MIT/WHOI 2003-19

Massachusetts Institute of Technology
Woods Hole Oceanographic Institution

Joint Program in Oceanography/ Applied Ocean Science and Engineering

DOCTORAL DISSERTATION

ZEBRAFISH CARDIOVASCULAR cDNA MICROARRAYS: EXPRESSION PROFILING AND GENE DISCOVERY IN EMBRYOS EXPOSED TO 2,3,7,8-TETRACHLORODIBENZO-*P*-DIOXIN

by

Heather Martin Handley

September 2003

DISTRIBUTION STATEMENT A
Approved for Public Release
Distribution Unlimited

BEST AVAILABLE COPY

20040810 009

MIT/WHOI
2003-19

ZEBRAFISH CARDIOVASCULAR cDNA MICROARRAYS: EXPRESSION PROFILING AND GENE DISCOVERY IN EMBRYOS EXPOSED TO 2,3,7,8-TETRACHLORODIBENZO-*P*-DIOXIN

by

Heather Martin Handley

Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

and

Woods Hole Oceanographic Institution
Woods Hole, Massachusetts 02543

September 2003

DOCTORAL DISSERTATION

Funding was provided by an NSF Coastal Research Traineeship, NIH Superfund Basic Research Program grant 5-P42-ES07381, EPA grant R827102-01-0, and WHOI Ocean Life Institute and Coastal Ocean Institute grant 39591300 and the WHOI Academic Programs Office.

Reproduction in whole or in part is permitted for any purpose of the United States Government. This thesis should be cited as: Heather Martin Handley, 2003. Zebrafish Cardiovascular cDNA Microarrays: Expression Profiling and Gene Discovery in Embryos Exposed to 2,3,7,8–Tetrachlorodibenzo-p-dioxin. Ph.D. Thesis. MIT/WHOI, 2003-19.

Approved for publication; distribution unlimited.

Approved for Distribution:

John J. Stegeman, Chair
Department of Biology

Paola Malanotte-Rizzoli
MIT Director of Joint Program

John W. Farrington
WHOI Dean of Graduate Studies

ZEBRAFISH CARDIOVASCULAR cDNA MICROARRAYS: EXPRESSION PROFILING AND GENE DISCOVERY IN EMBRYOS EXPOSED TO 2,3,7,8-TETRACHLORODIBENZO-*P*-DIOXIN

by

Heather Martin Handley
B.Sc., James Cook University (1997)

submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

and the

WOODS HOLE OCEANOGRAPHIC INSTITUTION

September 2003

© 2003 Woods Hole Oceanographic Institution
All rights reserved.

Signature of Author _____
Joint Program in Oceanography
Massachusetts Institute of Technology and
Woods Hole Oceanographic Institution
September 20003

Certified by _____
John J. Stegeman
Thesis Supervisor

Accepted by _____
John Waterbury
Chair, Joint Committee for Biological Oceanography
Woods Hole Oceanographic Institution

"Success is going from failure to failure without a loss of enthusiasm."

-Winston Churchill

Acknowledgements

I was twelve years old when I first set my heart on a Ph.D. from Woods Hole Oceanographic Institution. So I think I should begin by thanking everyone in Academic Programs, especially John Farrington, Judy McDowell, and Julia Westwater, for helping me make that dream a reality.

I need to thank my thesis supervisor, John Stegeman, for constant intellectual stimulation and scientific guidance. Even more important, though, has been the unflagging support and encouragement he has provided. The hugs and pep talks are uncountable, and kept me from walking away more than once. Thank you, John.

Mark Fishman welcomed me into his lab and, in doing so, really made this thesis work possible. Mark, your vision and dedication are inspiring.

Mark Hahn, Nancy Hopkins, and Iain Drummond have guided and supported my research. Their insightful questions and constructive criticisms are greatly appreciated.

I have led a dual life, jumping back and forth between Boston and Woods Hole, for nearly four years now. Thank you to **all** of the members of both Stegeman and Fishman labs for making both places feel like home. In particular, Bruce Woodin, Margaret Boulos, and Amy Dougherty have made it possible for me to be scientifically productive, even when I barely knew my way around the lab or was altogether absent! They have also become dear friends.

Matthew Grow and I were nearly inseparable through a grueling year of tedious technical work that I absolutely could not have done on my own. Ashok Srinivasan provided the cDNA libraries, without which there would be no microarrays. Tyler Aldredge and the staff at the Harvard Center for Genomics Research provided invaluable technical support. Randy Peterson and Jordan Shin have each taken me under their wing at one time or another. Jordan – your encouragement and your sense of humor have both been greatly appreciated!

Ann Michelle and Ru Morrison, Joanna Wilson, Andrew McArthur, Amy Kukulya, Catherine and Brant Davis, Ann and Ray Jones, Chris Morse, Brian Tivol, Rose Koch

and Brian Litofsky – you are esteemed colleagues and great friends. Thank you for many wonderful meals together. Your support and understanding have meant so much to me.

Mom and dad – I do not think that any parents in the world could have done more to foster curiosity and instill the deep respect and passion for learning that I carry with me. You have always encouraged me to follow my dreams, and your love has given me the strength to do so.

Jed, my best friend, soon to be my husband – You have been beside me from day one of this long, rocky trip called graduate school. Your love and support kept me going when my own enthusiasm failed.

This research was supported in part by the WHOI Academic Programs Office, an NSF Coastal Research Traineeship, NIH Superfund Basic Research Program grant 5-P42-ES07381, EPA grant R827102-01-0, and WHOI Ocean Life Institute and Coastal Ocean Institute grant 39591300.

ZEBRAFISH CARDIOVASCULAR cDNA MICROARRAYS: EXPRESSION PROFILING AND GENE DISCOVERY IN EMBRYOS EXPOSED TO 2,3,7,8-TETRACHLORODIBENZO-*P*-DIOXIN

by Heather Martin Handley

submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy

Abstract

2,3,7,8-Tetrachlorodibenzo-*p*-dioxin (TCDD) is a potent teratogen that impacts the developing cardiovascular system. Hallmarks of embryonic exposure include cardiac malformation, impaired circulation, loss of erythrocytes, pericardial and yolk sac edema, and early life stage mortality. However, the mechanism of TCDD cardiovascular embryotoxicity is poorly understood. The primary goal of this thesis was to identify TCDD-responsive genes likely to be involved in processes of toxicity.

We constructed microarrays using cDNA libraries derived from zebrafish embryonic and adult heart tissue. Embryonic heart arrays were used for protocol development. The resulting workflow was employed in the production of adult heart microarrays containing ~2800 unique cardiovascular genes.

These arrays were used to establish gene expression profiles of zebrafish embryos exposed to 1.84±0.42 or 10.74±0.1.38 ng TCDD/g embryo. Alterations in cardiovascular gene expression were limited; 44 genes or ESTs were significantly differentially expressed $\geq$1.8-fold (p-values $\leq 5 \times 10^{-4}$), and only CYP1A and CYP1B1 were induced >4-fold. Transcriptional responses to TCDD were highly dose-dependent, and adaptive responses were a prevalent feature of TCDD-modulated gene expression.

Microarray analyses indicated induction of genes in three major functional classes – xenobiotic detoxification, sarcomere structure, and energy transfer. TCDD-modulation of selected genes was verified by RT-PCR. Induction of mitochondrial electron transfer genes was variable and modest; such induction provides a possible pathway to reactive oxygen generation and cardiac pathology. Sarcomere genes were generally robustly induced, but RT-PCR indicated suppression of cardiac troponin T2. The current data suggest that TCDD causes cardiomyopathy in zebrafish embryos.

Investigation of a TCDD-induced EST cluster led to the discovery of a novel retroelement, EZR1. EZR1 elements lack genes necessary for autonomous retrotransposition, but are highly expressed in normal and TCDD-exposed cardiac tissue. Putative regulatory elements in LTR sequences may account for observed expression patterns. The function, if any, of EZR1 remains open to speculation.

Thesis Supervisor: John J. Stegeman, Senior Scientist and Chairman of Biology, WHOI

Table of Contents

Title Page 1
Acknowledgements 5
Abstract 7
Table of Contents 9
List of Figures 14
List of Tables 20

Chapter 1 Introduction and background 23
1.1 Planar halogenated aromatic hydrocarbons 24
1.2 TCDD cardiovascular embryotoxicity 25
Cardiovascular impacts in fish 26
Dilative cardiomyopathy in avian embryos 28
1.3 Molecular mechanism of TCDD teratogenesis 29
Role of aryl hydrocarbon receptor 29
Role of cytochrome P450 1A 30
Missing links 32
1.4 Toxicogenomics and TCDD embryotoxicity 32
TCDD expression profiling 34
1.5 Contributions of this thesis 35
Objectives and rationale 35
Thesis content 36

Chapter 2 Development of zebrafish cardiovascular cDNA microarrays 49
Abstract 49
2.1 Introduction 50

2.2 Methods 53
cDNA Libraries 53
PCR 53
PCR product purification 54
Microarray Production and Processing 55
Quality Control 56
RNA and cDNA preparation 56
Microarray hybridization 58
2.3 Results 58
cDNA probe sets 58
Proof of concept – MAZF001 61
Heart-Tox Chips (MAZF002 & 003) 62
PCR product purification 62
Target cDNA Preparation 63
AH001 & AH001A 64
Microarray Quality Control 64
Slide processing 66
AH002A & B 67
Data analysis tools 67
2.4 Discussion 68
Microarray production 68
Adult heart arrays 69
Microarray quality control 70
Dye bias and labeling protocols 71
Experimental design 71
Conclusions 72

Chapter 3 The gene expression profile of 2,3,7,8 tetrachlorodibenzo-p-dioxin in zebrafish embryos is consistent with dilated cardiomyopathy 121
Abstract 121
3.1 Introduction 122
3.2 Methods 124
Embryos and Chemicals 124
DNA Microarray Hybridizations 125
Microarray Data Analysis 126
Real-time RT-PCR 127
3.3 Results 127
TCDD cardiovascular embryotoxicity 127
Experimental design and sources of variability 128
≥2-fold differential expression 129
Statistical confidence and systematic variance 129
TCDD-induced differential gene expression 131
Follow-up by real-time RT-PCR 132
3.4 Discussion 134
Limits of detection 134
General trends in TCDD-modulated gene expression 135
Xenobiotic responsive gene expression 137
Perturbed cellular energetics 137
Altered cardiovascular gene expression 138
TCDD-regulated ESTs 140
Conclusions 140

