE cores, regions of AhR enrichment and temporal heatmaps of temporal gene expression responses.

## RESULTS

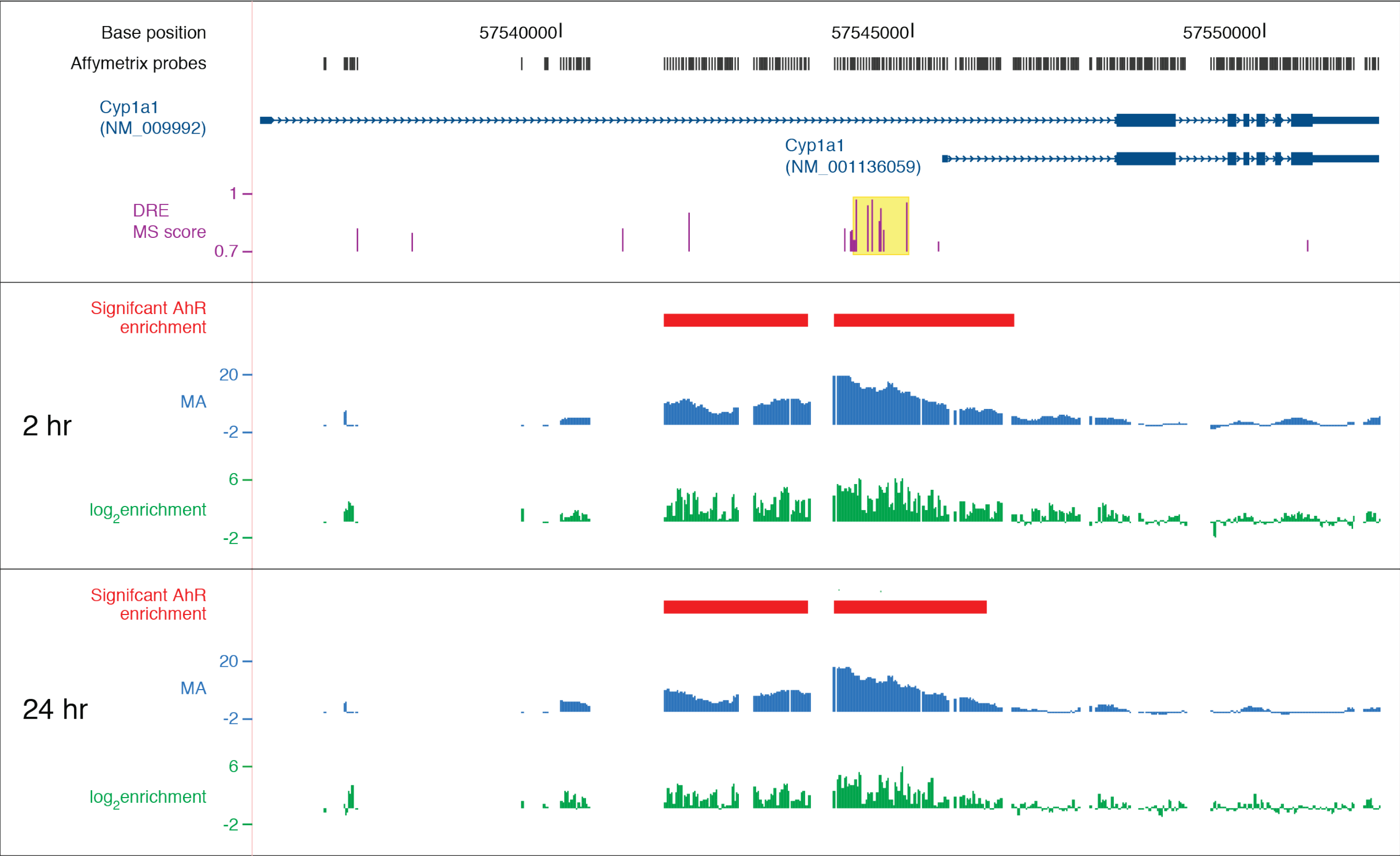
### ***IDENTIFICATION AND CHARACTERIZATION OF TCDD-ELICITED AHR ENRICHMENT***

In order to identify regions of AhR enrichment induced by TCDD across the genome, ChIP-chip assays were performed using hepatic tissue from immature ovariectomized mice orally gavaged with 30 µg/kg TCDD for 2 and 24 hrs. CisGenome [36] analysis identified 22,502 and 12,677 enriched regions at 2 and 24 hrs, respectively. Applying a conservative FDR of 0.01 resulted in 14,446 and 974 significant AhR enriched regions at 2 and 24 hrs, respectively (Supplementary Tables 2 and 3 provides a complete list of enriched regions). Fewer enriched regions at 24 hrs is consistent with the rapid degradation of the AhR within hours following TCDD treatment [43]. The distribution, location and enrichment values for each tiled probes across the *Cyp1a1* gene (represented by RefSeq sequences NM\_009992 and NM\_001136059) are summarized in Figure 19. MA value plots visualize the profile of the enriched region and log<sub>2</sub> fold-enrichment values for each probe are also illustrated (Figure 19). Note that the probes are unevenly tiled throughout the genome, resulting in gaps in genome coverage that may coincide with DRE core locations that may affect AhR enriched region identification. For example, two enriched regions were associated with *Cyp1a1* (Figure 19, red bars). However, the MA plots for 2 and 24 hrs suggest that there is only one large region of enrichment divided into two as a result of the uneven tiling. Consequently, uneven tiling and the lack of tiling in regions that contain DREs may affect the estimated number of AhR enriched regions.

Genomic regions with significant AhR enrichment were mapped to intragenic (10 kb upstream of a TSS plus the transcribed gene of mature RefSeq sequences) and non-coding intergenic regions (Table 13; Supplementary Figure 1). Most regions were enriched 5.7-fold with

**Figure 19. Summary of AhR enrichment within *Cyp1a1* genic region at 2 and 24 hrs.**

*Cyp1a1* is represented by two RefSeq sequences (NM\_009992 and NM\_001136059, dark blue tracks) that have different TSSs (dark blue box at far left). The rectangles and lines represent exons and introns, respectively, and the UTRs are depicted as the thinner rectangles. Arrowhead direction indicates the orientation of the gene. The grey boxes above represent the Affymetrix 2.0R mouse tiling array probe locations across the *Cyp1a1* genic regions. The location and matrix similarity (MS) scores of the consensus DREs are represented by the purple histogram. The highlighted yellow box identifies bona fide functional DREs (matrix similarity (MS) score  $\geq 0.8473$ ) involved in AhR-mediated *Cyp1a1* gene expression. The red boxes identify regions of significant AhR enrichments (FDR < 0.01) based on the moving average (MA) profile by TileMap. The green histogram plots the log<sub>2</sub> fold enrichment values for each individual probe.





**Table 13. Distribution and density analysis of TCDD-induced AhR enriched regions<sup>a</sup> in the mouse genome.**

|       |                                 | Genome | Intergenic<br>DNA <sup>b</sup> | Intragenic<br>DNA <sup>b</sup> | Genic Region <sup>c</sup> |        |       |        |
|-------|---------------------------------|--------|--------------------------------|--------------------------------|---------------------------|--------|-------|--------|
|       |                                 |        |                                |                                | 10kb upstream             | 5' UTR | CDS   | 3' UTR |
| 2 hr  | AhR enrichment                  | 14,446 | 4,163                          | 10,283                         | 4,601                     | 2,569  | 7,499 | 225    |
|       | Enrichment density <sup>d</sup> | 5.44   | 2.62                           | 9.64                           | 18.65                     | 17.29  | 7.21  | 7.29   |
| 24 hr | AhR enrichment                  | 974    | 344                            | 630                            | 306                       | 132    | 507   | 9      |
|       | Enrichment density <sup>d</sup> | 0.37   | 0.22                           | 0.59                           | 1.24                      | 0.89   | 0.49  | 0.29   |

<sup>a</sup> AhR enriched regions with a FDR < 0.01

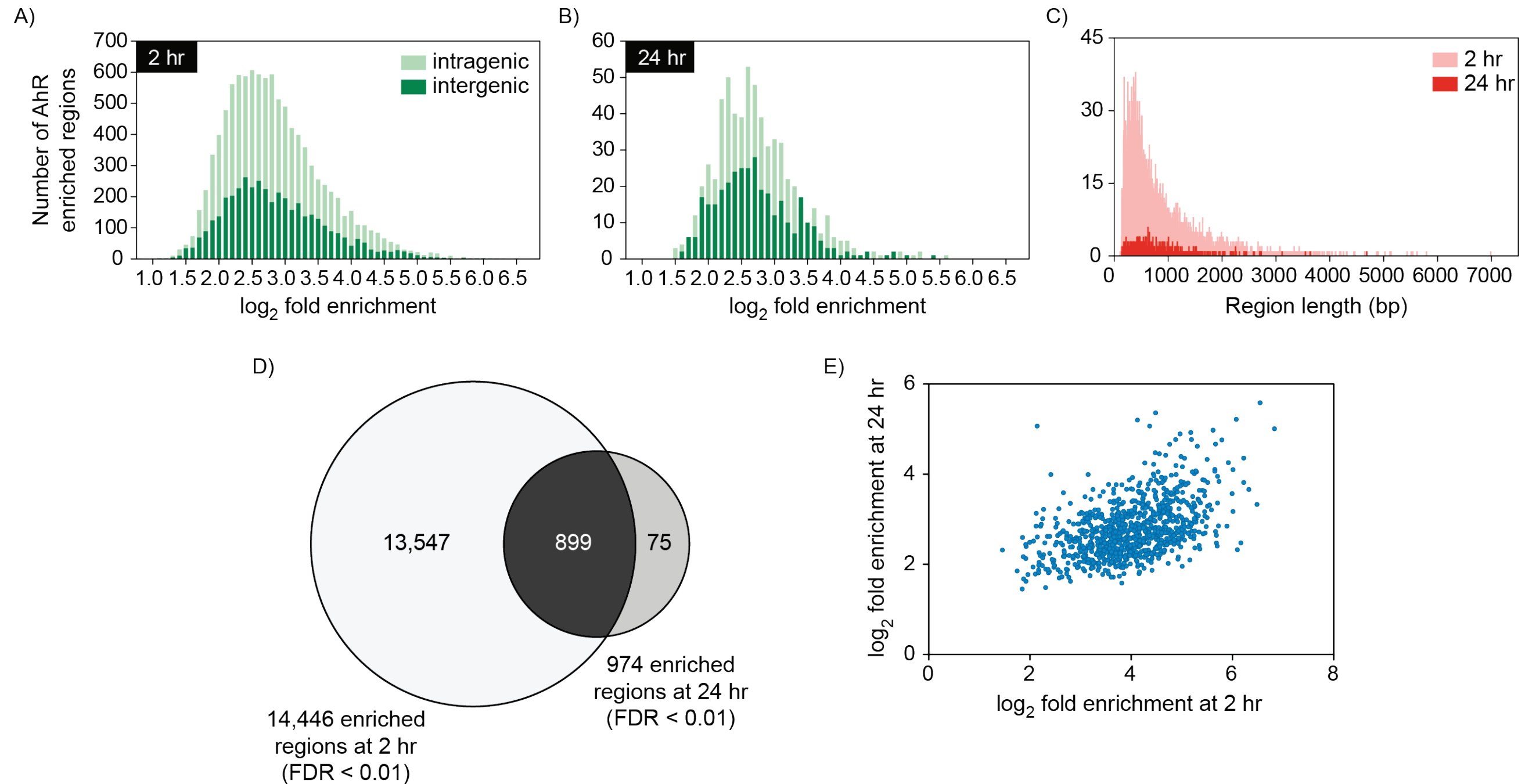
<sup>b</sup> intergenic, intragenic and gene regions are defined as described in Supplementary Figure 1

<sup>c</sup> regions are defined using the genomic locations in the refGene database from the UCSC Genome Browser