Chapter 4 EZR1: A novel unorthodox LTR-retroelement in zebrafish (*Danio rerio*) . 179
Abstract 179
4.1 Introduction 180

4.2 Methods 182
Zebrafish Embryos 182
in situ Hybridization 182
RNA Preparation 182
5' RACE and RT-PCR 183
DNA Sequencing 183
Radiation Hybrid Mapping 184
Sequence Analysis 184
4.3 Results 185
Characterization of TR004 transcripts 185
Putative EZR1 LTR structure 187
Genomic distribution of EZR1 188
Cardiac expression of EZR1 189
Putative LTR regulatory elements 190
4.4 Discussion 191
A novel class of non-autonomous LTR retroelements 191
LTR regulation of EZR1 expression 193
EZR1 induction by TCDD 194
Biological implications of EZR1 expression 195

Chapter 5 Conclusions and future work 217
5.1 Summary 218
5.2 Zebrafish adult heart cDNA microarrays 218
5.3 Expressed Zebrafish Retroelement 1 220
5.4 TCDD-induced dilated cardiomyopathy 222
5.5 TCDD and reactive oxygen species (ROS) 223
5.6 Alternative mechanisms of gene expression regulation 225
5.7 Conclusions 226

Appendix A Expression of vascular endothelial growth factor in early zebrafish embryos is unaffected by 2,3,7,8-tetrachlorodibenzo-p-dioxin exposure 227
A.1 Introduction 228
A.2 Methods 229
Embryos and RNA 229
Competitive RT-PCR 230
A.3 Results 232
Standard RT-PCR 232
Competitive RT-PCR 232
A.4 Discussion 233

Appendix B Preliminary evaluation of cross-species hybridization efficiency 249
B.1 Introduction 250
B.2 Methods 250
B.3 Results 251
B.4 Discussion 251

Appendix C The role of cytochrome P450 1A in TCDD embryotoxicity: preliminary results 259
C.1 Introduction 260
C.2 Methods 260
C.3 Results 261
C.4 Discussion 261

Literature Cited 267

List of Figures

Figure 1.1 Generic structure and substituent numbering system for halogenated dibenzo-*p*-dioxins (a), dibenzofurans (b), and biphenyls (c). 38

Figure 1.2 Relative sensitivity of vertebrate species to TCDD-induced early life-stage mortality, as indicated by LC_{50} (ng/g embryo) or LD_{50} (ng/g body weight) values. 40

Figure 1.3 Hallmarks of normal cardiovascular development (top) and TCDD-induced cardiovascular embryotoxicity (bottom) in zebrafish. 42

Figure 1.4 Schematic representation of AHR signaling, including activation by TCDD, nuclear translocation, dimerization with ARNT, and heterodimer binding to a consensus DRE sequence 44

Figure 1.5 Diagram of the CYP1A catalytic cycle showing hypothesized pHAH-induced uncoupling and reactive oxygen production, as well as possible points of reactive oxygen-mediated enzyme inactivation. 46

Figure 2.1 Scans of two AH001 microarrays hybridized with whole embryo cDNAs showing general lack of hybridization signal from regions containing genomic PCR products for mitochondrial, housekeeping, and many known cardiovascular clones (white rectangles). 88

Figure 2.2 PCR products for mitochondrial (a) and housekeeping (b) genes, amplified using primers described in Tables 2.5 and 2.6 90

Figure 2.3 MAZF001 microarray hybridized with 1 μg direct-labeled cDNA from zebrafish adult heart tissue (Cy5, red) and from whole adult zebrafish minus heart tissue (Cy3, green) 92

Figure 2.4 Comparison of results from duplicate features on the MAZF001 hybridization shown in Figure 2.3 94

Figure 2.5 Comparison of yields from isopropanol precipitations and filter purifications of ten PCR products. 96

Figure 2.6 Dye-swapped pair of AH001 hybridizations comparing cDNA from 72 hpf larval zebrafish exposed to either 1.7 ng/g or 1.1 ng/g TCDD.. 98

Figure 2.7 Comparison of fluorescent signal generated by homotypic hybridizations with (a) 500 ng amino-allyl post-labeled cDNA from adult heart total RNA or (b) 1 μg direct-labeled cDNA from adult heart mRNA. 100

Figure 2.8 Fluorescence intensity scatter plots for homotypic hybridizations using amino-allyl post-labeling of 1 μg cDNA from whole adult zebrafish (a, MAZF001) or 365 ng cDNA from adult heart tissue (b, AH001). 102

Figure 2.9 Ratio-intensity plots for homotypic hybridizations. 104

Figure 2.10 Effect of relative humidity during arraying on feature morphology, as indicated by salt autofluorescence (grey bars) or PicoGreen staining (black bars) 106

Figure 2.11 Quantitation of arrayed DNA by PicoGreen staining 108

Figure 2.12 Laser-excited fluorescence scan of a Syto22-stained AH001A microarray (top), and comparison of feature morphology observed by Syto22-staining to that observed on two randomly selected AH001A hybridizations (bottom). 110

Figure 2.13 Comparison of background fluorescence observed when arrays were processed and pre-hybridized according to four alternative protocols 112

Figure 2.14 Schematic representation of the data flagging process implemented by Flagger.pl. 114

Figure 2.15 Screen-dump of graphical user interface for microarray data analysis macros 116

Figure 2.16 Schematic diagram of microarray synthesis and hybridization workflow, showing techniques chosen for each production step and appropriate quality control measures. 118

Figure 3.1 Dose-response curves for TCDD-induced pericardial edema (a) and impairment of caudal circulation (b), as observed at 80 hpf 160

Figure 3.2 Schematic representation of experimental designs for TCDD exposure and microarray analyses. 162

Figure 3.3 Lack of effect of cDNA quantity on microarray gene expression results 164

Figure 3.4 Inverse relationship between degree of difference in embryo TCDD levels and correlation between low-dose replicate hybridizations (red, squares) and high-dose replicates (blue, diamonds) 166

Figure 3.5 Percentage of ≥2-fold changes made up by genes with known functions (black, with selected genes listed inset), ESTs with no significant similarity to known genes (dark grey), clones whose identity was undetermined due to low quality sequence (light grey), and clones that were found to have no insert (white) 168

Figure 3.6 Summary of gene expression results filtered according to a statistical confidence threshold of p-values $\leq 5 \times 10^{-4}$ 170

Figure 3.7 Frequency distribution of magnitude of statistically significant mean expression ratios at 0.5nM (gray) and 5.0nM TCDD (black). 172

Figure 3.8 Summary of TCDD-induced significantly differential expression. 174

Figure 3.9 Gene expression ratios for selected genes, as determined by real-time RT-PCR. 176

Figure 4.1 Schematic illustration of EZR1 sequence assembly (top), including TR004 ESTs (black), PCR (blue) and 5' RACE (red) products, and the resultant consensus sequence (bottom) 200

Figure 4.2 Multiple sequence alignments taken from major variable regions spanning positions 2567 to 2690 (a) and 3689 to 4114 (b) in the EZR1 sequence assembly.. 202

Figure 4.3 Dot-plot comparison of the complete EZR1 consensus sequence to zebrafish genomic sequence contig ctg9483, positions 90,000-95,000. 206

Figure 4.4 Schematic illustration of the putative EZR1 LTR structure, showing U3, R, and U5 domain boundaries and the complete R domain sequence (inset).. 208

Figure 4.5 Most probable tree morphology, as determined by Bayesian inference of phylogeny, for EZR1 LTR sequences..... 210

Figure 4.6 Representative photographs showing spatial distribution of EZR1 expression in 72hpf zebrafish embryos, as visualized by *in situ* hybridization with anti-sense probes..... 212

Figure 4.7 Schematic depiction of canonical domain structures for vertebrate endogenous retroviruses (a), LTR retrotransposons (b), zebrafish bhikhari (c), and EZR1 (d) . 214

Figure A.1 Example of competitive RT-PCR data, showing linear best-fit trendlines and equations used to calculate absolute copy number for target sequences 238

Figure A.2 Relative expression levels of vascular endothelial growth factor (striped) and cytochrome P450 1A (solid) in control and TCDD-injected zebrafish embryos.... 240

Figure A.3 Absolute β-actin mRNA expression levels, as measured in two replicates of competitive RT-PCR..... 242

Figure A.4 Absolute quantitation by of VEGF and CYP1A mRNA levels in control and TCDD-injected zebrafish embryos. 244

Figure A.5 Mean VEGF and CYP1A mRNA levels expressed as a proportion of β-actin values. 246

Figure B.1 Representative quadrants from zebrafish (which arrays) hybridized with cDNA from zebrafish heart tissue (left panel) or from *Fundulus heteroclitus* heart tissue (right panel)..... 254

Figure B.2 Correlation between Cy3 feature intensities generated by hybridization of AH001 arrays with either zebrafish or *Fundulus heteroclitus* adult heart cDNA.. 256

Figure C.1 CYP1A morphant phenotype, as observed at 50 hpf..... 264

List of Tables

Table 2.1 Names, sequences, and melting temperatures of primers used for PCR amplification of clones from embryonic and adult heart cDNA libraries. 74

Table 2.2 PCR primer sequences and fragment sizes for nuclear receptors and cytochrome P450s included on cDNA microarrays..... 76

Table 2.3 Clone information for fragments of cytochromes P450, nuclear receptors, and related genes contributed by various researchers..... 78

Table 2.4 Clones for genes involved in cell-cycle regulation and apoptosis..... 80

Table 2.5 PCR primer sequences and product lengths for zebrafish mitochondrial genes included on AH002 cDNA microarrays. 82

Table 2.6 Highly expressed housekeeping genes included on microarrays as biological negative controls 84

Table 2.7 *Arabidopsis thaliana* genes used as negative controls on zebrafish microarrays 86

Table 3.1 Primer sequences and PCR product information for real-time RT-PCR analyses. 142

Table 3.2 [^{3}H]TCDD levels (ng/g) in treated embryos, as determined by liquid scintillation counting..... 144

Table 3.3 Pairwise correlation coefficients for replicate hybridizations for 0.5nM TCDD samples (a) and 5.0nM TCDD samples (b) 146

Table 3.4 Clone information and mean expression ratios for positive and negative control features. 148

Table 3.5 Microarray expression data for redundantly represented genes determined to be differentially expressed at p-value $\leq 5x10^{-4}$. These genes fell into three functional classes – mitochondrial genes (a), sarcomeric proteins (b), and ESTs (c) 150

Table 3.6 Identities and mean expression ratios of genes whose expression was altered by embryonic TCDD exposure. 152

Table 4.1 Predicted transcription factor binding sites in putative EZR1 LTR sequences. 198

Table A.1 Number of embryos and total tissue weight in flash-frozen samples from which total RNA was isolated..... 236

CHAPTER 1

Introduction and background:

2,3,7,8-Tetrachlorodibenzo-*p*-dioxin cardiovascular embryotoxicity

1.1 Planar halogenated aromatic hydrocarbons

2,3,7,8-Tetrachlorodibenzo-*p*-dioxin (TCDD), often called the most toxic man-made chemical, is the archetypical halogenated aromatic hydrocarbon (HAH). Halogenated aromatic hydrocarbons constitute a large class of toxicologically important synthetic chemicals, including polychlorinated dibenzodioxins, dibenzofurans, and biphenyls (Figure 1.1). Of particular concern are laterally halogenated congeners, such as TCDD and PCBs bearing substitutions in positions 2-6 (and/or 2'-6'). These planar HAH (pHAH) are highly persistent in the environment and more potent toxicants than non-coplanar counterparts. The biological effects of pHAH in vertebrates include severe epithelial disorders, thymic atrophy and thyroid dysfunction, tumor promotion, endocrine disruption, and developmental abnormalities [1].

HAH are primarily anthropogenic in origin and have become ubiquitous contaminants in aquatic environments. Polychlorinated biphenyls (PCBs) were manufactured for industrial use as lubricants, coolants, diluents, and plasticizers. Dioxins and furans have never been deliberately produced, but are common contaminants in organochlorine syntheses (e.g., Agent Orange and PCBs) and are formed as by-products of chlorinated bleaching processes. Industrial processes and accidental spills have resulted in localized, high-level HAH contamination of certain inland and coastal waters (e.g., New Bedford Harbor, MA [2]). Recently, large-scale incineration of waste material, particularly chlorinated plastics, has become the leading source of HAH and has contributed significantly to their global distribution via atmospheric transport.

Planar HAH are largely recalcitrant to biological or chemical degradation and, due to their hydrophobicity, may be accumulated to high concentrations in animals' lipid stores. TCDD has been found in fish tissues at concentrations hundreds of thousands of times those found in the surrounding environment. A recent survey by the Environmental Protection Agency found TCDD in fish at 70 percent of 388 sites, with observed body burdens as high as 204 pg/g wet weight [3]. At least one dioxin or furan congener was found in fish at 89 percent of all sites surveyed. Similarly, pHAH have been found in soil and water, as well as fish, bird, and human tissues from around the globe [4-6]. The

widespread presence of pHAH in biological tissues warrants concern about potential health effects.

The embryotoxic effects of pHAH are of particular interest due to the potential for long-term adverse impacts on individual fitness and population success. Indeed, the importance of understanding processes of developmental and reproductive toxicity has been addressed in recent reports from the National Research Council [7, 8]. Given their proximity to human population centers, animals inhabiting coastal and inshore aquatic environments may be uniquely susceptible to the effects of anthropogenic pollutants. In the case of PCBs, 97 percent of the environmental burden is found in the coastal and open ocean [4]. Understanding toxicological impacts and mechanisms in aquatic organisms is a pressing problem, and will remain so as industrialization and human population levels in coastal regions continue to increase. This is particularly true for teleost fish, which are among the most sensitive of all animals to early life stage mortality caused by TCDD (Figure 1.2).

1.2 TCDD cardiovascular embryotoxicity

TCDD and other pHAH are potent developmental toxicants that target the cardiovascular system. The hallmarks of embryonic TCDD exposure are edema, hemorrhage, craniofacial malformations, and early life stage mortality. This suite of symptoms, similar to blue sac syndrome in salmonid fish, has been observed in over a dozen fish species exposed to TCDD and related pHAH [9-15]. The avian equivalent, GLEMEDS (Great Lakes embryo mortality, edema, and deformities syndrome), has been described in embryos of chicken, turkey, and several other domestic bird species experimentally treated with pHAH [16-20]. Edema and craniofacial deformities have been observed in pHAH-exposed rat, hamster, and guinea pig embryos [21].

While susceptibility to cardiovascular impacts by TCDD varies greatly, the similarity of TCDD-induced syndromes across taxa suggests that a common mechanism may be involved. In considering possible mechanisms of TCDD cardiovascular embryotoxicity, it is important to clearly define two terms – embryotoxicity and teratogenesis. Embryotoxicity includes all adverse effects of toxicant exposure, regardless of

originating mechanism. In contrast, teratogenesis refers to the production of (usually) irreversible morphological malformations by specific disruption of a normal developmental event. Edema and hemorrhage are generally thought to be secondary manifestations of an underlying teratogenic impact on the developing cardiovascular system. Significant progress has been made toward unraveling the sequence and relatedness of teratogenic and embryotoxic events in zebrafish (*Danio rerio*) and in the chick embryo.

Cardiovascular impacts in fish

The timing of onset of specific embryotoxic endpoints has been closely scrutinized in zebrafish, and may provide clues regarding causative teratogenic events (Figure 1.3). The first overt sign of TCDD toxicity is congestion and reduced circulation in peripheral vascular beds. Subtle, transient reductions in red blood cell perfusion rate in the brain and trunk can be detected as early as 48 hours post fertilization (hpf), approximately concurrent with hatching [22, 23]. By 60-72 hpf, blood flow in the tail is significantly slowed and blood begins to pool in the caudal vein [24]. Pericardial edema is first observable at 72 hpf, followed by yolk sac edema several hours later [10]. These conditions increase in severity both time- and dose-dependently, culminating in complete circulatory failure. The relative timing and progression of circulatory failure and edema is similar in other fish species, including Japanese medaka and rainbow trout [14, 25]. In zebrafish, circulatory impairment is exacerbated by gradual loss of erythrocytes at 80-96 hpf, resulting from disruption of definitive erythropoiesis [24].

TCDD does not appear to impact early cardiovascular patterning events, as the window of susceptibility for cardiovascular toxicity falls between 48 and 96 hpf. The suite of circulatory impacts described above can be produced by exposure of zebrafish embryos to TCDD at any point up to 48 hpf, and onset is only slightly delayed when exposure occurs at 72 hpf [24]. In contrast, exposure at or after 96 hpf produces no effect on cardiovascular performance. Thus, TCDD appears to be specifically modulating processes that take place between 48 and 96 hpf (e.g., definitive erythropoiesis). Similarly, in Japanese medaka, the window of susceptibility for cardiovascular effects by

TCDD is limited to the period during which cardiovascular lesions are observed, on days 4 and 5 of development [14].

It has been proposed that circulatory failure might result from malformation of blood vessels. This hypothesis was based on the similarity of the TCDD-induced phenotype to that of genetic mutants with vascular defects, such as *cloche*, and the sensitivity of vascular endothelial cells to enzyme induction, apoptosis, and morphological alteration caused by TCDD [26-28]. However, in accordance with the observed window of susceptibility, the molecular pathways responsible for vasculogenesis are unaffected prior to 48 hpf, and blood vessel size, number, and patterning is normal in TCDD-treated zebrafish [24]. Vascular damage appears to be, in itself, a toxic endpoint of TCDD exposure that may constitute *one* proximal mechanism for generation of edema. However, it is unlikely to be the cause of circulatory failure.

Alternatively, it is possible that edema and hemorrhage might be secondary effects of circulatory failure caused by cardiac deficiencies. Indeed, edema is a common phenotype among zebrafish genetic mutants with cardiac defects [29, 30], and there is some evidence to suggest that TCDD impacts cardiac morphology and function in developing fish. As expected, early cardiac development, including heart tube formation and cardiac looping, appears to be unaffected by TCDD. Heart rate also remains normal until after 96 hpf (Figure 1.3), at which point reduced heart rate is most likely a reflection of impending mortality. However, TCDD treatment results in reduced contractile strength as early as ~50 hpf [Handley, unpublished data], and reductions in heart size are apparent by 72 hpf [31]. Significant reduction in heart size has also been observed in TCDD-exposed sac fry of rainbow trout [25]. There has been some debate regarding whether such cardiac derangements might be attributable solely to physical forcing by pericardial edema. However, rearing TCDD-treated zebrafish embryos in iso-osmotic sugar solutions reduces edema without rescuing the cardiac or circulatory phenotypes [31].

Thus, there are preliminary indications that TCDD exerts a direct teratogenic impact on cardiac growth and muscle development between 48 and 96 hpf. Unfortunately, this phase of zebrafish cardiac development has received little attention and is poorly

understood. At hatching, cardiac looping is complete and the four chambers of the teleost heart (sinus venosus, atrium, ventricle, and bulbus arteriosus) are distinct. By 96 hpf, the zebrafish heart is essentially "adult." Presumably, the intervening period is one of proliferation and, possibly, further differentiation. However, given the current lack of knowledge, it is difficult to speculate as to the nature of processes that might be impacted by TCDD during this period.