<sup>d</sup> density calculated per million base pairs

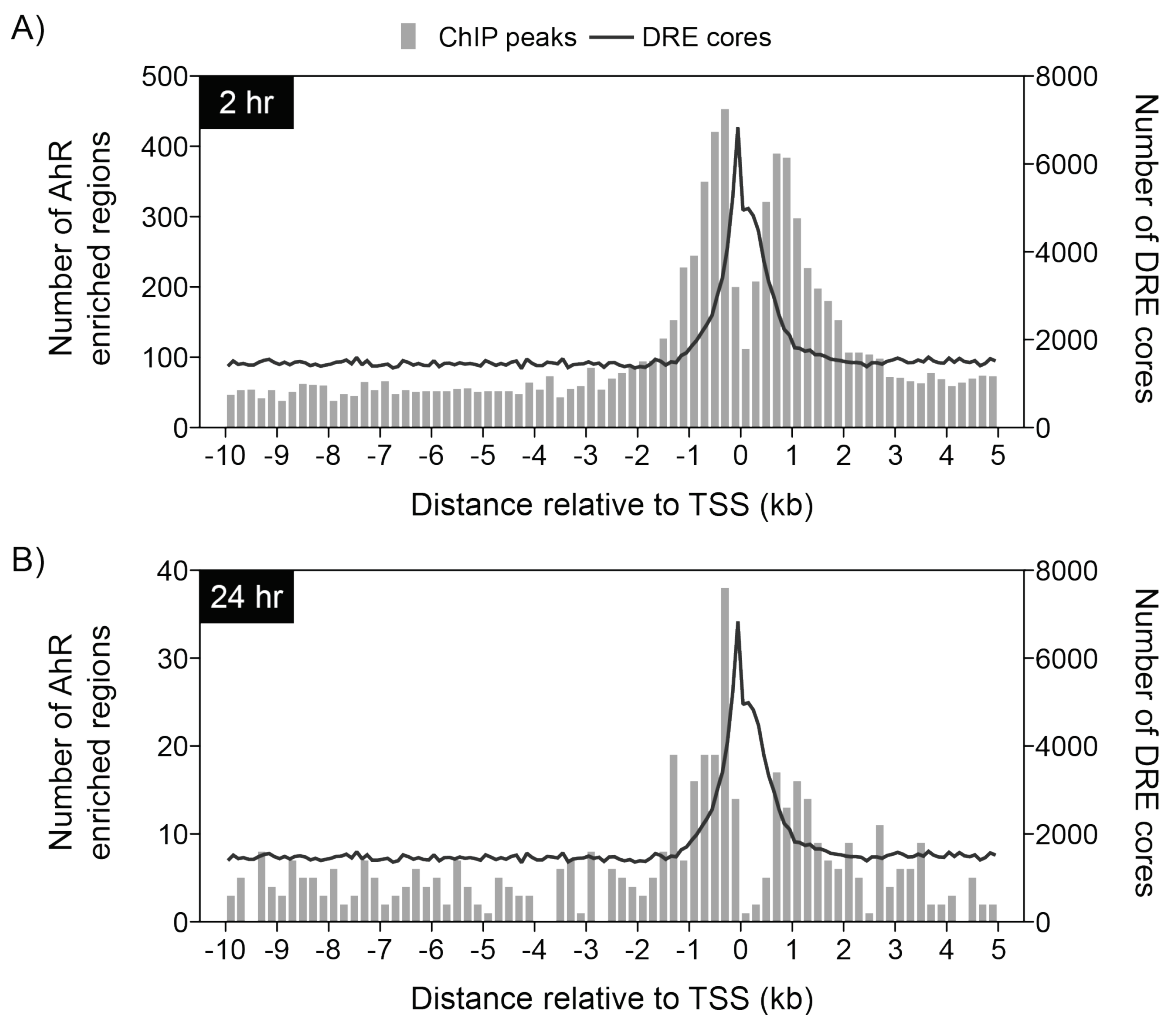
values ranging from 1.7- to 111.4-fold (Figures 20A-B). Enriched regions varied in width from 108 to 6,990 bp (Figure 20C) with 90.5% spanning  $\leq 1,500$  bp. There was no correlation between fold enrichment and region width (data not shown). Of the 974 significantly enriched regions at 24 hr, 899 of them overlapped with a 2 hr enriched region (Figure 20D), consistent with reports of constant shuttling of the AhR between the nucleus and cytoplasm [44], and AhR promoter occupancy of targeted genes in untreated cells [33]. Relaxing the FDR to 0.05 increased the overlap to 906, while reducing the number of 24 hr specific enriched regions to 68. Comparable overlaps were identified in promoter-specific ChIP-chip studies of TCDD-induced AhR enrichment at 2 and 24 hrs in the livers of intact C57BL/6 mice, which identified 1,397 number of genes with 403 overlap (Lo *et al.*, manuscript in preparation). Further analysis of the 899 enriched regions found that the fold enrichment values from both time points were positively correlated (Pearson correlation coefficient = 0.4853, two-tailed p-value < 0.0001; Figure 20E).

Although only 40% of the mouse genome consists of intragenic DNA, 71.2% and 64.7% of all sites with significant AhR enrichment at 2 hrs and 24 hrs, respectively, were within this region. The density of AhR enrichment (per million base pairs [Mbp]) was calculated across the entire genome in order to consider the cumulative intergenic and intragenic DNA region lengths (Table 13). Genome and chromosomal analyses (Supplementary Tables 4 and 5) revealed increased enrichment within intragenic regions compared to non-coding intergenic regions further illustrating a bias for gene encoding regions. However, these values may be inflated due to incomplete probe coverage in the intergenic regions and sequence gaps in the genome. Specific analysis of the 10 kb upstream, 5' and 3' UTRs and CDS regions revealed the highest density of AhR enrichment was proximal to the TSS (Table 13 and Supplementary Tables 4 and 5). AhR enrichment density was greatest within  $\pm 1.5$  kb at 2 and 24 hrs (Figures 21A-B),



**Figure 20. Characterization of TCDD-induced AhR enriched regions at 2 and 24 hrs (FDR < 0.01).**

Frequency analysis of enriched regions relative to log<sub>2</sub> fold enrichment at 2 hr (A) and 24 hr (B) illustrating enrichment values in intragenic (light green) and intragenic (dark green) DNA regions. Distribution of enriched regions relative to region width (C) at 2 hrs (light red) and 24 hrs (dark red) identified 90.5% of enriched sites were ≤ 1,500 bp. Comparison of AhR enriched regions at 2 and 24 hrs identified 899 overlapping regions (D). Analysis of the fold enrichment values for the 899 overlapping regions at 2 and 24 hrs identified a positive correlation (two-tailed P-value < 0.0001, Pearson correlation coefficient = 0.4853; E).



**Figure 21. TCDD-induced AhR enrichment (FDR < 0.01) densities in the proximal promoter (10 kb upstream and 5 kb downstream of a TSS) at 2 hrs (A) and 24 hrs (B).**

The bars represent the number of enriched regions in each 200 bp window. The number of DRE cores in 100 bp non-overlapping windows is superimposed (line) illustrating the overlap between AhR enriched regions and DRE cores in the proximal promoter region.

coinciding with proximal promoter DRE core densities (Dere *et al.*, in submission) and RNA polymerase II binding at the TSSs [9]. Interestingly, there is a notable cleft in AhR enrichment 200 bp directly upstream and downstream of the TSSs, possibly to accommodate general transcription machinery. Both global and proximal promoter density analyses illustrate TCDD-induced AhR enrichments are more prominent in regions directly associated with a gene. Nevertheless, there are a significant number of distally located enrichment sites that may also be functionally relevant.

### **CONFIRMATION OF AHR CHIP-CHIP ENRICHMENT ANALYSIS**

Selected regions of AhR enrichment identified by ChIP-chip analysis at 2 hrs were confirmed by ChIP-PCR (Table 14 and Figure 22). Three representative ChIP-chip enrichments from each genomic region (intergenic, 10 kb upstream of a TSS, 5' UTR, CDS and 3' UTR) were selected to validate AhR enrichments with and without a DRE core at different positions relative to the TSS. ChIP-PCR and ChIP-chip analysis of DRE containing regions exhibited similar levels of AhR enrichment relative to IgG<sub>TCDD</sub> controls and were significantly greater than vehicle controls relative to IgG<sub>vehicle</sub>. AhR enriched regions without the DRE core were also verified, further demonstrating that the AhR can interact with DNA independent of a DRE core, but does not eliminate the possibility of AhR interaction through DNA looping or protein tethering. Interestingly, the fold enrichment values for regions without the DRE core were consistently lower than those with a DRE core, suggesting AhR interactions are stronger in regions containing a DRE.

### **DRE ANALYSIS OF AHR ENRICHED REGIONS**

TCDD-elicited changes in gene expression are mediated through AhR signaling via binding to the substitution intolerant DRE core sequence (5'-GCGTG-3'). Overlaying TCDD-

**Table 14. List of AhR enriched regions<sup>a</sup> identified by ChIP-chip analysis at 2 hrs confirmed by ChIP-PCR.**

| Enrichment ID                 | Genomic region | DRE <sup>b</sup> | Gene symbol                  | Refseq                                       | Enrichment ID                   | Genomic region | DRE <sup>b</sup> | Gene symbol                 | Refseq                              |
|-------------------------------|----------------|------------------|------------------------------|----------------------------------------------|---------------------------------|----------------|------------------|-----------------------------|-------------------------------------|
| AhR_106<br>AhR_129<br>AhR_130 | Intergenic     | Yes              | NA                           | NA                                           | AhR_82<br>AhR_159<br>AhR_216    | Intergenic     | No               | NA                          | NA                                  |
| AhR_106<br>AhR_129<br>AhR_130 | 10 kb upstream | Yes              | Cyp1a1<br>Tiparp<br>Cyp1b1   | NM_009992<br>NM_178892<br>NM_009994          | AhR_4<br>AhR_35<br>AhR_1456     | 10 kb upstream | No               | Slc35d1<br>Pcp4l1<br>Cyp1a1 | NM_177732<br>NM_025557<br>NM_009992 |
| AhR_106<br>AhR_129<br>AhR_130 | 5' UTR         | Yes              | Myo1b<br>Edc3<br>Sfxn1       | NM_010863<br>NM_153799<br>NM_027324          | AhR_51<br>AhR_581<br>AhR_857    | 5' UTR         | No               | Bach2<br>Atxn1<br>Tbc1d16   | NM_007521<br>NM_009124<br>NM_172443 |
| AhR_106<br>AhR_129<br>AhR_130 | CDS            | Yes              | Prom1<br>Pkm2<br>Lmo7        | NM_008935<br>NM_011099<br>NM_201529          | AhR_146<br>AhR_201<br>AhR_908   | CDS            | No               | Npc1<br>Derl1<br>ErbB3      | NM_008720<br>NM_008720<br>NM_010153 |
| AhR_106<br>AhR_129<br>AhR_130 | 3' UTR         | Yes              | Al464131<br>Irf2bp2<br>Wipf3 | NM_001085515<br>NM_001164598<br>NM_001167860 | AhR_143<br>AhR_3220<br>AhR_5515 | 3' UTR         | No               | Rbm4<br>Irs1<br>Cebpa       | NM_009032<br>NM_009032<br>NM_007678 |

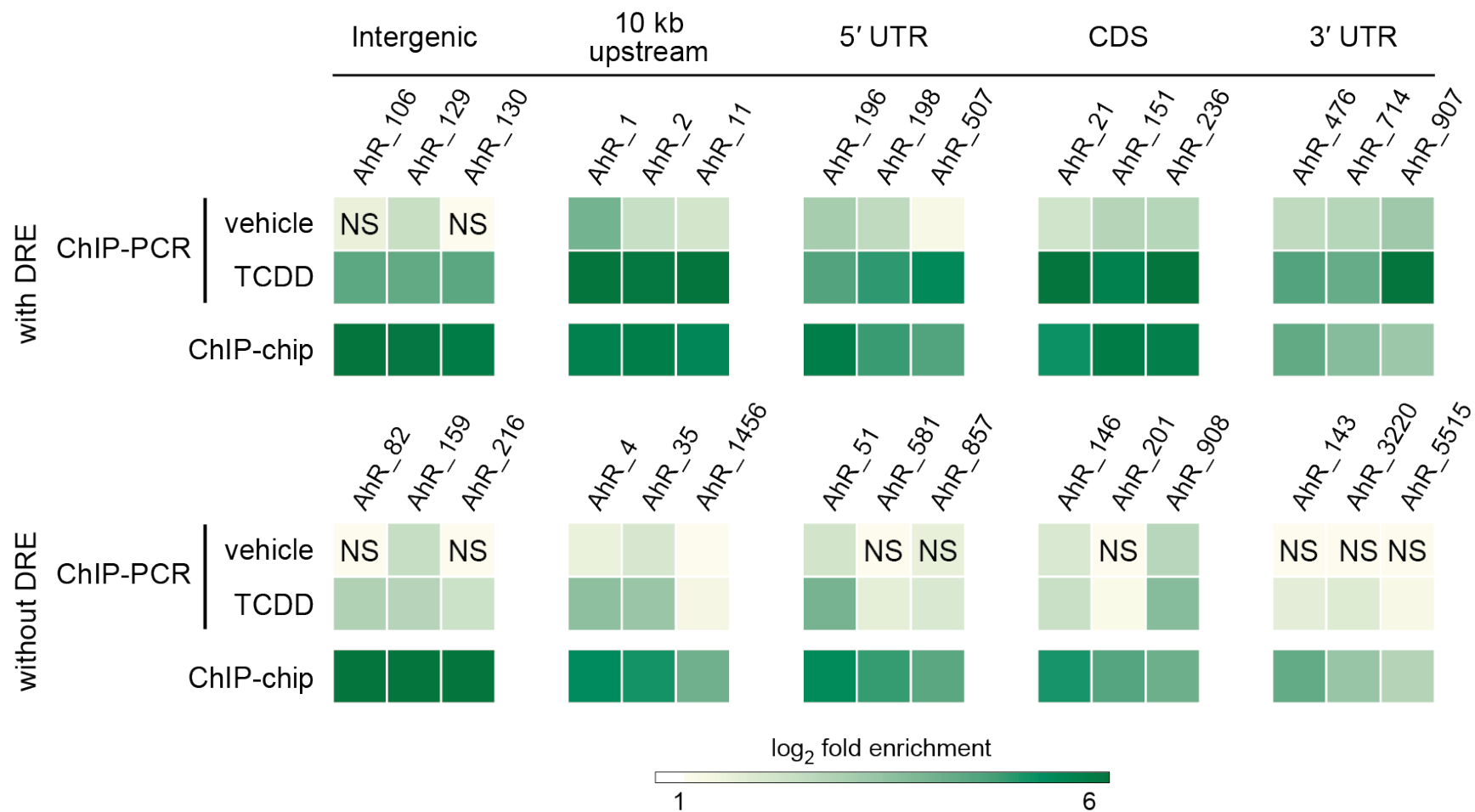
<sup>a</sup> AhR enriched regions with a FDR < 0.01

<sup>b</sup> 5'-GCGTG-3' core DRE sequence

NA: Not applicable

**Figure 22. Confirmation of hepatic TCDD-induced AhR enrichment identified by ChIP-chip analysis (FDR < 0.01) at 2 hrs by ChIP-PCR.**

AhR enriched regions listed in Table 14 were chosen for verification based on position relative to a TSS, ChIP-chip fold enrichment and the presence or lack of a DRE core within the region of enrichment. Immunoprecipitated DNA was measured by QRT-PCR and AhR enrichment was calculated as fold induction above IgG controls. The color intensity of each box represents the mean value of three independent replicates. NS = not significant compared to IgG controls ( $p < 0.05$ ). 2 hr ChIP-chip enrichment values are provided in Supplementary Table 2.





induced AhR enrichment with DRE core locations throughout the mouse genome (Dere *et al.*, in submission) identified 57.8% and 48.5% of the enriched regions did not contain a DRE core regions at 2 and 24 hrs, respectively (Table 15 and Figures 23A-B). Other promoter-specific ChIP-chip studies have also reported DRE cores in ~50% of the AhR enriched regions [32, 34]. The remaining enriched regions possessed at least one and as many as 16 DRE cores (Table 15). AhR enriched regions with or without a DRE core exhibited similar widths and levels of enrichment.