Dilative cardiomyopathy in avian embryos

In the avian embryo, edema and hemorrhage are secondary effects of TCDD-induced dilative cardiomyopathy. At dose levels low enough to avoid significant edema or hemorrhage, TCDD causes enlargement of the heart due to increase in the size of ventricular luminal cavities, but not in ventricular wall thickness (i.e., dilation) [32]. Prior to observable dilation, cardiomyocyte proliferation is inhibited and apoptosis is increased in specific cardiac structures [33]. As cardiotoxicity progresses, the number and size of coronary arteries is reduced, atrial natriuretic factor mRNA expression is induced, β-adrenergic chronotropic (heart rate) responsiveness is suppressed, and subcutaneous edema is observed [32, 33]. Overall, these observations are consistent with TCDD-induced dilated cardiomyopathy that leads to congestive heart failure.

There appears to be a window of susceptibility for cardiotoxic impacts that coincides with a period of myocardial remodeling in the embryonic avian heart. Signs of TCDD cardiotoxicity (molecular or morphological) are not manifest until Day 8 (D8), reach maximal severity by D12, and cannot be induced by treatment on or after D14 [33, 34]. In normal avian development, D8-D10 is a period of extraordinary proliferation and rearrangement of ventricular myocytes [35]. The outer, compact layer of ventricular myocardial cells quadruples in thickness in this two-day period, before settling into a more moderate rate of growth. This thickening triggers increased coronary artery invasion. Also during this period, the compact layer of myocardial cells develops a highly organized, multi-layer system of spiral myocardial fibers. This new architecture is necessary to maintain increased hemodynamic pressure. Based on current understanding

of TCDD cardiotoxicity in the chick, it seems likely that TCDD specifically blocks the process of ventricular compact layer thickening, and possibly remodeling.

Processes of cardiac remodeling are poorly understood in lower vertebrates (i.e., fish). The adult morphology of teleost hearts is extremely variable. Many large, fast-swimming fish, such as tuna, have hearts with similar myocardial architecture and extensive coronary vascularization. Others, like zebrafish, are so small that these elaborations are unnecessary. However, given the fact that *some* fish develop cardiac muscle morphology comparable to that of higher vertebrates, it seems likely that homologous (if simplified) genetic pathways exist in fish. Furthermore, the similarity in phenotype and ontogeny of TCDD cardiovascular embryotoxicity across taxa suggests a common underlying mechanism.

1.3 Molecular mechanism of TCDD teratogenesis

Role of aryl hydrocarbon receptor

The aryl hydrocarbon receptor (AHR) is a basic-helix-loop-helix Per-ARNT-Sim (bHLH-PAS) protein that functions as a ligand-activated transcription factor with a broad affinity for aromatic hydrocarbons [36]. AHR homologs have been identified in most animal lineages, including arthropods, nematodes, bivalves, agnathans, cartilaginous and bony fishes, amphibians, reptiles, birds, and mammals [37]. The mechanism of ligand-activated AHR signaling is well understood, and is highly conserved across vertebrate taxa [36-38]. Following ligand binding, cytosolic AHR undergoes a conformational shift that facilitates release of cofactors, including hsp90 and Ara9, and translocation of AHR into the nucleus. Nuclear AHR interacts with aryl hydrocarbon receptor nuclear translocator (ARNT) protein to form a heterodimeric transcription factor that binds enhancer sequences known variously as AHR-, dioxin-, or xenobiotic-response elements (AHRE, DRE or XRE) (Figure 1.4).

AHR plays a significant role in cardiovascular development and function in vertebrates. AHR is highly expressed in the heart and vasculature of fish and birds [39-42]. AHR-null knock-out mice manifest transient alterations in fetal and neonatal cardiac

morphology, as well as progressive hypertension and cardiac hypertrophy beginning soon after birth [43-45]. It is interesting that the murine AHR-null phenotype is, in many ways, opposite of that resulting from activation of AHR by pHAH. Similarly, AHR expression is increased in hearts of human patients suffering ischemic or dilative cardiomyopathy [46].

Transcriptional modulation by AHR is the primary means by which TCDD effects toxicity. The toxic potency of specific aromatic hydrocarbons is strongly correlated to their ability to bind and activate AHR [47]. Strain-dependent or inter-specific differences in TCDD sensitivity also depend largely on properties of AHR [48-50]. For example, AHR expression and signaling is altered in strains of the salt-marsh killifish, *Fundulus heteroclitus*, which have acquired heritable resistance to PCBs and other HAH [51, 52]. Furthermore, chemical antagonists and genetic knock-down/out technologies have provided direct evidence of the necessity of AHR for TCDD toxicity in zebrafish embryos [22, 53, 54] and in both embryonic and adult mice [55-59].

Role of cytochrome P450 1A

While AHR is capable of regulating expression of numerous genes, inducing cytochrome P450 1A (CYP1A) gene expression appears to be the primary means by which AHR mediates HAH toxicity. Induction of CYP1A enzymes by aromatic hydrocarbons was first reported more than thirty years ago [60, 61], and has since been shown to be strictly AHR-dependent [62, 63]. CYP1A proteins are phase I xenobiotic metabolizing enzymes whose primary function is oxidative modification of hydrophobic organic substrates. Such metabolism is intended to facilitate elimination of exogenous toxicants from the cell, but may have alternative consequences, such as bioactivation or reactive oxygen production [64], that can contribute to toxicity. It has long been thought, based on correlations between CYP1A induction and pHAH-induced toxic impacts, that CYP1A enzymes may be involved in processes of pHAH toxicity.

Patterns of evolutionary variation in CYP1A gene complement can be correlated to susceptibility to pHAH toxicity. Whereas AHR is present in most animals, biochemical, molecular, and bioinformatics-based surveys have failed to identify CYP1A genes in any

invertebrate species. Correspondingly, pHAH exposure does not produce overt toxicity in invertebrates. The number and type of CYP1A genes present may also account for differential sensitivity to TCDD embryotoxicity among vertebrates. As a rule, teleosts possess a single CYP1A gene and are extremely sensitive to early life-stage toxicity by TCDD (Figure 1.2). In contrast, mammals possess two distinct CYP1A genes, CYP1A1 and CYP1A2, and are relatively insensitive to TCDD embryotoxicity and early life-stage mortality (Figure 1.2). Birds, which are also highly sensitive to TCDD (Figure 1.2), also have two CYP1A genes, CYP1A4 and CYP1A5 [65, 66]. However, as is implied by their names, these genes are the result of an avian-specific gene duplication and are not orthologues of mammalian CYP1A1 and CYP1A2 [66, 67].

Xenobiotic induction of CYP1A genes is also correlated with toxic impacts in vertebrates on temporal, spatial, and dose-dependent bases. For example, in lake trout embryos exposed to TCDD, CYP1A protein levels were greatly enhanced in vascular endothelium at the time of onset of cardiovascular malfunction, and the dose-dependence of CYP1A induction was closely correlated with that of sac fry mortality [27]. Furthermore, CYP1A induction in vascular endothelium co-localizes with regions of TCDD-induced apoptosis associated with embryotoxicity [26, 68]. CYP1A4 induction in chick embryos is similarly correlated with cardiovascular toxicity [69]. Finally, blocking CYP1A induction at the level of mRNA expression [70], protein expression [54], or enzymatic activity [22], protects against pHAH-induced toxicity.

The most likely mode of toxic action for CYP1A enzymes is production of reactive oxygen species (ROS). pHAH exposure results in elevated intracellular reactive oxygen levels and increased rates of oxidative damage in a variety of biological systems [64, 71-85], and these processes have been implicated in aspects of pHAH toxicity [26, 86, 87]. pHAH-induced reactive oxygen production is largely CYP1A-dependent, as evidenced by the reduction of oxidative stress in CYP1A-null knock-out mice exposed to TCDD [70]. It is thought that imperfect substrates, such as HAH with multiple lateral chlorine substitutions, become lodged in the CYP1A active site and uncouple the CYP1A catalytic cycle, causing production of superoxide radicals without subsequent substrate

oxygenation (Figure 1.5) [64, 88]. Alternatively, superoxide production could result from direct electron withdrawal by chlorine substituents. Superoxide may then be converted to hydrogen peroxide, a longer-lived species capable of diffusing from cell to cell and generating highly reactive hydroxyl radicals [Goldstone, pers. comm.].

Missing links

While nearly every aspect of cardiovascular development and function is exquisitely sensitive to intracellular oxygen conditions, the downstream effectors of TCDD embryotoxicity have remained elusive. The traditional candidate gene approach has, thus far, proved unsuccessful in this endeavor. For example, vascular endothelial growth factor and hipoxia-inducible factor 1α (VEGF and HIF1α) have been subject to intensive investigation based on the potential for modulation by reactive oxygen, or by ARNT-mediated cross-talk between AHR and HIF1α [89, 90]. However, competition for ARNT does not significantly impact downstream signaling by AHR and HIF1α [90-92], and TCDD-induced alterations in VEGF expression are extremely variable [93-95]. In TCDD-treated zebrafish, VEGF expression is unaffected up to 24 hpf (Appendix A). Thus, this line of investigation has been largely uninformative with regard to processes of TCDD embryotoxicity. The sheer abundance of possibilities may play a large role in obscuring the relevant pathways.

1.4 Toxicogenomics and TCDD embryotoxicity

The recent advent of genomics has revolutionized every area of the biological sciences, not least of all toxicology. In the past few years, the number of so-called 'omics' has grown steadily to include transcriptomics, proteomics, and metabolomics, to name a few. The abundance and rapid expansion of the 'omics' represents a widespread ideological shift toward examination of biological processes on broader scales than previously possible. This movement has been thoroughly adopted by the toxicology

community, which has coined its own 'omics', *toxicogenomics* [96]. Selkirk and Tennant [97] have defined toxicogenomics as

> "... a new scientific field that elucidates how the entire genome is involved in biological responses of organisms exposed to environmental toxicants/stressors. It combines information from studies of genomic-scale mRNA profiling, cell-wide or tissue-wide protein profiling (proteomics), genetic susceptibility, and computational models to understand the roles of gene-environment interactions in disease."

This definition, while not absolutely all-encompassing, stresses the variety of data types that might contribute to understanding a single toxicant or pathology. Similarly, Ballatori and colleagues [98] expressed the idea that disparate data sources might be combined under the umbrella of toxicogenomics in order to provide "... a unified framework for understanding the biochemical and genetic basis for various diseases."

While such a synthesis is still distant, the field of toxicogenomics is already making great strides in the areas of elucidating molecular mechanisms of toxicity and defining chemical-specific expression profiles [99, 100], with the ultimate goal of developing diagnostic and predictive biomarkers for pre-clinical, clinical, and environmental use [101-103]. Toward these ends, the National Center for Toxicogenomics, a recently established subsidiary of the National Institute for Environmental Health Sciences, has developed the Chemical Effects in Biological Systems database to house and integrate genomics, proteomics, and metabonomics data with conventional toxicological data [104].

Broad-scale gene expression profiling has become one of the primary tools employed in toxicogenomics research. Methods for assessing gene expression on a genomic scale include DNA microarrays, serial analysis of gene expression (SAGE) [105-107], differential display reverse transcriptase polymerase chain reaction (DD RT-PCR) [108, 109], and subtraction hybridization [110, 111]. The variety of gene expression profiling techniques facilitates adaptation to nearly any organism or question of interest. The use of DNA microarrays has become common among researchers studying human biology or model mammalian species. Subtraction hybridization and differential display RT-PCR

have been used to identify genes of interest for generation of DNA arrays in species for which genomic or mRNA sequence information is limiting [112, 113]. SAGE may be the method of choice in laboratories where high-throughput sequencing capacity is more readily available than the instrumentation required for microarray analysis. Nonetheless, DNA microarrays are probably the most commonly utilized gene expression profiling technology. Currently, the number of microarray publications exceeds those for any other gene expression profiling method by approximately 4:1.

TCDD expression profiling

The application of DNA microarray technology to the problem of understanding TCDD toxicity has vastly expanded the repertoire of known TCDD-responsive genes. To date, conventional methods have identified nearly 50 genes whose expression is modulated by TCDD exposure, many directly by AHR. While differential display RT-PCR and suppression subtractive hybridization are global in scope (i.e., no gene(s) is targeted *a priori*), the necessity for laborious follow-up work has limited the informational yield of such studies; nearly a dozen investigations have yielded a similar number of novel TCDD-regulated genes [114-122]. In contrast, five microarray studies and one SAGE experiment have identified several hundred TCDD-responsive genes [93-95, 123, 124]. Most TCDD-related gene expression profiling work has focused on liver tissue and cultured hepatocytes [93, 95, 123, 124], due to the primacy of liver in TCDD effects such as CYP1A enzyme induction. However, spleen and thymus tissues [95], and cultured lung epithelial cells [94] have also been interrogated.

Despite difficulties imposed by disparate gene representation among microarray platforms, comparison of gene expression profiling results is elucidating general trends in TCDD molecular responses. Multiple researchers have observed induction of plasminogen activator inhibitor I [93, 95, 123, 125], and metallothionein [93, 123, 124, 126, 127]. Metallothionein is known to have antioxidant activity, and may be expressed as part of a protective response to TCDD-induced reactive oxygen production. On a broader scale, TCDD appears to consistently perturb a multitude of basic cellular processes, including signal transduction (i.e., phosphorylation and Ca^{2+}), transcriptional

and translational machinery, cell cycle regulation and apoptosis, and fatty acid disposition.

Differences between gene expression profiling results are further emphasizing the complex, pleiotropic nature of TCDD impacts. In comparing three TCDD concentrations spanning two orders of magnitude, Martinez and colleagues [94]) found that more than half of all TCDD-regulated genes were differentially expressed at only one dose level and that many genes manifest non-traditional dose-response curves (e.g., induction at one dose, suppression at another). Cell- or tissue-type, and the state of cells with regard to tumorigenesis, also influence TCDD-responsiveness. For example, vascular endothelial growth factor (VEGF) expression was found to be increased in non-tumorigenic HPL1A lung cells, unchanged in the malignant, tumorigenic lung cell line A549, and decreased in HepG2 hepatoma cells [93, 94]. VEGF expression was induced in an isoform-specific manner in mouse thymus and liver tissues [95]. A wealth of other (as yet unexplored) biological factors, such as gender and age (developmental stage), are likely to be significant in shaping molecular responses to TCDD exposure.

Thus, determining a universal TCDD signature will require synthesis of gene expression profiles from numerous biological conditions. Conversely, elucidating the mechanism of TCDD toxicity in a particular system will require specific characterization of transcriptional responses in that system. This is especially true in the case of developmental toxicity, as the molecular and cellular complexity of embryogenesis cannot be mimicked by any *in vitro* system in existence.

1.5 Contributions of this thesis

Objectives and rationale

The goal of this work was to characterize the cardiovascular-specific gene expression profile of TCDD exposure in zebrafish embryos. In particular, it was hoped that identifying genes whose expression is altered during TCDD-induced cardiovascular

toxicity would begin to address outstanding questions in two areas of TCDD embryotoxicity:

1) What is the nature of cardiac teratogenesis in fish? How does this compare to effects seen in other species?
2) What is the molecular mechanism of TCDD embryotoxicity? Specifically, what are the downstream effectors of AHR and CYP1A?

The zebrafish (*Danio rerio*) was selected for this work based not only on its proven utility as a model for the study of developmental genetics, but also timely and dramatic increases in resources available for genetic and genomic research in zebrafish. The zebrafish has been subject to extensive embryological and genetic investigation over the past three decades, and has recently become a major model organism for toxicological work [128]. At the outset of this work, large-scale chemical mutagenesis screens were coming to fruition, providing a wealth of information about the roles of individual genes in cardiovascular development [29, 30, 129]. A number of EST sequencing projects were underway, and a genome sequencing project imminent. Thus, toxicogenomic investigation in zebrafish seemed feasible, timely, and relevant to a rapidly growing community of researchers.

Thesis content

cDNA Microarrays. Chapter 2 describes the design and construction of cardiovascular-specific cDNA microarrays, and work to optimize protocols for their use. This work has provided the technical groundwork necessary to allow toxicogenomic interrogation of TCDD embryotoxicity in zebrafish. In addition, as pre-fabricated arrays have only become commercially available in the past several months, these microarrays constituted a significant resource for the zebrafish community. As a result, several collaborations have developed around the use of these arrays for investigation of both effects of genetic mutations and mechanisms of toxicity in zebrafish; these projects are beyond the scope of this thesis, but are described briefly in Chapter 5. I have also explored the possibility that zebrafish microarrays might be suitable for use as

biomarkers of environmental contaminant exposure in fish via cross-species hybridization (Appendix B).

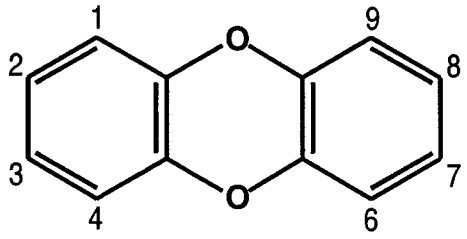
TCDD cardiovascular embryotoxicity. The central biological question driving this thesis, namely the nature of transcriptional responses to embryonic TCDD exposure, is addressed in Chapter 3. Gene expression profiling of zebrafish embryos exposed to two doses of TCDD revealed several general trends in TCDD-modulated transcription, such as a high level of dose-specificity and rather limited alterations in cardiovascular gene expression. This work also identified TCDD-induced changes in expression of cardiac sarcomere proteins and energy production enzymes that suggest dilated cardiomyopathy is likely in zebrafish embryos. Furthermore, TCDD exposure influenced expression of a number of ESTs with undetermined function; these ESTs are exciting in their potential for revealing novel aspects of TCDD toxicity.

Novel gene discovery. The work described in Chapter 4 follows from an unexpected result of cDNA microarray analyses (Chapter 3). The highly represented, TCDD-induced EST cluster TR004 was found to constitute a novel retroelement in the zebrafish genome, named EZR1. This finding was of toxicological interest, as induction of retrotransposons and endogenous retroviruses is coming to be associated with both environmental stress and a variety of disease states, including cardiac pathologies. Furthermore, the discovery of a previously unknown genetic element highlighted one of the major advantages of the chosen microarray strategy.

Future work. It is the nature of microarrays, indeed most high-throughput screening assays, to provide more questions than answers. The most important contribution of this thesis may well be the generated body of strong, observation-based hypotheses upon which further investigation of TCDD embryotoxicity can be built. Further discussion of the questions posed by this work, and possible approaches for addressing these questions, can be found in Chapter 5.

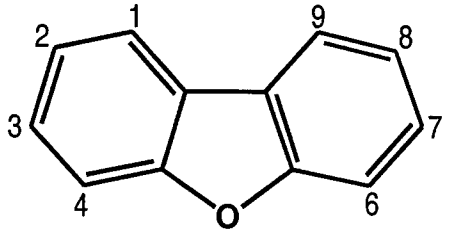
Figure 1.1 Generic structure and substituent numbering system for halogenated dibenzo-*p*-dioxins (a), dibenzofurans (b), and biphenyls (c).

a)



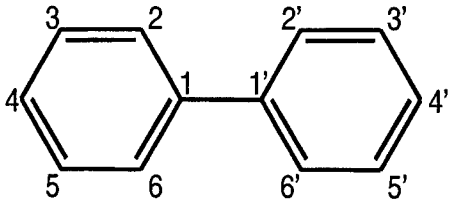
dibenzo-*p*-dioxin

b)



dibenzofuran

c)



biphenyl

Figure 1.2 Relative sensitivity of vertebrate species to TCDD-induced mortality, as indicated by LC_{50} (ng/g embryo, fish and chicken) or LD_{50} (ng/g body weight, mammals) values. Typical cytochrome P450 1A gene complement for each taxon is indicated by the color of bars (black = CYP1A1 and CYP1A2, striped = CYP1A4 and CYP1A5, grey = CYP1A). Data were taken from Poland and Knutson (1982) [21], Allred and Strange (1977) [130], Kennedy *et al.* (1996) [131], Henry *et al.* (1997) [10], Elonen *et al.* (1998) [9]. Where multiple measurements are available, standard deviation is shown.