Matrix similarity (MS) scores have been calculated for each 19 bp DRE sequence within the mouse genome using a position weight matrix (PWM) constructed from *bona fide* functional DREs (Dere *et al.*, in submission). Of the 6,595 significant AhR enriched regions containing a DRE core (6,093 from 2 hr and 502 from 24 hr), 90.7% were within 500 bp of a DRE core (i.e. distance of maximum enrichment within the region to an underlying DRE core) with half of these positions located within 135 bp of a DRE core. However, only 8.3% and 17.8% of the AhR enriched regions at 2 and 24 hrs, respectively, possessed a putative functional (high scoring) DRE sequence (MS score  $\geq 0.8473$ ) suggesting the AhR may bind other degenerate sequence elements.

AhR binding to an alternate response element (5'-CATGN<sub>6</sub>C[T|A]TG-3') has also been reported [45, 46]. Of the 8,353 and 472 enriched regions at 2 and 24 hrs, respectively, that did not contain a DRE core, 482 and 237, respectively, contained the alternate DRE sequence (5.8% and 50.2%, respectively). The higher incidence of AhR enriched regions at 24 hrs containing the alternate response element may represent tertiary AhR binding sites resulting from conformational changes and crowding of the promoter with the general transcription machinery [47, 48].

**Table 15. Distribution of DRE cores in AhR enriched regions<sup>a</sup>.**

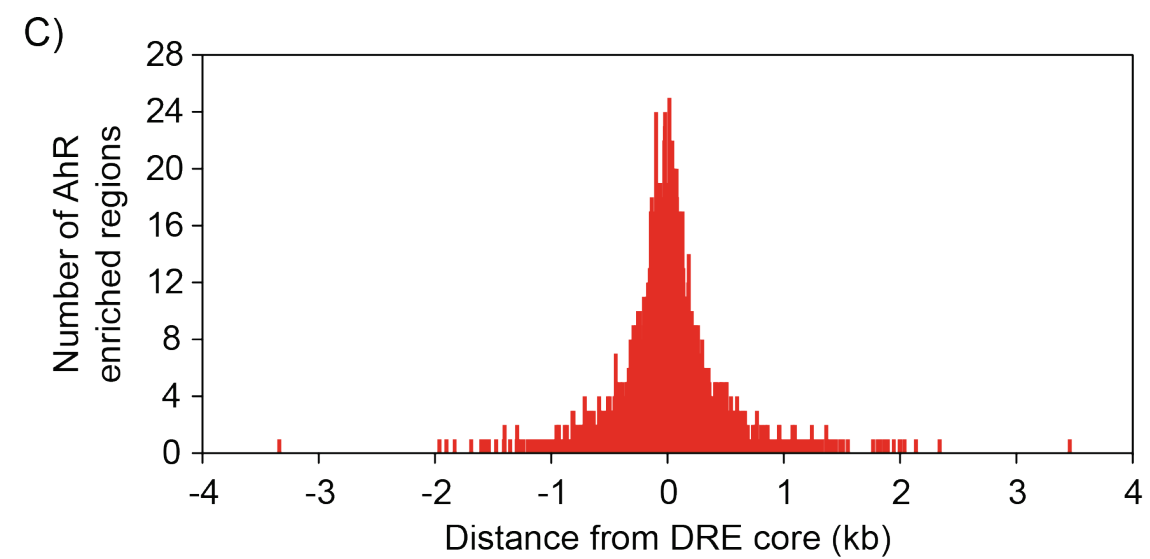
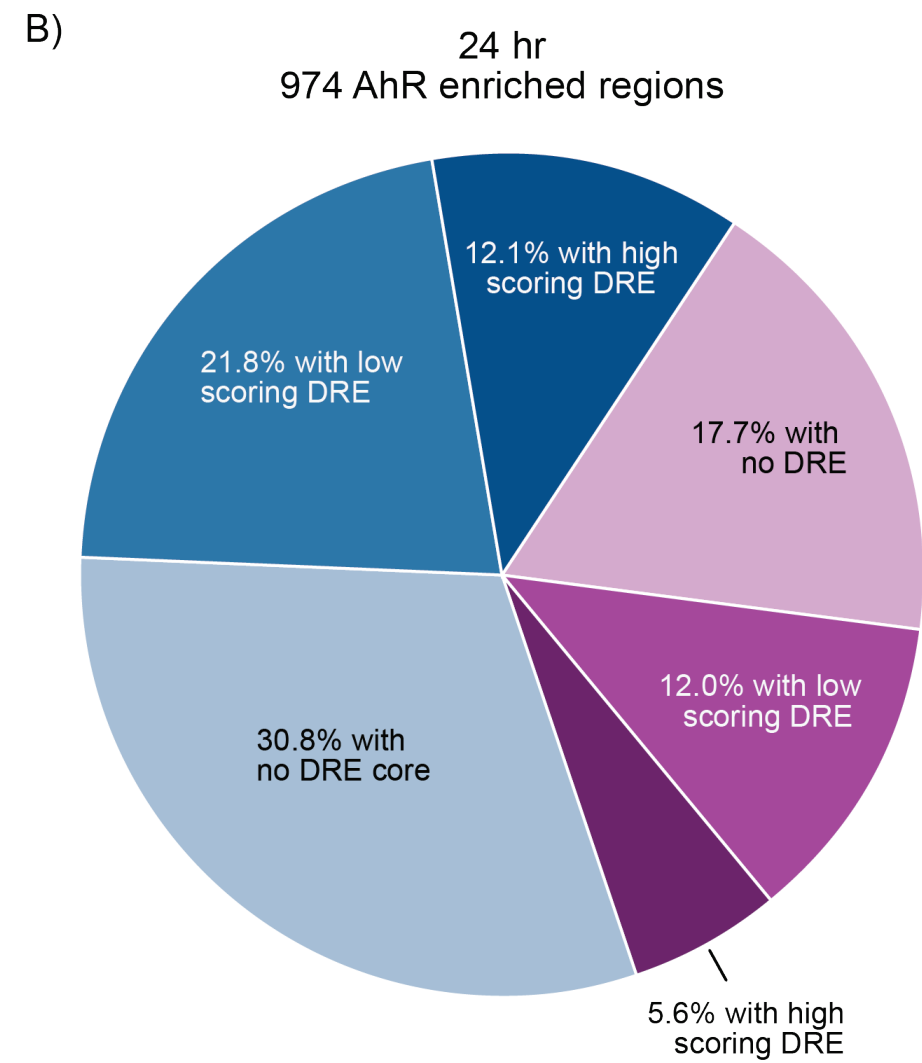
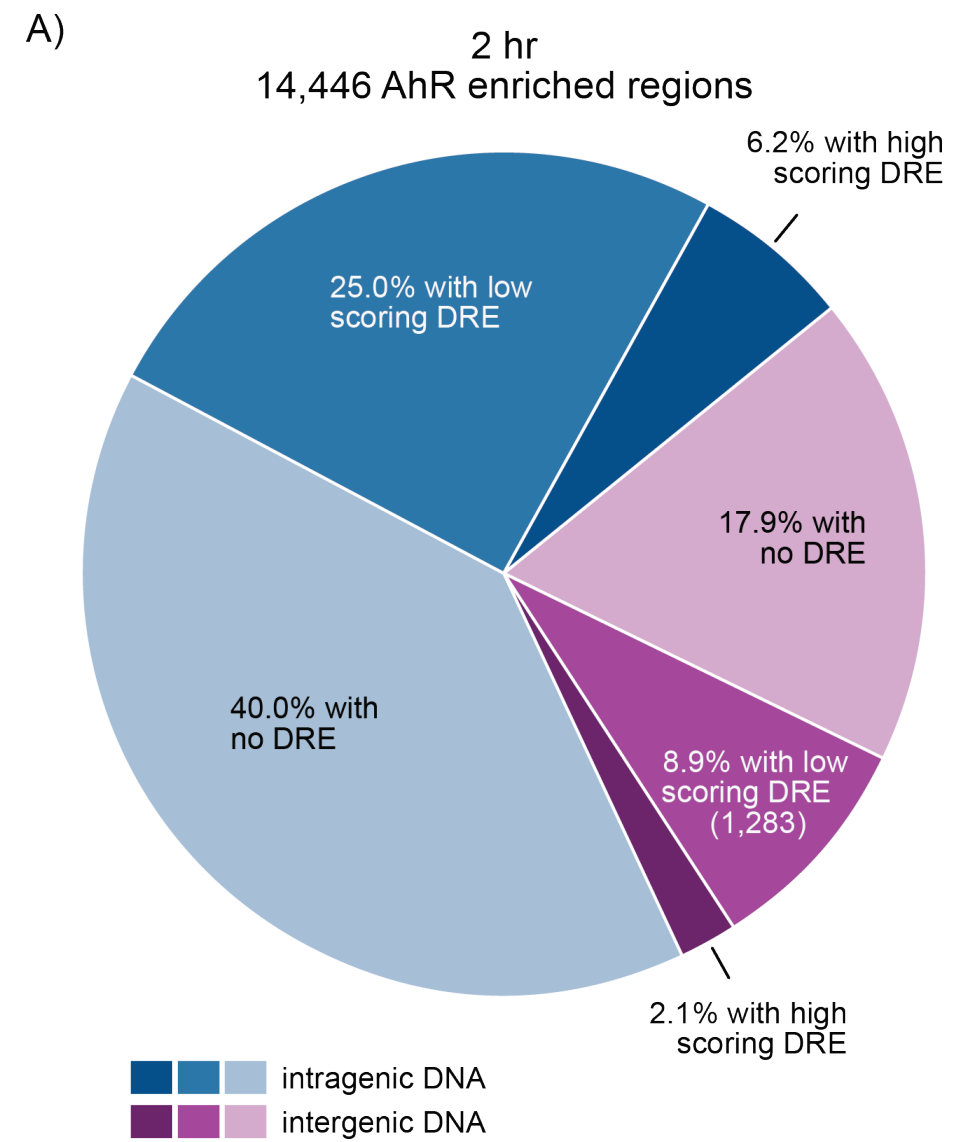
| Number of<br>DRE cores <sup>b</sup> | Number of AhR enriched regions |            |
|-------------------------------------|--------------------------------|------------|
|                                     | 2 hr                           | 24 hr      |
| 0                                   | 8,353                          | 472        |
| 1                                   | 3,705                          | 289        |
| 2                                   | 1,372                          | 121        |
| 3                                   | 544                            | 46         |
| 4                                   | 223                            | 16         |
| 5                                   | 109                            | 12         |
| 6                                   | 67                             | 7          |
| 7                                   | 25                             | 5          |
| 8                                   | 15                             | 0          |
| 9                                   | 11                             | 3          |
| 10                                  | 7                              | 1          |
| 11                                  | 5                              | 0          |
| 12                                  | 3                              | 1          |
| 13                                  | 0                              | 0          |
| 14                                  | 3                              | 1          |
| 15                                  | 3                              | 0          |
| 16                                  | 1                              | 0          |
| <b>Total</b>                        | <b>14,446</b>                  | <b>974</b> |

<sup>a</sup> AhR enriched regions with a FDR < 0.01

<sup>b</sup> 5'-GCGTG-3' core sequence

**Figure 23. Mapping TCDD-induced AhR enriched regions (FDR < 0.01) with DRE locations.**

Regions of enrichment identified in the intergenic (purple) and intragenic (blue) DNA regions of the genome at 2 hrs (A) and 24 hrs (B) were searched for high scoring (putative functional) DRE sequences (matrix similarity score  $\geq 0.8473$ ; dark blue and dark purple segments) and low scoring DRE sequences (matrix similarity score  $< 0.8473$ ; mid blue and mid purple segments) using a position weight matrix developed from *bona fide* functional DREs (Dere *et al.*, in submission). Light blue and light purple segments represent regions with no DRE core sequence. A total of 6,595 enriched regions (6,093 at 2 hrs and 502 at 24 hrs) contained at least one DRE core (5'-GCGTG-3'). 50% of these regions were within 135 bp of a DRE core (based on the location of maximum enrichment within the enriched region; C).