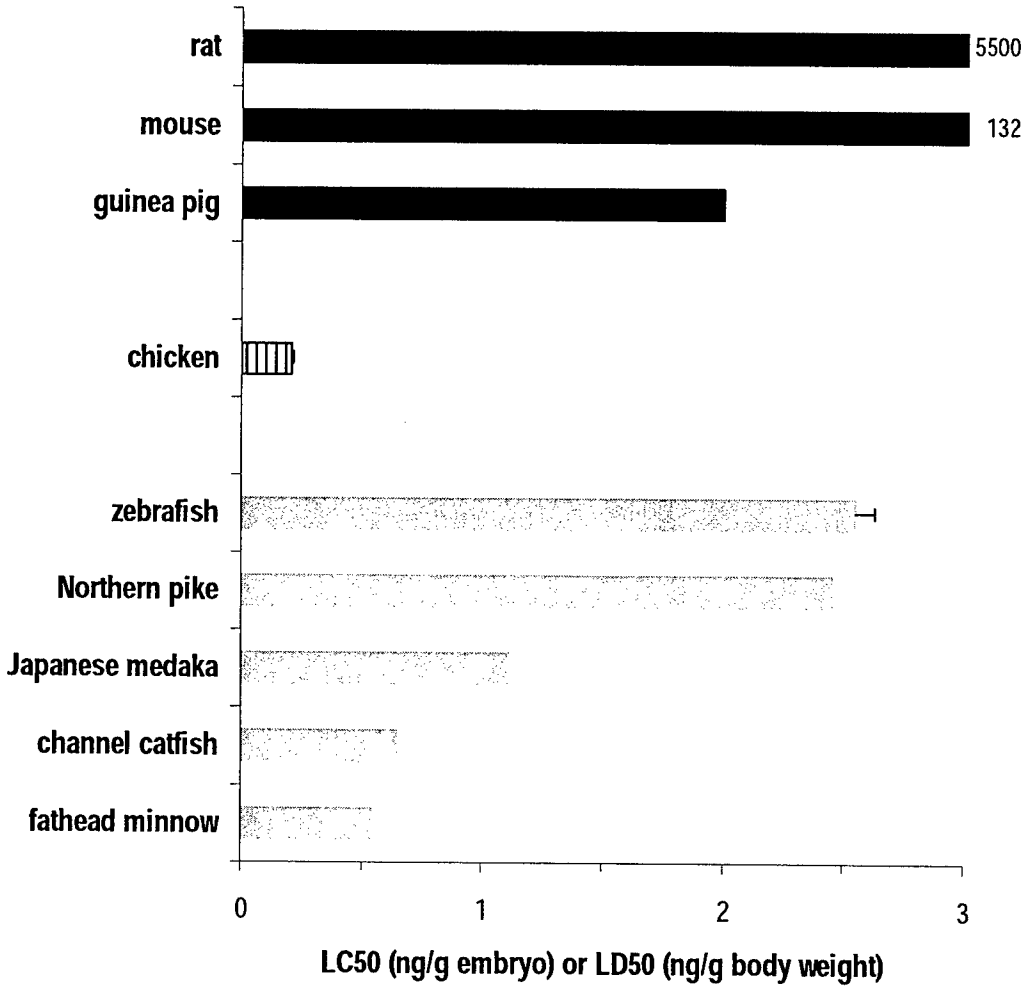

rat
mouse
guinea pig
chicken
zebrafish
Northern pike
Japanese medaka
channel catfish
fathead minnow
5500
132
0
1
2
3
LC50 (ng/g embryo) or LD50 (ng/g body weight)

Figure 1.3 Hallmarks of normal cardiovascular development (top) and TCDD-induced cardiovascular embryotoxicity (bottom) in zebrafish.

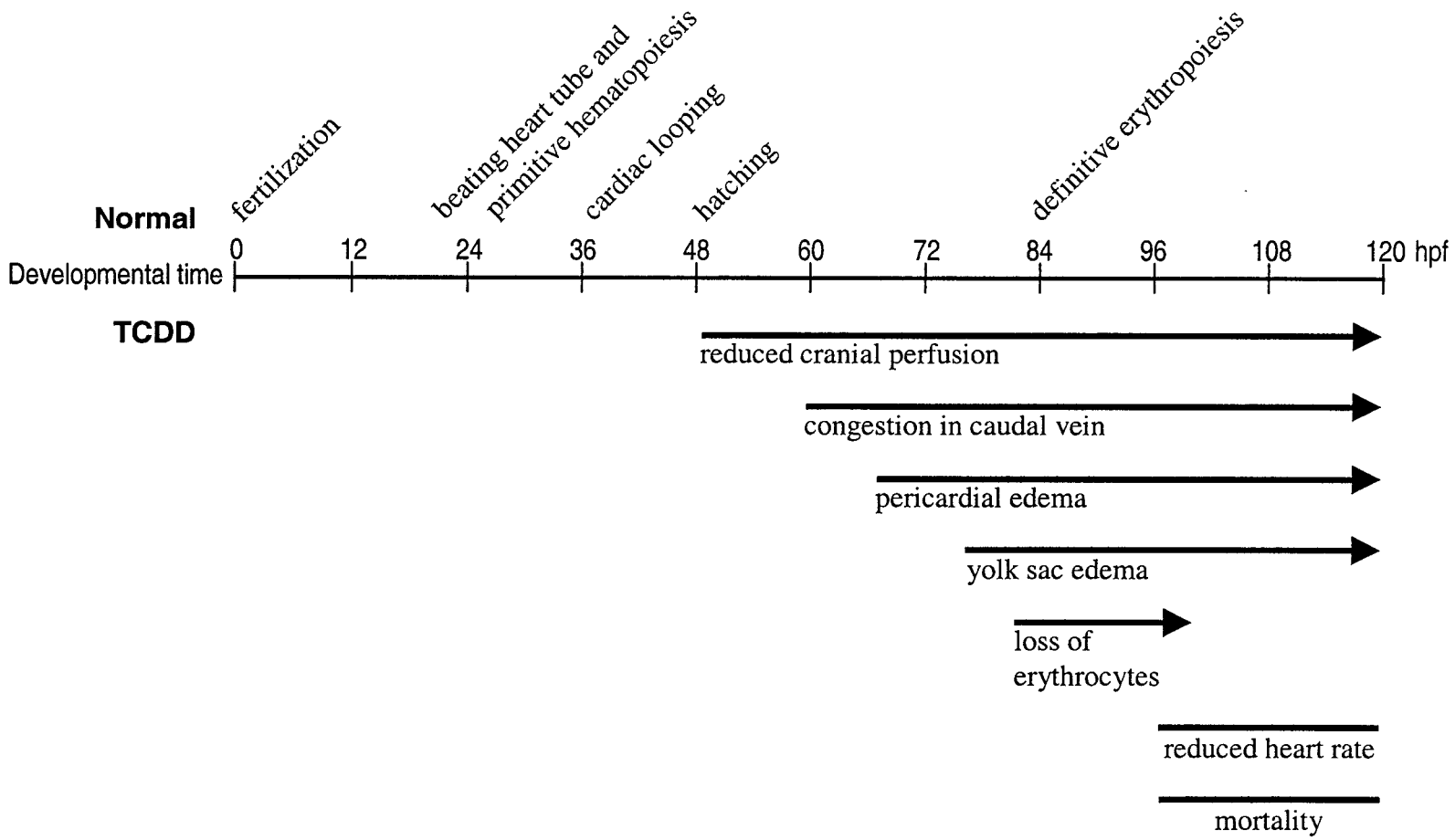

Normal
fertilization
beating heart tube and primitive hematopoiesis
cardiac looping
hatching
definitive erythropoiesis
Developmental time
0
12
24
36
48
60
72
84
96
108
120 hpf
TCDD
reduced cranial perfusion
congestion in caudal vein
pericardial edema
yolk sac edema
loss of erythrocytes
reduced heart rate
mortality

Figure 1.4 Schematic representation of AHR signaling, including activation by TCDD, nuclear translocation, dimerization with ARNT, and heterodimer binding to a consensus DRE sequence. The gray arrow indicates transcriptional activation of a downstream gene, such as CYP1A.