### **TRANSCRIPTION FACTOR BINDING SITE OVER-REPRESENTATION ANALYSIS**

Significantly AhR enriched regions were computationally analyzed for over-represented response elements for known TF binding site (TFBS) families using RegionMiner (Genomatix). DREs as well other sites for early growth response (EGR), E2F, nuclear respiratory factor 1 (NRF1), nuclear receptor subfamily 2 factors (NR2F/COUP-TF) and peroxisome proliferator-activated receptor (PPAR) were over-represented within AhR enriched regions (Table 16; complete list of over-represented TFBS are provided in Supplementary Tables 6 and 7). Many of these TF sites were enriched proximally to a DRE core (i.e. within 10-50 bp; Supplementary Table 8) suggesting possible interactions. Studies have previously reported interactions between AhR and many of these TFs [33, 49, 50]. For example, AhR complexes with EGR-1 following treatment of human HUVEC cells with high glucose concentrations [50]. In addition, AhR aggregates with E2F1 to inhibit E2F1-induced apoptosis [51]. AhR also directly interacts with COUP-TF to repress ER-mediated gene expression [52].

### **DE NOVO MOTIF ANALYSIS**

Approximately 50% of enriched regions lacked the DRE core sequence (Figures 23A-B) suggesting AhR interacts with DNA using alternate strategies. *De novo* motif analysis of these regions using the Gibbs motif sampler in CisGenome identified over-representation of comparable repetitive elements in both the intergenic and intragenic DNA regions (Supplementary Figure 2). Comparison of over-represented non-repetitive motifs to existing TF binding motifs in JASPAR and TRANSFAC [39, 40] using STAMP [41] identified similarities to COUP-TF, hepatocyte nuclear factor 4 (HNF4), liver receptor homolog 1 (LRH1/NR5A2) and PPAR binding sites (Figure 24). Interestingly, COUP-TF and HNF4 belong to the NR2F family identified in the TFBS over-representation analysis of all AhR enriched regions (Table 16). The

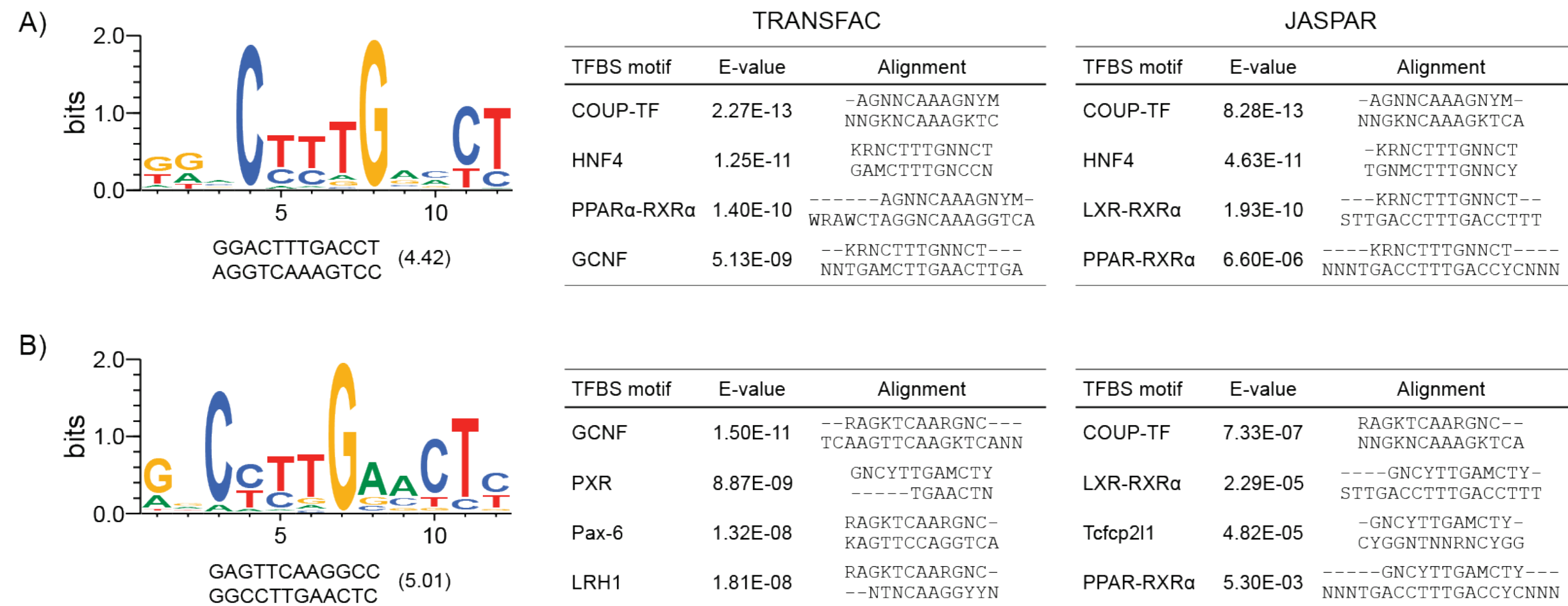
Table 16. Significantly over-represent transcription factor module families in TCDD-induced AhR enriched regions<sup>a</sup>.

| TF Module Family | Module Description                                | 2 hr AhR enriched regions |                       |                                  |         | 24 hr AhR enriched regions |                       |                                  |         |
|------------------|---------------------------------------------------|---------------------------|-----------------------|----------------------------------|---------|----------------------------|-----------------------|----------------------------------|---------|
|                  |                                                   | # of matches              | Expected # of matches | Over-representation <sup>b</sup> | Z-Score | # of matches               | Expected # of matches | Over-representation <sup>b</sup> | Z-Score |
| AHR              | AhR-ARNT heterodimer                              | 9,447                     | 4,278.30              | 2.21                             | 79.03   | 851                        | 297.44                | 2.86                             | 32.07   |
| SP1              | GC-Box factors SP1/GC                             | 19,356                    | 12,839.04             | 1.51                             | 57.54   | 1,346                      | 892.61                | 1.51                             | 15.17   |
| HIF              | Hypoxia inducible factor, bHLH/PAS protein family | 8,763                     | 4,841.13              | 1.81                             | 56.37   | 569                        | 336.57                | 1.69                             | 12.64   |
| E2F              | E2F-Myc activator/cell cycle regulator            | 18,247                    | 12,444.88             | 1.47                             | 52.04   | 1,266                      | 865.21                | 1.46                             | 13.62   |
| ZBP              | Zinc binding protein factors                      | 25,542                    | 18,518.20             | 1.38                             | 51.65   | 1,739                      | 1,287.45              | 1.35                             | 12.58   |
| NRF1             | Nuclear respiratory factor 1                      | 3,475                     | 1,494.81              | 2.32                             | 51.21   | 252                        | 103.92                | 2.42                             | 14.48   |
| ZF5              | ZF5 POZ domain zinc finger                        | 3,156                     | 1,442.92              | 2.19                             | 45.09   | 205                        | 100.32                | 2.04                             | 10.40   |
| NF1              | Nuclear factor 1                                  | 13,047                    | 8,876.90              | 1.47                             | 44.27   | 886                        | 617.15                | 1.44                             | 10.81   |
| NR2F             | Nuclear receptor subfamily 2 factors              | 44,774                    | 36,390.64             | 1.23                             | 44.02   | 3,180                      | 2,530.00              | 1.26                             | 12.93   |
| EGR              | EGR/nerve growth factor induced protein C         | 22,224                    | 16,794.62             | 1.32                             | 41.92   | 1,541                      | 1,167.62              | 1.32                             | 10.92   |
| PPAR             | Peroxisome proliferator-activated receptor        | 24,808                    | 19,035.70             | 1.30                             | 41.87   | 1,752                      | 1,323.43              | 1.32                             | 11.78   |
| RXR              | RXR heterodimer binding sites                     | 41,027                    | 33,441.00             | 1.23                             | 41.54   | 2,932                      | 2,324.93              | 1.26                             | 12.60   |
| WHN              | Winged helix binding sites                        | 3,030                     | 1,477.23              | 2.05                             | 40.39   | 206                        | 102.70                | 2.01                             | 10.14   |

<sup>a</sup> AhR enriched regions with a FDR < 0.01

<sup>b</sup> 5'-GCGTG-3' core sequence

Complete list of over-represented TF module families are provided in Supplementary Tables 5 and 6



**Figure 24. *De novo* motif analysis of intragenic (A) and intergenic (B) AhR enriched regions lacking a DRE core.** The non-repetitive over-represented motifs from each region are shown with their consensus and reverse complement sequence, and the Gibbs motif sampler score. Over-represented motifs were associated with specific TFBSs in JASPAR and TRANSFAC based on the consensus sequence alignments and E-value scores.

presence of these binding motifs in non-DRE containing regions of AhR enrichment further suggests that AhR-DNA interactions occur through a tethering mechanism involving other TFs or by tertiary looping of DNA.

### **GENE LEVEL ANALYSIS OF AHR ENRICHMENT**

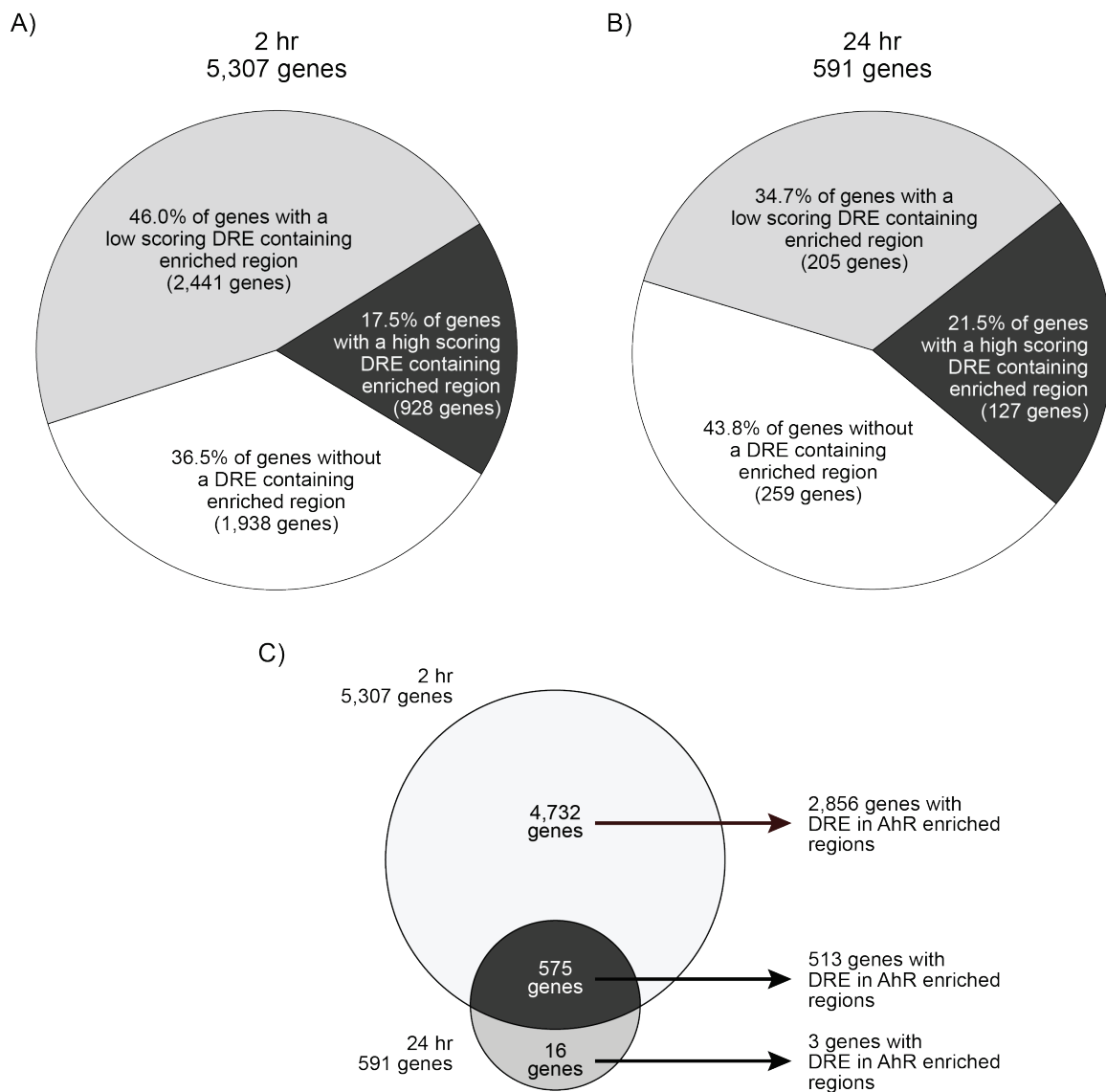
Of the 10,283 enrichments identified in the intragenic DNA regions, 43.9% (4,510/10,283) containing a DRE core at 2 hrs, and 52.4% (330/630) at 24 hrs (Figure 22, areas shaded blue). These intragenic AhR enriched regions mapped to 5,307 and 591 unique genes at 2 and 24 hrs, respectively (AhR targeted genes are provided as gene annotated enriched regions in Supplementary Tables 2 and 3). Analysis of genes associated with AhR enrichment identified 63.5 and 56.2% contained a DRE core within the 2 and 24 hr data sets, respectively (Figure 25). The higher percentage of genes containing a DRE core compared to enriched regions with a DRE core is due to multiple regions of AhR enrichment associated with a single gene (as illustrated for *Cyp1a1* in Figure 19). The remaining genes (36.5% and 54.8% at 2 and 24 hrs, respectively) with significant AhR enrichment were targeted independently of a DRE core.

At both 2 and 24 hrs, 575 genes had AhR enrichment, with 513 possessing DRE cores in the AhR enriched region (Figure 25C). Only 16 genes exhibited AhR enrichment solely at 24 hrs, with three containing a DRE core. In contrast, 4,732 genes possessed significant AhR enrichment with 60.4% (2,856) containing a DRE core within the region of enrichment at 2 hrs. Due to the large overlap of enriched regions at 2 and 24 hrs, the remaining analysis focuses predominantly on the AhR enrichment at 2 hr.

### **COMPARISON OF TRANSCRIPTIONAL RESPONSES WITH AHR ENRICHMENT**

Gene expression analysis at 2, 4, 8, 12, 18, 24, 72, and 168 hrs identified 1,896 differentially expressed genes ( $|\text{fold change}| \geq 1.5$  and  $P1(t) > 0.999$ ) at one or more time points.





**Figure 25. Mapping TCDD-induced AhR enriched regions (FDR < 0.01) and DRE analysis to genes.**

The 10,283 and 660 AhR enrichments within the intragenic DNA regions at 2 and 24 hrs (blue shaded areas in Figures 23A-B) mapped to 5,307 (A) and 591 (B) distinct genes based on the refGene data from the UCSC Genome Browser. These genes were searched for the presence of high (matrix similarity score (MS)  $\geq 0.8473$ ; dark grey areas) and low (MS score  $< 0.8473$ ; light grey areas) scoring DRE sequences, and the absence of a DRE core (white areas) within the region of AhR enrichment. Comparing 2 and 24 hrs data identified 575 overlapping genes with AhR enrichment and 513 of these genes contained a DRE core within the region of enrichment (C).

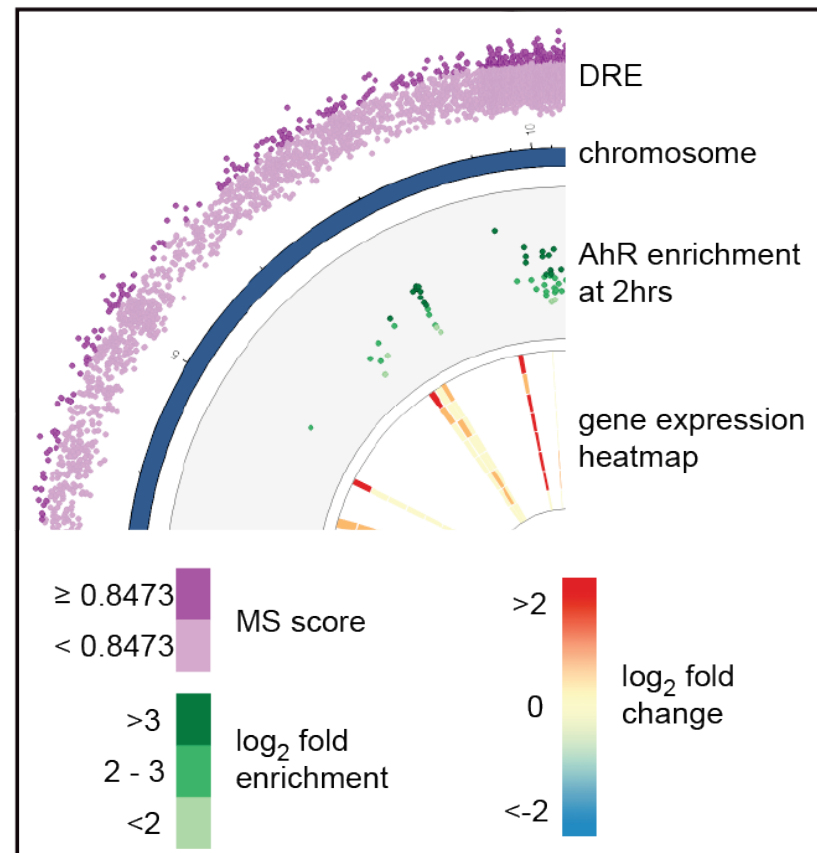
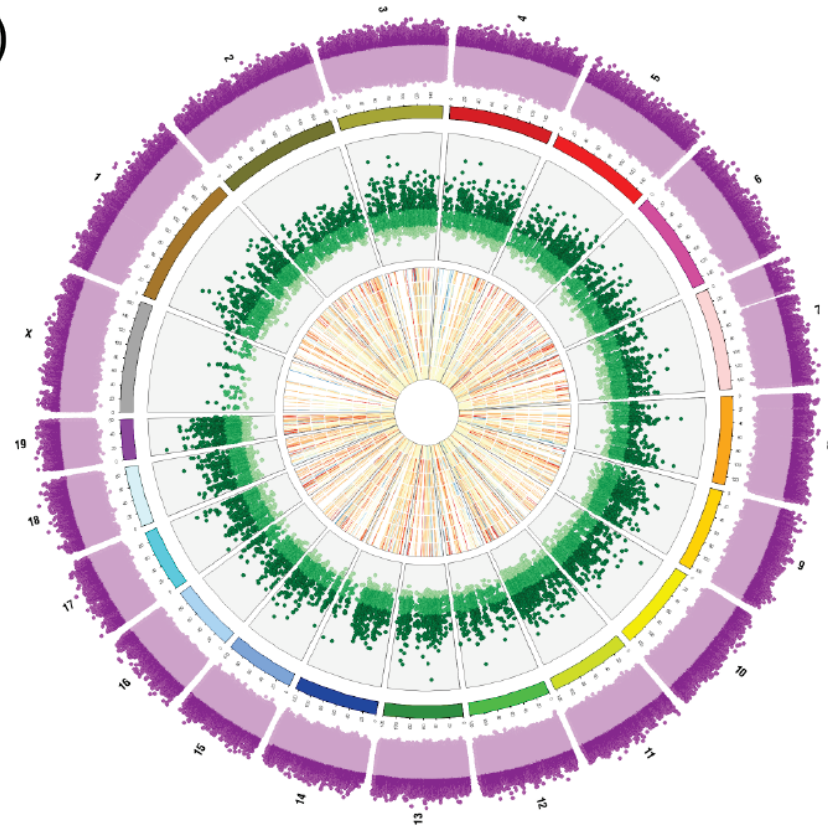
Of the 1,896 TCDD-responsive genes, 900 genes (47.5%) possessed significant AhR enrichment within the intragenic region (10 kb upstream of the TSS to the end of the transcript). Moreover, of the 900 genes exhibiting AhR enrichment at 2 hrs, 625 contained a DRE core sequence, suggesting these responses are AhR-mediated. In order to concisely visualize the integration of the DRE, ChIP-chip and gene expression analyses, Circos plots were generated for the genome and individual chromosomes (Figure 26 and Supplementary Figure 3). The plots further illustrate the diversity in AhR enrichment locations in relation to the genomic position of dysregulated genes. Further analysis of the responsive genes found that most were induced by TCDD (Table 17) at all time points. Greater than 82% of the induced genes at 2 or 4 hrs had significant AhR enrichment, and more than 62% of them contained at least one DRE core suggesting that regulation is DRE-dependent fashion. In contrast, only 35% of the 691 genes induced at 168 hrs, exhibited AhR enrichment with 26% possessing a DRE core suggesting that these are secondary gene expression responses. Interestingly, down-regulated genes associated with AhR enrichment were relatively consistent across all time points. Approximately one third of the down-regulated genes appear to be AhR regulated with DRE involvement.

Functional analysis of the 900 differentially expressed genes associated with AhR enrichment was performed using DAVID [53]. The most over-represented functions were associated with lipid metabolic processes (enrichment score of 7.34, Table 18), consistent with the induced fatty liver phenotype [35, 54]. In addition, *de novo* motif analysis (Figure 24) identified binding sites for TFs associated with lipid metabolism and transport. The induction of AhR regulated xenobiotic enzymes, such as cytochrome P450s, glutathione S-transferases (Gsts) and UDP-glucuronosyltransferases (Ugts), hallmarks of TCDD exposure, were also identified as an enriched cluster (enrichment score of 3.54).

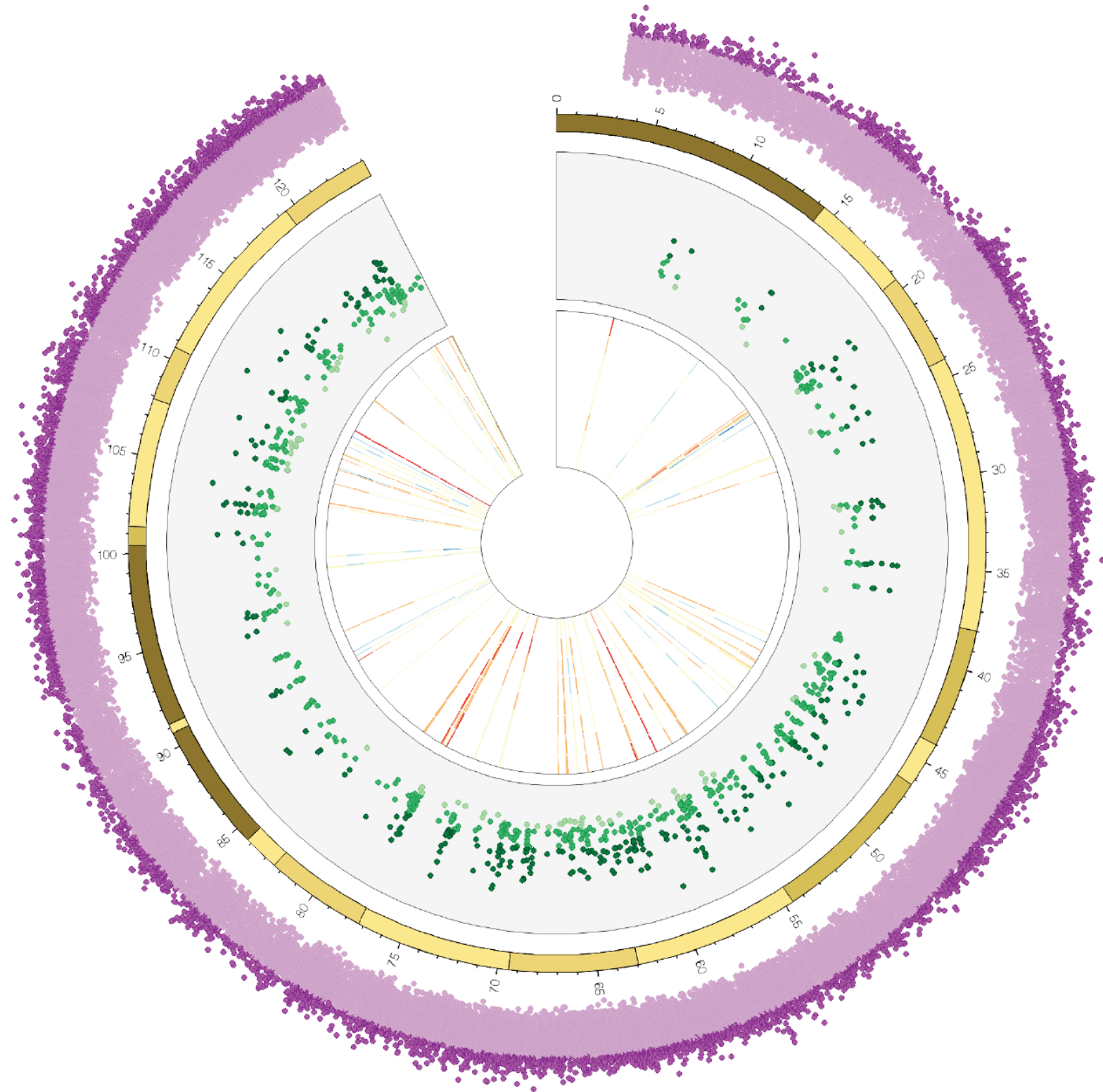
**Figure 26. Circos plots integrating DRE analysis, AhR enrichment (2 hrs; FDR < 0.01) and heatmaps for hepatic differential gene expression responses ( $|\text{fold change}| \geq 1.5$  and  $P1(t) > 0.999$ ) induced by TCDD across the genome (A) and chromosome 9 (B).**

The inset legend image provides information represented by each data ring. DRE matrix similarity (MS) scores and AhR enrichment values increase radially outward. The time points for the gene expression heatmaps also increase radial outward. The arc of each heatmap wedge maps directly to the location of the gene in the genome. The arc length is proportional to the length of the transcribed region. Circos plots for the other chromosomes are provided in Supplementary Figure 3.