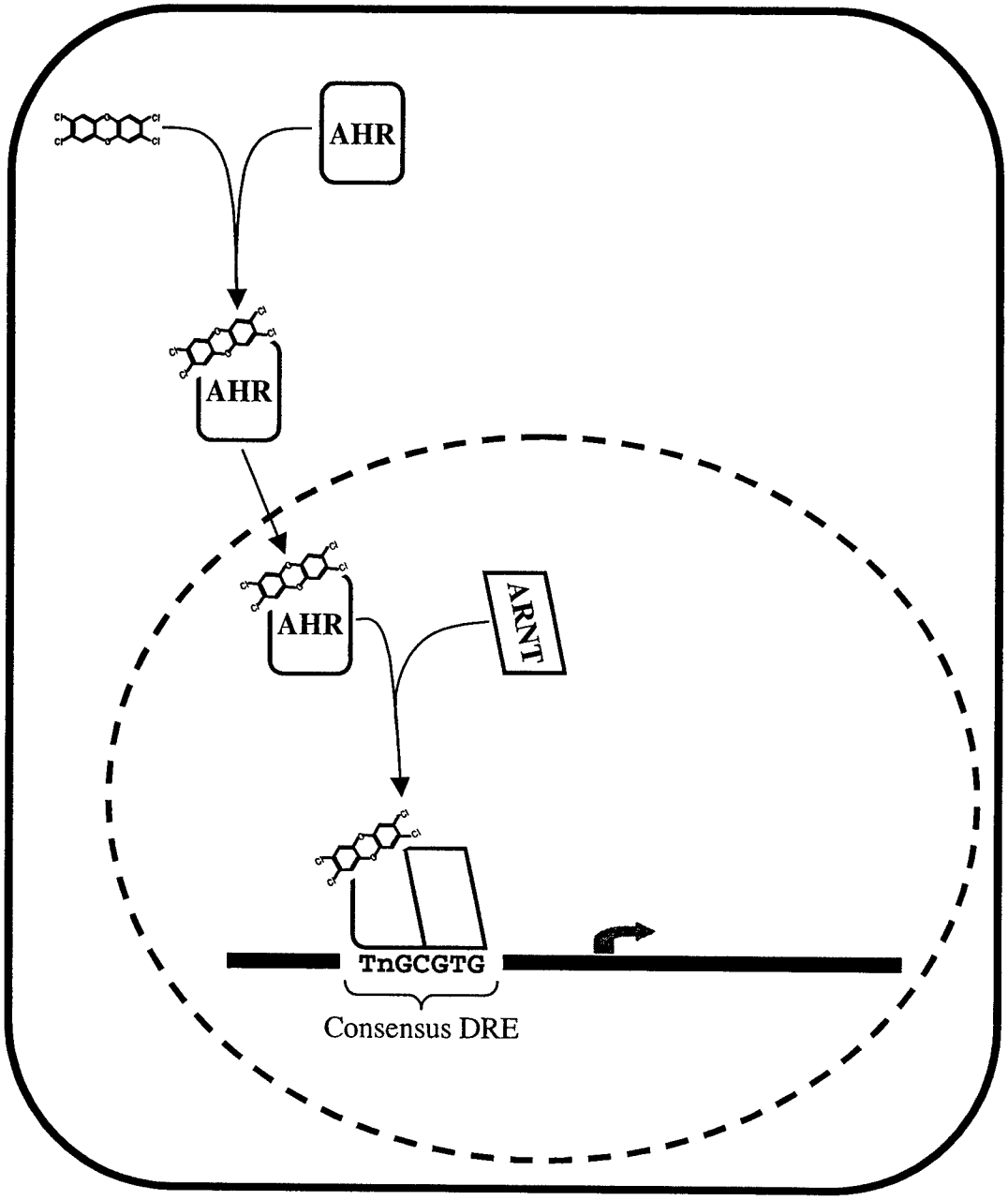
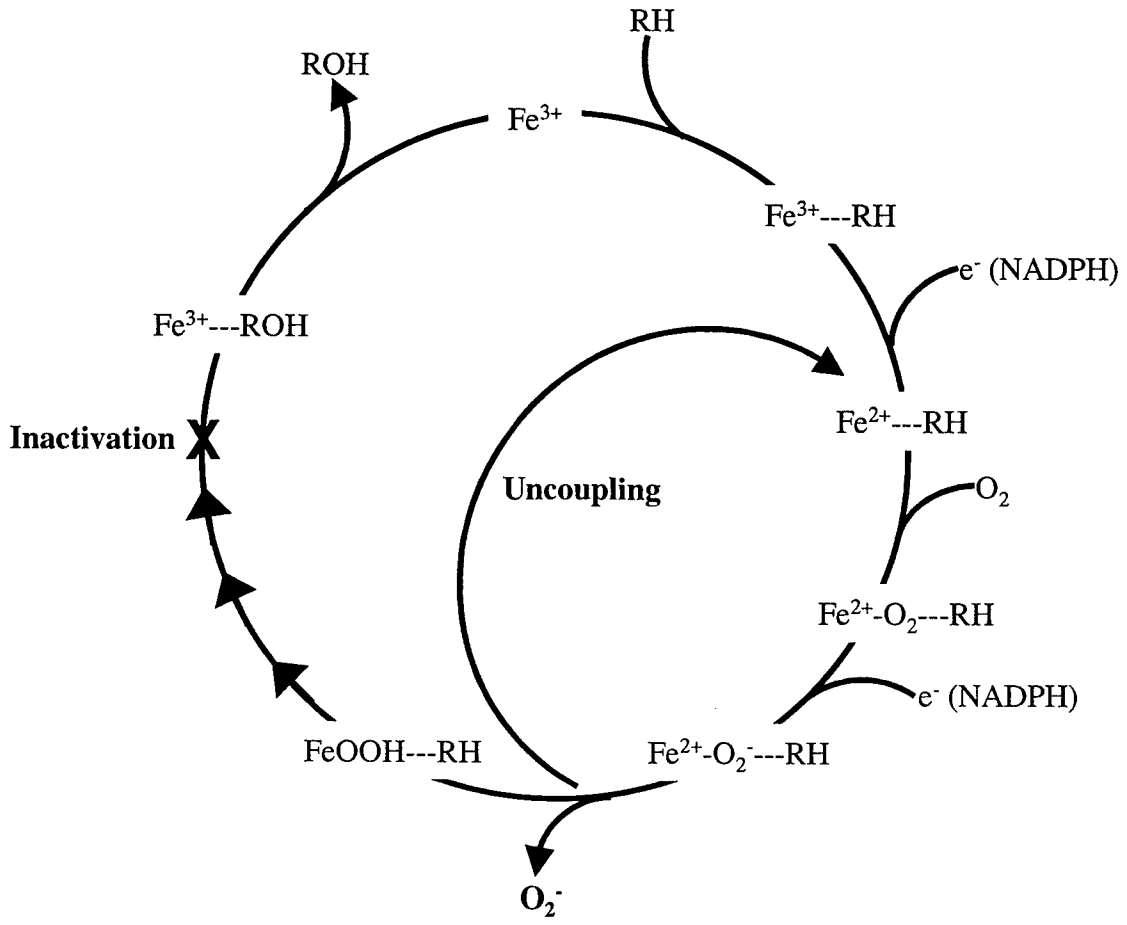

AHR
AHR
AHR
ARNT
TnGCGTG
Consensus DRE

Figure 1.5 Diagram of the CYP1A catalytic cycle showing hypothesized pHAH-induced uncoupling and reactive oxygen production, as well as possible points of reactive oxygen-mediated enzyme inactivation.


RH
ROH
Fe^{3+}
Fe^{3+}---RH
e^- (NADPH)
Fe^{3+}---ROH
Fe^{2+}---RH
Inactivation
Uncoupling
O_2
Fe^{2+}-O_2---RH
e^- (NADPH)
FeOOH---RH
Fe^{2+}-O_2^----RH
O_2^-

CHAPTER 2

Development of zebrafish cardiovascular cDNA microarrays

Abstract

DNA microarray technology has revolutionized the study of gene expression and transcriptional regulation. However, commercially available arrays are limited to a small number of species, such as yeast and human, for which there is an abundance of genome sequence data and a significant community of researchers to support the cost of technical development; zebrafish microarrays have only become available within the past year. In order to enable high-throughput analyses of cardiovascular gene expression in zebrafish embryos, we constructed spotted cDNA arrays using two cDNA libraries derived from embryonic and adult heart tissue.

We compared alternative protocols at several steps in the process of synthesizing and using cDNA microarrays. Filter purification was found to provide a superior method of cDNA probe purification over traditional isopropanol precipitations. Similarly, amino-allyl post-labeling of cDNA for microarray hybridizations manifest several advantages over traditional direct dye incorporation protocols. In contrast, most methods of processing microarrays and blocking background fluorescence performed comparably. Based on this work, we have compiled an effective workflow for cDNA microarray synthesis and hybridization. In addition, we developed methods for routine quality control during and after microarray synthesis. Our methodology is critically discussed in the context of the most recent developments in microarray technology.

In all, we have produced seven full-scale print lots, MAZF001-003, AH001/A, and AH002A/B. With the exceptions of MAZF002 and 003, which suffered from severe technical difficulties, all arrays manifest robust feature morphology and strong signal:noise ratios. MAZF001 arrays were primarily used for methodological development. AH001 and AH001A arrays contain 4896 adult heart cDNA clones likely to represent ~2800 unique cardiovascular genes; these arrays have been used for gene expression profiling of zebrafish embryos exposed to TCDD (Chapter 3). AH002 arrays contain a smaller collection of adult heart cDNAs (3456 clones, ~2000 genes) and are being used in ongoing investigations of genetic mutations and toxicological processes in zebrafish.

2.1 Introduction

The invention of DNA microarrays has been one of the most influential technical developments in the recent history of biology. DNA microarray technology was first developed less than a decade ago for the purpose of measuring gene expression levels on a high-throughput basis [132]. Since that time, DNA microarray platforms have been adapted to a number of other uses, including comparative genomic hybridization [133, 134], high-throughput identification of protein interaction sequences by chromatin immunoprecipitation (i.e., ChIp-on-chip) [135], and most recently, large-scale DNA sequencing. However, massively parallel detection of differential gene expression, so-called gene expression profiling, remains the most common use of DNA microarrays.

DNA arrays can be separated into three fundamental categories – membrane (macro)arrays, spotted microarrays, and *in situ* synthesized oligonucleotide arrays. Membrane arrays consist of DNA probes deposited, often manually, onto nylon or nitrocellulose membranes at relatively low densities. Because of issues of size and sensitivity, large quantities (i.e., several micrograms) of radiolabeled sample DNA may be required for membrane hybridizations. Thus, the utility of membrane arrays is primarily limited to situations in which the number of genes of interest is limited and sample tissue is abundant. In contrast, *in situ* synthesis of oligonucleotides can be used to generate arrays of stunningly high densities (e.g., $\leq$100,000 features per array), but the capability to manufacture these arrays is strictly limited to commercial facilities.

Spotted arrays consist of either cDNA fragments, generally PCR products, or long (40-80 bp) oligonucleotides deposited onto a solid substrate, usually a chemically coated glass microscope slide, using high-speed precision robotics. Probe sets (i.e., cDNA libraries or oligonucleotides) can be purchased commercially or produced in-house. Most major academic and research institutions now possess microarray spotting and scanning equipment, and numerous such commercial services are available. Depending on the robotics available, maximum feature density may reach 10-30,000 per microarray slide. Thus, spotted DNA arrays provide a flexible, widely accessible, intermediate-density platform for gene expression profiling.

While spotted oligonucleotide arrays are becoming increasingly popular, cDNA arrays present certain advantages. Extensive oligonucleotide design and quality control is required to avoid cross-hybridization with non-target sequences and produce consistent gene expression results [136, 137]. Large quantities of genomic and expressed sequence data are needed to provide sufficient material for high quality oligonucleotide design and to allow probe specificity to be assessed accurately. cDNA clones, which are generally at least 500 bp in length, appear to be more robust to slight changes in probe sequence or hybridization conditions [136] and are generally not susceptible to non-specific cross-hybridization with sequences that are >75-85% identical [138]. Furthermore, no difference in detection sensitivity between the two technologies has been shown [137]. Thus, with relatively little design input, cDNA arrays can provide sensitive, specific detection of gene expression.