A)



B)



**Table 17. Distribution and AhR enrichment and DRE analyses of differentially expressed genes elicited by TCDD.**

|                                                          |                    |                                                             | 2 hr | 4 hr | 8 hr | 12 hr | 18 hr | 24 hr | 72 hr | 168 hr |
|----------------------------------------------------------|--------------------|-------------------------------------------------------------|------|------|------|-------|-------|-------|-------|--------|
| Number of differentially<br>expressed genes <sup>a</sup> | Up-<br>regulated   | Total                                                       | 65   | 244  | 328  | 204   | 224   | 276   | 255   | 691    |
|                                                          |                    | With AhR enrichment <sup>b</sup>                            | 55   | 200  | 202  | 148   | 172   | 186   | 164   | 243    |
|                                                          |                    | With AhR enrichment <sup>b</sup> + DRE<br>core <sup>c</sup> | 47   | 168  | 171  | 126   | 146   | 156   | 135   | 181    |
|                                                          | Down-<br>regulated | Total                                                       | 17   | 220  | 110  | 163   | 101   | 224   | 246   | 210    |
|                                                          |                    | With AhR enrichment <sup>b</sup>                            | 10   | 123  | 76   | 102   | 59    | 131   | 137   | 123    |
|                                                          |                    | With AhR enrichment <sup>b</sup> + DRE<br>core <sup>c</sup> | 6    | 79   | 49   | 63    | 30    | 72    | 79    | 70     |

<sup>a</sup> |fold-change| ≥ 1.5 and P1(t) > 0.999

<sup>b</sup> AhR enriched regions at 2 hrs with FDR < 0.01

<sup>c</sup> 5'-GCGTG-3' core sequence within AhR enriched region

Although AhR mediates the expression of enzymes involved in xenobiotic metabolizing enzymes, including NADP(H) dehydrogenase, quinone 1 (*Nqo1*) and UDP-glucose dehydrogenase (*Ugdh*) as well as several Ugt and Gst isoforms, they are also regulated by nuclear factor, erythroid derived 2, like 2 (Nrf2) via antioxidant response elements in response to oxidative stress [55, 56]. Recent studies with AhR and Nrf2 null mice report that TCDD induction of *Nqo1* is AhR and Nrf2 dependent [57]. Furthermore, specific Ugt and Gst isoforms induced by TCDD require Nrf2. Collectively, these responses are referred to as the “TCDD-inducible AhR-Nrf2 gene battery.” ChIP-chip and gene expression analysis indicates that *Nqo1*, *Gstm1*, *Gstm2*, *Ugdh* and *Nrf2* induction is associated with AhR enrichment. Although supportive of the Nrf2-dependency model, these data do not distinguish if these are secondary responses mediated by Nrf2 alone, or involve an AhR-Nrf2 interaction. In contrast, *Gsta1* and *Ugt2b35* induction occurred independently of AhR enrichment, suggesting they may only be dependent on Nrf2 [57].

Immune cell accumulation following a single acute dose of TCDD at 168 hrs is presumed to be a secondary response to hepatic injury or fatty acid accumulation [35, 54]. DAVID analysis of genes induced at 168 hrs identified multiple over-represented immune-related clusters (enrichment scores > 2). However, several of the genes including complement component 1, q subcomponent, beta polypeptide (*C1qb*), CD36 antigen (*Cd36*), complement component 4A (C4a) and interferon regulatory factor 8 (Irf8), did not exhibit accompanying AhR enrichment within their intragenic region (10 kb upstream of the TSS to the end of the 3' UTR). Only 26 out of 105 differentially regulated genes in the enriched immune clusters exhibited AhR enrichment. Collectively, these data suggest that gene expression associated with immune function is a consequence of immune cell infiltration into the liver.

**Table 18. Functional enrichment analysis of differentially regulated<sup>a</sup> genes with AhR enrichment<sup>b</sup> using DAVID.**

| Category                      | Term                                                                    | Gene Count | Fold enrichment | P-value  |
|-------------------------------|-------------------------------------------------------------------------|------------|-----------------|----------|
| <b>Enrichment score: 7.34</b> |                                                                         |            |                 |          |
| GOTERM_BP_3                   | GO:0006629 ~ lipid metabolic process                                    | 76         | 2.23            | 5.53E-11 |
| GOTERM_BP_3                   | GO:0044255 ~ cellular lipid metabolic process                           | 53         | 2.23            | 7.80E-08 |
| GOTERM_BP_3                   | GO:0008610 ~ lipid biosynthetic process                                 | 32         | 2.30            | 2.28E-05 |
| <b>Enrichment score: 4.12</b> |                                                                         |            |                 |          |
| GOTERM_BP_3                   | GO:0048523 ~ negative regulation of cellular process                    | 94         | 1.60            | 4.46E-06 |
| GOTERM_BP_3                   | GO:0048519 ~ negative regulation of biological process                  | 101        | 1.54            | 7.83E-06 |
| GOTERM_BP_3                   | GO:0031324 ~ negative regulation of cellular metabolic process          | 45         | 1.85            | 8.99E-05 |
| GOTERM_BP_3                   | GO:0051172 ~ negative regulation of nitrogen compound metabolic process | 39         | 1.96            | 9.68E-05 |
| GOTERM_BP_3                   | GO:0009890 ~ negative regulation of biosynthetic process                | 40         | 1.86            | 2.33E-04 |
| GOTERM_BP_3                   | GO:0009892 ~ negative regulation of metabolic process                   | 46         | 1.74            | 3.20E-04 |
| GOTERM_BP_3                   | GO:0010605 ~ negative regulation of macromolecule metabolic process     | 43         | 1.72            | 6.88E-04 |
| <b>Enrichment score: 3.54</b> |                                                                         |            |                 |          |
| GOTERM_BP_3                   | GO:0009410 ~ response to xenobiotic stimulus                            | 8          | 10.59           | 3.54E-06 |
| GOTERM_BP_3                   | GO:0006805 ~ xenobiotic metabolic process                               | 7          | 11.59           | 1.12E-05 |
| GOTERM_BP_3                   | GO:0018894 ~ dibenzo-p-dioxin metabolic process                         | 3          | 19.86           | 7.31E-03 |
| GOTERM_BP_3                   | GO:0009404 ~ toxin metabolic process                                    | 3          | 11.92           | 2.28E-02 |



**Table 18: (cont'd).**

| Category                      | Term                                                                    | Gene Count | Fold enrichment | P-value  |
|-------------------------------|-------------------------------------------------------------------------|------------|-----------------|----------|
| <b>Enrichment score: 2.70</b> |                                                                         |            |                 |          |
| GOTERM_BP_3                   | GO:0051272 ~ positive regulation of cell motion                         | 9          | 4.58            | 6.00E-04 |
| GOTERM_BP_3                   | GO:0051270 ~ regulation of cell motion                                  | 15         | 2.84            | 7.55E-04 |
| GOTERM_BP_3                   | GO:0040017 ~ positive regulation of locomotion                          | 9          | 4.26            | 1.01E-03 |
| GOTERM_BP_3                   | GO:0030334 ~ regulation of cell migration                               | 13         | 2.87            | 1.76E-03 |
| GOTERM_BP_3                   | GO:0040012 ~ regulation of locomotion                                   | 13         | 2.39            | 7.93E-03 |
| GOTERM_BP_3                   | GO:0032879 ~ regulation of localization                                 | 32         | 1.61            | 9.36E-03 |
| <b>Enrichment score: 2.55</b> |                                                                         |            |                 |          |
| GOTERM_BP_3                   | GO:0048518 ~ positive regulation of biological process                  | 106        | 1.44            | 7.45E-05 |
| GOTERM_BP_3                   | GO:0048522 ~ positive regulation of cellular process                    | 94         | 1.45            | 2.03E-04 |
| GOTERM_BP_3                   | GO:0009893 ~ positive regulation of metabolic process                   | 51         | 1.53            | 2.64E-03 |
| GOTERM_BP_3                   | GO:0010604 ~ positive regulation of macromolecule metabolic process     | 48         | 1.54            | 3.02E-03 |
| GOTERM_BP_3                   | GO:0031325 ~ positive regulation of cellular metabolic process          | 47         | 1.48            | 6.67E-03 |
| GOTERM_BP_3                   | GO:0009891 ~ positive regulation of biosynthetic process                | 38         | 1.38            | 3.96E-02 |
| GOTERM_BP_3                   | GO:0051173 ~ positive regulation of nitrogen compound metabolic process | 36         | 1.39            | 4.40E-02 |

<sup>a</sup> |fold-change| ≥ 1.5 and P1(t) > 0.999

<sup>b</sup> AhR enriched regions at 2 hrs with FDR < 0.01



## DISCUSSION

This study further elucidates the role of the AhR in mediating the hepatic effects of TCDD in C57BL6 mice. Recent studies have mapped AhR binding using promoter-focused ChIP-chip arrays and found that ~50% of the AhR enriched regions were devoid of the DRE core [31-33]. Additionally, ChIP-seq experiments for other TFs have also demonstrated enrichment in remote genome regions, which may serve important regulatory roles [9, 10, 13, 16]. Collectively these data suggest the AhR uses different mechanisms to regulate gene expression. Moreover, the integration of genome-wide *in silico* DRE search, with *de novo* motif analysis and TCDD-elicited hepatic temporal gene expression data has further elucidated the hepatic AhR gene regulatory network.

ChIP-chip analysis identified 14,446 TCDD-induced AhR regions at 2 hrs and 974 regions at 24 hrs, consistent with the rapid nuclear export and subsequent degradation of the AhR following TCDD activation [43]. Approximately half of these regions were within intragenic regions (10 kb upstream of a TSS to the end of the 3' UTR). Furthermore, 25% of these enriched regions at 2 hrs and 19% at 24 hrs were within 2 kb of a TSS, indicating that a large subset of AhR enrichment occurs adjacent to a TSS. Unlike other studies that report a normal distribution of TF binding centered around the TSS [14, 58-60], the AhR density profile exhibited a cleft immediately adjacent to the TSS, possibly to accommodate recruited transcriptional machinery.

Although most AhR enrichment regions are intragenic, a significant number are located in distal intergenic regions (i.e. 4,163 of 14,446 at 2 hrs and 344 of 974 at 24 hrs). Studies with the ER, p53 and forkhead box protein A1 [9, 10, 13, 16] suggest distal TF binding may have distinct regulatory roles. Binding proximal to the TSS is presumed to stabilize the general transcriptional machinery, while distal binding regulates transcription by a looping mechanism or

by altering chromatin structure [8, 61, 62]. Consequently, AhR binding outside of the proximal promoter region may have important regulatory roles that remain largely uninvestigated.

Comparing AhR enriched regions with DRE cores revealed that their intergenic, intragenic and genic (10 kb upstream, UTRs, and CDS) density distributions were similar. The greatest density of AhR enrichment associated with a DRE core occurred within the proximal promoter. Both exhibited comparable distribution profiles except for the cleft in enrichment at the TSS. The decrease in AhR enrichment at the TSS coincides with RNA polymerase II binding at the TSSs [9] of transcriptionally responsive genes. Although TCDD-elicited differential gene expression is thought to be mediated by the substitution intolerant DRE core sequence (5'-GCGTG-3'), only ~50% of the AhR enriched regions contained a DRE core, consistent with findings in other promoter targeted AhR ChIP-chip studies [32, 34] (Lo *et al.*, manuscript in preparation). Moreover, relatively few alternative AhR response elements (5'-CATGN<sub>6</sub>C[T|A]TG-3') [45, 46] were identified in AhR enriched regions lacking a DRE core sequence. Enrichment in regions lacking DRE cores provides additional evidence of AhR-DNA interactions that don not involve the basic bHLH domain [63], such as tethering to other DNA interacting TFs and/or tertiary interactions with looping DNA.

Integration of gene expression, ChIP-chip, and DRE distribution data suggested that approximately 35% of all differentially expressed genes are mediated by direct AhR binding to a DRE. Consequently, 65% of the gene expression responses elicited by TCDD do not involve DRE binding. However, TF binding analyses based on tiling arrays is limited by the extent of probe coverage (Figure 19). Genomic regions lacking probe coverage may falsely inflate the number of DRE-absent AhR enriched regions, thus underestimating the number of AhR regulated genes involving a DRE.

TCDD induces hepatic vacuolization and lipid accumulation with differential gene expression associated with fatty acid metabolism and transport in the mouse [24, 35]. Functional annotation of differentially regulated genes with significant AhR enrichment identified clusters related to fatty acid and lipid metabolism. Computational analysis also identified over-represented binding motifs for TFs involved in the regulation of lipid and cholesterol metabolism, including sites for HNF4, liver X receptor (LXR/NR1H), pregnane X receptor (PXR), PPAR and COUP-TF. COUP-TF is a potent repressor that antagonizes transcriptional responses mediated by other nuclear receptors including HNF4, PPAR, estrogen receptor (ER), retinoic acid receptor (RAR) and vitamin D receptor (VDR) [64]. For example, COUP-TF antagonizes HNF4-mediated responses by binding HNF4 response elements [65-69]. Furthermore, AhR interactions with COUP-TF repress ER-mediated gene expression responses [52]. Data for *Cyp7a1*, the HNF4-regulated rate-limiting enzyme in the conversion of cholesterol into bile acids, demonstrated that TCDD elicited a 9.2-fold enrichment of AhR in the intragenic region and 2.3-fold repression in gene expression at 168 hrs. Furthermore results from the *de novo* motif analysis identified an over-represented sequence that was highly similar to the HNF4 binding motif. Studies examining AhR-COUP-TF interactions and their effects on HNF4 target gene expression are being investigated further.

This study identified the genome-wide locations of TCDD-induced hepatic AhR enrichment *in vivo* and incorporates DRE distribution and differential gene expression data to further elucidate the hepatic AhR regulatory network. In addition to identifying interactions in regions associated with genes, AhR enrichment in distal non-coding intergenic regions was characterized. The functional significance of these distal interactions is unknown but intergenic binding has been reported for other TFs, and warrants further investigation. Moreover, only

~50% of all AhR enriched regions involved a DRE, suggesting that indirect AhR binding to DNA plays a significant role in the AhR regulatory network.

## REFERENCES

## REFERENCES

1. Gu Y, Hogenesch J, Bradfield C: **The PAS superfamily: sensors of environmental and developmental signals.** *Annu Rev Pharmacol Toxicol* 2000, **40**:519-561.
2. Denison MS, Heath-Pagliuso S: **The Ah receptor: a regulator of the biochemical and toxicological actions of structurally diverse chemicals.** *Bulletin of environmental contamination and toxicology* 1998, **61**:557-568.
3. Poland A, Knutson JC: **2,3,7,8-tetrachlorodibenzo-p-dioxin and related halogenated aromatic hydrocarbons: examination of the mechanism of toxicity.** *Annu Rev Pharmacol Toxicol* 1982, **22**:517-554.
4. Hankinson O: **The aryl hydrocarbon receptor complex.** *Annu Rev Pharmacol Toxicol* 1995, **35**:307-340.
5. Okey AB, Vella LM, Harper PA: **Detection and characterization of a low affinity form of cytosolic Ah receptor in livers of mice nonresponsive to induction of cytochrome P1-450 by 3-methylcholanthrene.** *Molecular Pharmacology* 1989, **35**:823-830.
6. Gonzalez F, Fernandez-Salguero P: **The aryl hydrocarbon receptor: studies using the AHR-null mice.** *Drug Metab Dispos* 1998, **26**:1194-1198.
7. Swanson H, Chan W, Bradfield C: **DNA binding specificities and pairing rules of the Ah receptor, ARNT, and SIM proteins.** *J Biol Chem* 1995, **270**:26292-26302.
8. Farnham PJ: **Insights from genomic profiling of transcription factors.** *Nat Rev Genet* 2009, **10**:605-616.
9. Carroll JS, Meyer CA, Song J, Li W, Geistlinger TR, Eeckhoutte J, Brodsky AS, Keeton EK, Fertuck KC, Hall GF, et al: **Genome-wide analysis of estrogen receptor binding sites.** *Nature Genetics* 2006, **38**:1289-1297.
10. Carroll JS, Liu XS, Brodsky AS, Li W, Meyer CA, Szary AJ, Eeckhoutte J, Shao W, Hestermann EV, Geistlinger TR, et al: **Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1.** *Cell* 2005, **122**:33-43.
11. Deblois G, Hall JA, Perry M-C, Laganière J, Ghahremani M, Park M, Hallett M, Giguère V: **Genome-wide identification of direct target genes implicates estrogen-related receptor alpha as a determinant of breast cancer heterogeneity.** *Cancer Res* 2009, **69**:6149-6157.
12. Laganière J, Deblois G, Lefebvre C, Bataille AR, Robert F, Giguère V: **From the Cover: Location analysis of estrogen receptor alpha target promoters reveals that FOXA1**

- defines a domain of the estrogen response.** *Proc Natl Acad Sci USA* 2005, **102**:11651-11656.
13. Lin C-Y, Vega VB, Thomsen JS, Zhang T, Kong SL, Xie M, Chiu KP, Lipovich L, Barnett DH, Stossi F, et al: **Whole-genome cartography of estrogen receptor alpha binding sites.** *PLoS Genet* 2007, **3**:e87.
  14. van der Meer DLM, Degenhardt T, Väisänen S, de Groot PJ, Heinäniemi M, de Vries SC, Müller M, Carlberg C, Kersten S: **Profiling of promoter occupancy by PPARalpha in human hepatoma cells via ChIP-chip analysis.** *Nucleic Acids Res* 2010, **38**:2839-2850.
  15. Chong HK, Infante AM, Seo Y-K, Jeon T-I, Zhang Y, Edwards PA, Xie X, Osborne TF: **Genome-wide interrogation of hepatic FXR reveals an asymmetric IR-1 motif and synergy with LRH-1.** *Nucleic Acids Research* 2010.
  16. Wederell ED, Bilenky M, Cullum R, Thiessen N, Dagpinar M, Delaney A, Varhol R, Zhao Y, Zeng T, Bernier B, et al: **Global analysis of in vivo Foxa2-binding sites in mouse adult liver using massively parallel sequencing.** *Nucleic Acids Res* 2008, **36**:4549-4564.
  17. Wei G-H, Badis G, Berger MF, Kivioja T, Palin K, Enge M, Bonke M, Jolma A, Varjosalo M, Gehrke AR, et al: **Genome-wide analysis of ETS-family DNA-binding in vitro and in vivo.** *The EMBO journal* 2010.
  18. Malhotra D, Portales-Casamar E, Singh A, Srivastava S, Arenillas D, Happel C, Shyr C, Wakabayashi N, Kensler TW, Wasserman WW, Biswal S: **Global mapping of binding sites for Nrf2 identifies novel targets in cell survival response through ChIP-Seq profiling and network analysis.** *Nucleic Acids Research* 2010.
  19. Schmidt D, Wilson MD, Ballester B, Schwalie PC, Brown GD, Marshall A, Kutter C, Watt S, Martinez-Jimenez CP, Mackay S, et al: **Five-vertebrate ChIP-seq reveals the evolutionary dynamics of transcription factor binding.** *Science* 2010, **328**:1036-1040.
  20. Ramos YFM, Hestand MS, Verlaan M, Krabbendam E, Ariyurek Y, van Galen M, van Dam H, van Ommen G-JB, den Dunnen JT, Zantema A, 't Hoen PAC: **Genome-wide assessment of differential roles for p300 and CBP in transcription regulation.** *Nucleic Acids Research* 2010.
  21. Gupta R, Wikramasinghe P, Bhattacharyya A, Perez FA, Pal S, Davuluri RV: **Annotation of gene promoters by integrative data-mining of ChIP-seq Pol-II enrichment data.** *BMC Bioinformatics* 2010, **11 Suppl 1**:S65.
  22. Tallack MR, Whittington T, Shan Yuen W, Wainwright EN, Keys JR, Gardiner BB, Nourbakhsh E, Cloonan N, Grimmond SM, Bailey TL, Perkins AC: **A global role for KLF1 in erythropoiesis revealed by ChIP-seq in primary erythroid cells.** *Genome research* 2010.

23. Consortium EP, Birney E, Stamatoyannopoulos JA, Dutta A, Guigó R, Gingeras TR, Margulies EH, Weng Z, Snyder M, Dermitzakis ET, et al: **Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project.** *Nature* 2007, **447**:799-816.
24. Boverhof DR, Burgoon LD, Tashiro C, Sharratt B, Chittim B, Harkema JR, Mendrick DL, Zacharewski TR: **Comparative toxicogenomic analysis of the hepatotoxic effects of TCDD in Sprague Dawley rats and C57BL/6 mice.** *Toxicol Sci* 2006, **94**:398-416.
25. Dere E, Boverhof DR, Burgoon LD, Zacharewski TR: **In vivo-in vitro toxicogenomic comparison of TCDD-elicited gene expression in Hepa1c1c7 mouse hepatoma cells and C57BL/6 hepatic tissue.** *BMC Genomics* 2006, **7**:80.
26. Puga A, Maier A, Medvedovic M: **The transcriptional signature of dioxin in human hepatoma HepG2 cells.** *Biochem Pharmacol* 2000, **60**:1129-1142.
27. Hayes K, Zastrow G, Nukaya M, Pande K, Glover E, Maufort J, Liss A, Liu Y, Moran S, Vollrath A, Bradfield C: **Hepatic transcriptional networks induced by exposure to 2,3,7,8-tetrachlorodibenzo-p-dioxin.** *Chem Res Toxicol* 2007, **20**:1573-1581.
28. Boutros PC, Yan R, Moffat ID, Pohjanvirta R, Okey AB: **Transcriptomic responses to 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD) in liver: comparison of rat and mouse.** *BMC Genomics* 2008, **9**:419.
29. Flaveny CA, Murray IA, Perdew GH: **Differential gene regulation by the human and mouse aryl hydrocarbon receptor.** *Toxicological sciences : an official journal of the Society of Toxicology* 2010, **114**:217-225.
30. Whitlock J: **Induction of cytochrome P4501A1.** *Annu Rev Pharmacol Toxicol* 1999, **39**:103-125.
31. Ahmed S, Valen E, Sandelin A, Matthews J: **Dioxin increases the interaction between aryl hydrocarbon receptor and estrogen receptor alpha at human promoters.** *Toxicol Sci* 2009, **111**:254-266.
32. Pansoy A, Ahmed S, Valen E, Sandelin A, Matthews J: **3-methylcholanthrene induces differential recruitment of aryl hydrocarbon receptor to human promoters.** *Toxicological sciences : an official journal of the Society of Toxicology* 2010.
33. Sartor MA, Schnekenburger M, Marlowe JL, Reichard JF, Wang Y, Fan Y, Ma C, Karyala S, Halbleib D, Liu X, et al: **Genomewide analysis of aryl hydrocarbon receptor binding targets reveals an extensive array of gene clusters that control morphogenetic and developmental programs.** *Environ Health Perspect* 2009, **117**:1139-1146.
34. Kinehara M, Fukuda I, Yoshida K-I, Ashida H: **High-throughput evaluation of aryl hydrocarbon receptor-binding sites selected via chromatin immunoprecipitation-**



- based screening in Hepa-1c1c7 cells stimulated with 2,3,7,8-tetrachlorodibenzo-p-dioxin.** *Genes Genet Syst* 2008, **83**:455-468.
35. Boverhof DR, Burgoon LD, Tashiro C, Chittim B, Harkema JR, Jump DB, Zacharewski TR: **Temporal and dose-dependent hepatic gene expression patterns in mice provide new insights into TCDD-Mediated hepatotoxicity.** *Toxicol Sci* 2005, **85**:1048-1063.
  36. Ji H, Jiang H, Ma W, Johnson DS, Myers RM, Wong WH: **An integrated software system for analyzing ChIP-chip and ChIP-seq data.** *Nat Biotechnol* 2008, **26**:1293-1300.
  37. Ji H, Wong W: **TileMap: create chromosomal map of tiling array hybridizations.** *Bioinformatics* 2005, **21**:3629.
  38. Rhead B, Karolchik D, Kuhn R, Hinrichs A, Zweig A, Fujita P, Diekhans M, Smith K, Rosenbloom K, Raney B, et al: **The UCSC Genome Browser database: update 2010.** *Nucleic Acids Research* 2010, **38**:D613.
  39. Portales-Casamar E, Thongjuea S, Kwon AT, Arenillas D, Zhao X, Valen E, Yusuf D, Lenhard B, Wasserman WW, Sandelin A: **JASPAR 2010: the greatly expanded open-access database of transcription factor binding profiles.** *Nucleic Acids Research* 2010, **38**:D105-110.
  40. Matys V, Fricke E, Geffers R, Gössling E, Haubrock M, Hehl R, Hornischer K, Karas D, Kel AE, Kel-Margoulis OV, et al: **TRANSFAC: transcriptional regulation, from patterns to profiles.** *Nucleic Acids Res* 2003, **31**:374-378.
  41. Mahony S, Auron PE, Benos PV: **DNA familial binding profiles made easy: comparison of various motif alignment and clustering strategies.** *PLoS Comput Biol* 2007, **3**:e61.
  42. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA: **Circos: an information aesthetic for comparative genomics.** *Genome research* 2009, **19**:1639-1645.
  43. Ma Q, Baldwin K: **2,3,7,8-tetrachlorodibenzo-p-dioxin-induced degradation of aryl hydrocarbon receptor (AhR) by the ubiquitin-proteasome pathway. Role of the transcription activaton and DNA binding of AhR.** *J Biol Chem* 2000, **275**:8432-8438.
  44. Song Z, Pollenz R: **Ligand-dependent and independent modulation of aryl hydrocarbon receptor localization, degradation, and gene regulation.** *Mol Pharmacol* 2002, **62**:806-816.
  45. Boutros PC, Moffat ID, Franc MA, Tijet N, Tuomisto J, Pohjanvirta R, Okey AB: **Dioxin-responsive AHRE-II gene battery: identification by phylogenetic footprinting.** *Biochem Biophys Res Commun* 2004, **321**:707-715.

46. Sogawa K, Numayama-Tsuruta K, Takahashi T, Matsushita N, Miura C, Nikawa J-i, Gotoh O, Kikuchi Y, Fujii-Kuriyama Y: **A novel induction mechanism of the rat CYP1A2 gene mediated by Ah receptor-Arnt heterodimer.** *Biochem Biophys Res Commun* 2004, **318**:746-755.
47. Richter K, Nessling M, Lichter P: **Macromolecular crowding and its potential impact on nuclear function.** *Biochim Biophys Acta* 2008, **1783**:2100-2107.
48. Li G-W, Berg OG, Elf J: **Effects of macromolecular crowding and DNA looping on gene regulation kinetics.** *Nature Physics* 2009, **5**:294.
49. Kobayashi A, Sogawa K, Fujii-Kuriyama Y: **Cooperative interaction between AhR.Arnt and Sp1 for the drug-inducible expression of CYP1A1 gene.** *The Journal of biological chemistry* 1996, **271**:12310-12316.
50. Dabir P, Marinic TE, Krukovets I, Stenina OI: **Aryl hydrocarbon receptor is activated by glucose and regulates the thrombospondin-1 gene promoter in endothelial cells.** *Circ Res* 2008, **102**:1558-1565.
51. Marlowe J, Fan Y, Chang X, Peng L, Knudsen E, Xia Y, Puga A: **The aryl hydrocarbon receptor binds to E2F1 and inhibits E2F1-induced apoptosis.** *Mol Biol Cell* 2008, **19**:3263-3271.
52. Klinge CM, Kaur K, Swanson HI: **The aryl hydrocarbon receptor interacts with estrogen receptor alpha and orphan receptors COUP-TFI and ERRalpha1.** *Arch Biochem Biophys* 2000, **373**:163-174.
53. Dennis G, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, Lempicki RA: **DAVID: Database for Annotation, Visualization, and Integrated Discovery.** *Genome Biol* 2003, **4**:P3.
54. Kopec AK, Boverhof DR, Burgoon LD, Ibrahim-Aibo D, Harkema JR, Tashiro C, Chittim B, Zacharewski TR: **Comparative toxicogenomic examination of the hepatic effects of PCB126 and TCDD in immature, ovariectomized C57BL/6 mice.** *Toxicol Sci* 2008, **102**:61-75.
55. Köhle C, Bock KW: **Coordinate regulation of Phase I and II xenobiotic metabolisms by the Ah receptor and Nrf2.** *Biochemical Pharmacology* 2007, **73**:1853-1862.
56. Nioi P, Hayes JD: **Contribution of NAD(P)H:quinone oxidoreductase 1 to protection against carcinogenesis, and regulation of its gene by the Nrf2 basic-region leucine zipper and the arylhydrocarbon receptor basic helix-loop-helix transcription factors.** *Mutation research* 2004, **555**:149-171.
57. Yeager RL, Reisman SA, Aleksunes LM, Klaassen CD: **Introducing the "TCDD-inducible AhR-Nrf2 gene battery".** *Toxicological sciences : an official journal of the Society of Toxicology* 2009, **111**:238-246.

58. Hamza MS, Pott S, Vega VB, Thomsen JS, Kandhadayar GS, Ng PWP, Chiu KP, Pettersson S, Wei CL, Ruan Y, Liu ET: **De-novo identification of PPARgamma/RXR binding sites and direct targets during adipogenesis.** *PLoS ONE* 2009, **4**:e4907.
59. O'Geen H, Squazzo SL, Iyengar S, Blahnik K, Rinn JL, Chang HY, Green R, Farnham PJ: **Genome-wide analysis of KAP1 binding suggests autoregulation of KRAB-ZNFs.** *PLoS Genet* 2007, **3**:e89.
60. Xu X, Bieda M, Jin VX, Rabinovich A, Oberley MJ, Green R, Farnham PJ: **A comprehensive ChIP-chip analysis of E2F1, E2F4, and E2F6 in normal and tumor cells reveals interchangeable roles of E2F family members.** *Genome research* 2007, **17**:1550-1561.
61. Li Q, Barkess G, Qian H: **Chromatin looping and the probability of transcription.** *Trends Genet* 2006, **22**:197-202.
62. Long X, Miano JM: **Remote control of gene expression.** *J Biol Chem* 2007, **282**:15941-15945.
63. Murray IA, Morales JL, Flaveny CA, Dinatale BC, Chiaro C, Gowdahalli K, Amin S, Perdew GH: **Evidence for ligand-mediated selective modulation of aryl hydrocarbon receptor activity.** *Molecular Pharmacology* 2010, **77**:247-254.
64. Giguère V: **Orphan nuclear receptors: from gene to function.** *Endocr Rev* 1999, **20**:689-725.
65. Viollet B, Kahn A, Raymondjean M: **Protein kinase A-dependent phosphorylation modulates DNA-binding activity of hepatocyte nuclear factor 4.** *Molecular and Cellular Biology* 1997, **17**:4208-4219.
66. Mietus-Snyder M, Sladek FM, Ginsburg GS, Kuo CF, Ladas JA, Darnell JE, Karathanasis SK: **Antagonism between apolipoprotein AI regulatory protein 1, Ear3/COUP-TF, and hepatocyte nuclear factor 4 modulates apolipoprotein CIII gene expression in liver and intestinal cells.** *Molecular and Cellular Biology* 1992, **12**:1708-1718.
67. Ladas JA, Hadzopoulou-Cladaras M, Kardassis D, Cardot P, Cheng J, Zannis V, Cladaras C: **Transcriptional regulation of human apolipoprotein genes ApoB, ApoCIII, and ApoAII by members of the steroid hormone receptor superfamily HNF-4, ARP-1, EAR-2, and EAR-3.** *The Journal of biological chemistry* 1992, **267**:15849-15860.
68. Pineda Torra I, Jamshidi Y, Flavell D, Fruchart J-C, Staels B: **Characterization of the Human PPAR{alpha} Promoter: Identification of a Functional Nuclear Receptor Response Element.** *Molecular Endocrinology* 2002, **16**:1013.

69. You M, Fischer M, Cho WK, Crabb D: **Transcriptional control of the human aldehyde dehydrogenase 2 promoter by hepatocyte nuclear factor 4: inhibition by cyclic AMP and COUP transcription factors.** *Arch Biochem Biophys* 2002, **398**:79-86.

## CHAPTER 7

## **CHAPTER 7**

### **CONCLUSIONS AND FUTURE RESEARCH**

The preceding studies have utilized a toxicogenomic approach to integrate disparate and complementary techniques to further elucidate the hepatic AhR regulatory network elicited by TCDD. These data have identified a small set of responsive orthologous genes between humans, mice and rats, with functions related to xenobiotic metabolism. Meanwhile, a significant number of elicited responses occurred a species-specific manner and DRE analysis of these orthologous genes identified differences in their promoter context suggesting that species-specific regulons may account for the lack of conserved responses between species. Furthermore, genome-wide AhR enrichment and DRE analyses in mouse hepatic tissue have globally mapped TCDD-induced AhR-DNA interactions that are both DRE-dependent and -independent, and observed in all regions of the genome. Enriched regions devoid of the DRE core sequence suggest that the binding motif of the activated AhR/ARNT heterodimer is more promiscuous than originally believed and/or there are alternative mechanisms of AhR gene regulation. Data integration and extrapolations across species on the basis of the conserved AhR mode of action suggest that the AhR directly mediates a small set of conserved responsive genes. However, the work embodied in this dissertation has expanded the hepatic AhR regulatory network in mice and similar approaches can be applied to other species, and to reduce the uncertainties in cross-species extrapolations and risk assessment in general.

#### **COMPARATIVE GENE EXPRESSION ANALYSIS**

TCDD-elicited toxicity occurs in a species-specific manner, and therefore a more thorough understanding of the conserved and non-conserved gene expression responses between

species is integral to understanding the underlying mechanism of action. The comparative cross-species analysis of TCDD-elicited gene expression responses in human, mouse and rat hepatoma cell lines described in Chapter 4 highlighted the similarities and differences between species. Furthermore, sequence analysis of DREs in the promoter regions of responsive orthologous genes suggest that species-specific regulon exist that may account for non-conserved and divergent responses between species. Although these *in vitro* models are routinely used to investigate AhR-mediated effects, data from Chapter 3 have shown that immortalized cell lines may not adequately reflect the responses occurring *in vivo*. Studies utilizing freshly isolated primary hepatocytes from human, mouse and rat liver tissues that addresses this concern are currently in progress. The results from the rodent hepatocytes can be compared with pre-existing *in vivo* data, analogous to comparisons performed in Chapter 3, to assess the validity of primary hepatocytes as an *in vitro* model. Additionally, comparative cross-species analyses of the microarray data are expected to identify species-conserved and -divergent gene expression responses that may be responsible for the differences in sensitivity and differential *in vivo* toxicity observed between rodents and humans.

## **GLOBAL AHR ENRICHMENT ANALYSIS**

Recent promoter-focused and genome-wide ChIP-chip studies, including those presented in Chapter 6, have demonstrated that AhR enrichment occurs in the promoter region of both TCDD-responsive and non-responsive genes, and also in the intergenic regions of the genome. Furthermore, sequence analysis of the AhR enriched regions found that the DRE core sequence is not necessary for AhR enrichment. However, current ChIP-chip arrays are limited in terms of their probe coverage that results in relatively large unrepresented gaps in the genome and make it difficult to precisely map the location of all sites of AhR enrichment. ChIP coupled with next

generation sequencing (ChIP-seq) alleviates these problems and should be considered in future studies to provide a more comprehensive analysis of TCDD-elicited AhR-DNA interactions.

Additionally, AhR enrichment analysis in primary human hepatocytes elicited by TCDD would further enhance the current understanding of AhR gene regulation. Comparative analysis of the enrichment profiles in the primary hepatocytes with those identified in mouse hepatic tissue would help delineate the underlying cause for species differences in sensitivity to TCDD between mice and humans. Integrating these data with the ongoing expression profiling in primary human hepatocytes and genome-wide DRE analyses will provide mechanistic insight into the hepatic AhR regulatory network within humans. Collectively, these proposed studies would assist in identifying mechanistically based biomarkers for TCDD-induced hepatotoxicity in humans.

## **AHR INTERACTIONS WITH OTHER TRANSCRIPTION FACTORS**

Previous studies have demonstrated that the AhR can regulate gene expression responses in coordination with other transcription factors and receptors, including the hepatic nuclear factor 4, alpha (HNF4 $\alpha$ ). HNF4 $\alpha$  is known to mediate the expression of *Cyp7a1* and *Gck*, genes that have important roles in lipid and cholesterol metabolism in the liver. For instance, *Cyp7a1* expression is protective of high-fat diet induced obesity, fatty liver and insulin resistance [1]. Meanwhile, *Gck* is responsible for phosphorylating glucose during the initial step of glycolysis and mutations in *Gck* result in reduced kinase activity, which is associated with insulin resistance and maturity onset diabetes of young 2 (MODY2) in humans [2-4]. Microarray analysis detected the down-regulation of both of these genes in C57BL/6 mice and may contribute to their TCDD-induced hepatic vacuolization and lipid accumulation. Interestingly, the AhR has also been reported to interact with COUP-TF, a TF that antagonizes HNF4 $\alpha$ -mediated responses by



binding HNF4 $\alpha$  response elements. Studies have demonstrated that AhR interactions with COUP-TF inhibit estrogen-mediated gene expression. *In silico* TF binding site (TFBS) analysis of TCDD-induced AhR enrichment identified over-representation of both COUP-TF and HNF4 binding motifs. Combined, these data suggest that the AhR may interact with COUP-TF to regulate lipid and fatty acid metabolism by blocking HNF4 $\alpha$  target gene expression. COUP-TF antibodies are readily available that could be used to further investigate TCDD-induced co-localization and/or tethering of COUP-TF with the AhR using re-ChIP assays. Additionally, integrating these studies with *in silico* HNF4 $\alpha$  response element searches would further characterize the potential AhR/COUP-TF cross-talk mechanism involved in TCDD-induced hepatotoxicity.

## REFERENCES

## REFERENCES

1. Li T, Owsley E, Matozel M, Hsu P, Novak CM, Chiang JYL: **Transgenic expression of cholesterol 7 $\alpha$ -hydroxylase in the liver prevents high-fat diet-induced obesity and insulin resistance in mice.** *Hepatology* 2010, **52**:678-690.
2. Caro JF, Triester S, Patel VK, Tapscott EB, Frazier NL, Dohm GL: **Liver glucokinase: decreased activity in patients with type II diabetes.** *Horm Metab Res* 1995, **27**:19-22.
3. Stoffel M, Froguel P, Takeda J, Zouali H, Vionnet N, Nishi S, Weber IT, Harrison RW, Pilkis SJ, Lesage S: **Human glucokinase gene: isolation, characterization, and identification of two missense mutations linked to early-onset non-insulin-dependent (type 2) diabetes mellitus.** *Proc Natl Acad Sci USA* 1992, **89**:7698-7702.
4. Marks SD, Couch RM: **Identification of two new mutations in the glucokinase gene that result in maturity-onset diabetes of the young.** *Diabetes Care* 2010, **33**:e94.