The construction of cDNA microarrays requires five major steps: (1) selection of sequences of interest, (2) synthesis of cDNA probes by PCR, (3) purification of PCR products, (4) robotic arraying of cDNA probes, and (5) pre-hybridization processing of microarrays; this last step includes immobilization and denaturation of spotted probes, as well as treatment of slides to reduce background noise. At each step of the process, the technical options are myriad and there is currently no consensus regarding an "optimal protocol" for cDNA microarray synthesis. This is, in large part, because the best solution to any technical difficulty will depend on the specific questions to be addressed with microarrays, as well as available resources, financial and otherwise.

Of course, production of microarrays is only part of obtaining gene expression data; hybridization protocols and data analysis also present significant challenges. Quantitation of relative gene expression by microarray analysis is generally accomplished by competitive hybridization. RNA from two experimental samples, such as two tissue types, are used to make differentially fluorescently labeled cDNA; most often, cDNA is conjugated to Cy3 (green fluorescence) or Cy5 (red fluorescence). The two labeled cDNA populations, referred to as *target*, are hybridized to a microarray (immobilized probes), and the ratio of gene expression is inferred from the intensity of red and green

fluorescence at a given feature. The nature of RNA samples (i.e. total or messenger RNA), methods of Cy-dye conjugation, algorithms used for image analysis, and statistical manipulation of microarray datasets are all fodder for debate.

The goal of this work was to generate cDNA microarrays (and accompanying protocols) appropriate for investigation of processes of cardiovascular development in zebrafish (*Danio rerio*). In particular, these arrays were intended for use in interrogating mechanisms of embryonic cardiovascular dysfunction caused by chemical toxicity and genetic mutants. At the outset of this project, publicly available cDNA clone sets suffered from clone misidentification rates as high as 90%. Furthermore, the number of cloned, named genes from zebrafish was less than 1000, many of which had no known relevance to cardiovascular biology. Thus, for microarray probes, we relied heavily on two cDNA libraries, from embryonic and adult heart tissue, generated in-house. These libraries were supplemented with probes for genes with known roles in toxicological or developmental processes, as well as zebrafish housekeeping genes and *Arabidopsis thaliana* chloroplast genes to be used as controls.

In all, we have generated seven sets of zebrafish cardiovascular-specific cDNA microarrays, three with embryonic heart clones and four with adult heart cDNAs. In synthesizing these arrays and optimizing protocols for their use, we drew on information from the published literature, as well as personal communication with staff at nearby genomic research facilities and, of course, wet-lab comparisons of available methods and reagents. This chapter addresses data from technical comparisons, and provides vital statistics for probe sets and microarray print lots. An overview of the complete workflow developed is included, and the advantages of this approach are discussed with regard to the latest developments in microarray technologies.

2.2 Methods

cDNA Libraries

Embryonic heart

Preliminary work made use of a previously described zebrafish embryonic heart cDNA library [139]. Briefly, this library consisted of 5102 fully sequenced clones estimated to represent 3690 unique transcripts, including 1242 known genes. An aliquot of the complete gridded library (i.e. 5102 single-clone bacteriophage cultures) provided template material for PCR amplification (below).

Adult heart

Aliquots of DH5α *E. coli* cells transformed with an uncharacterized zebrafish adult heart cDNA library in were obtained from Dr. Ashok Srinivasan. These aliquots were spread onto plates of LB-Agar with 100 μg/ml ampicillin, and grown overnight at 37°C. Individual colonies were robotically picked (Genetix Q-bot) into 384-well microplates containing 65 μl Luria Broth with 100 μg/ml carbenicillin, with or without 1x HMF. Liquid cultures were grown overnight at 37°C, then sealed with adhesive foil and stored at either 4°C or -80°C. These single-clone bacterial cultures provided material for direct PCR amplification, as well as inoculation of larger volume cultures for preparation of plasmid DNA.

PCR

For print lot MAZF001, embryonic heart ESTs were amplified from 5 μl phage stock in 50 μl reactions containing 1x PCR buffer (1.5mM $MgCl_2$), 200μM each dNTP, 0.5μM each primer (Table 2.1), and 1.25U Taq polymerase (all reagents supplied by Qiagen). PCR conditions for all other microarray print lots were altered to provide a final concentration of 2.0mM $MgCl_2$. When amplifying directly from phage stocks or bacterial cultures, 2-5 μl was used for template. For clones in plasmids, 1 μl mini-prepped DNA was used.

Embryonic heart clones were amplified using either ZapExPCR-f and ZapExPCR-r primers (MAZF001) or T3 and T7 primers (MAZF002 and MAZF003), while adult heart library clones were amplified using universal SP6 and T7 primers (Table 2.1). An equimolar mixture of SP6, T3, and T7 primers was used to amplify additional clones in a variety of plasmids. PCR products for mitochondrial (Table 2.5) and housekeeping (Table 2.6) genes were obtained using gene-specific primers

For MAZF001, an initial DNA denaturation step (96°C, 5 min) was followed by 35 amplification cycles (30 sec at 94°C, 30 sec at 58°C, 3 min at 72°C) and a final 10 min extension period at 72°C. For later print lots, thermocycler conditions were altered slightly to increase PCR yields; extension time was limited to 2 min per cycle and 40 cycles were run.

PCR product purification

Isopropanol precipitation

PCR products were precipitated by addition of 1-2 volumes of cold isopropanol and ≥30 min at -20°C, followed by centrifugation for 45 min at 2,500 rpm. In some cases, NaCl was added to a final concentration of 200 mM prior to chilling and centrifugation. Supernatants were removed by aspiration, and DNA pellets were air-dried and reconstituted in aqueous spotting buffer (3x SSC + 0.1% Sarkosyl).

Filter purification

PCR products were filter-purified using Multiscreen-96 PCR Purification plates (Millipore). 100 μl reactions were transferred to filter plates. Vacuum pressure (650 mbar, or 20 inches Hg) was applied for 5 min, filter plates were blotted dry on paper towels, and vacuum pressure was applied for another 2 min. 100 μl nuclease-free water was added to each well, and DNA was resuspended by either repetitive pipetting or 10 min agitation at 500 rpm. Purified PCR products were removed to clean micro-well plates, dried by vacuum centrifugation (Savant Speed-Vac Concentrator), and reconstituted in spotting buffer (3x SSC + 0.1% Sarkosyl, or 50% DMSO).

Microarray Production and Processing

cDNA probes were printed onto CMT-GAPS slides (Corning) using one of three arrayers – GMS 417 (Genetic Microsystems), OmniGrid (GeneMachines), or a custom-build split-pin arrayer at the Harvard Center for Genomics Research. During arraying, probes were left at room temperature. Before and after, plates were stored at -20°C. Newly printed arrays were allowed to dry ≥30 min following the end of each print run before being transferred to storage cassettes. Arrays were stored in the dark in a room temperature dessication chamber.

Processing protocol #1

Arrays were individually held face-down over a steaming water bath ≤10 sec, then snap-dried by placing face-up on a 95°C heat block. Following this rehydration step, DNA was immobilized onto slides by UV cross-linking (Stratalinker, auto cross-linking function). Cross-linked slides were soaked for 15 min in a freshly prepared succinic anhydride/sodium borate solution (5 grams succinic anhydride in 315 mL of n-methyl-pyrrolidinone and 35 mL 0.2M sodium borate). Arrays held in a glass slide rack were placed in a larger glass container with a magnetic stir bar to provide gentle circulation. Upon removal from the succinate solution, arrays were washed 2 min each in 95°C nuclease-free water and 95% ethanol, then air-dried. Processed arrays were stored in darkness with dessication.

Processing protocol #2

Printed cDNAs were immobilized on the slide surface by UV cross-linking, as above, then washed 2 min each in 95°C nuclease-free water and 95% ethanol. Arrays were air-dried prior to storage in a dark dessication chamber.

Quality Control

PicoGreen staining

Aliquots (55 μl) of 1:200 dilutions of PicoGreen reagent (Molecular Probes) in 1x TE were pipetted onto the face of microarrays and allowed ~20 min in complete darkness to reach equilibrium binding. PicoGreen staining was visualized and photographed using high-power (2000x magnification) fluorescence microscopy coupled to a digital camera.

Syto22 staining

Syto22 dye (Molecular Probes) was diluted 1:100 in 1x TE, and 55 μl aliquots were pipetted onto microarrays and covered with glass cover slips. After ~1hr incubation at room temperature in complete darkness, slides were rinsed briefly in water and air-dried. Stained arrays were stored in darkness prior to laser-excited fluorescence scanning (Axon 4100B or 4200A).

RNA and cDNA preparation

Total RNA was extracted from embryo homogenates using TriZol reagent (Invitrogen) according to manufacturer's protocol. For long-term storage, RNA pellets were kept in 70% ethanol at -80°C. After removal of ethanol, total RNA was dissolved in water and stored frozen. mRNA was isolated from total RNA using the OligoTex mRNA system (Qiagen).

Direct Cy-dye incorporation

mRNA (1–2 μg) was spiked with *A. thaliana* chloroplast mRNA (100, 250, and 500 ng of Cab, RCA, and rbcL RNA, respectively; SpotReport®-3 Array Validation System, Stratagene) and incubated with 2μg oligo-$dT_{(20)}N$ primer for 10 min at 70°C, then chilled on wet ice. Reverse transcription reactions were run 2 hrs at 42°C in 1x first strand buffer plus 10mM DTT, 0.5mM dATP/dGTP/dTTP, 0.2mM dCTP, 0.3mM Cy-dCTP, and 400U Superscript II (Invitrogen). cDNA samples were rid of RNA contamination by alkaline hydrolysis (15 μl 0.1N NaOH added, 10 min at 70°C), then neutralized with 15

μl 0.1N HCl. TE (10mM Tris, 1mM EDTA) was added to a final volume of 500 μl, and cDNAs were purified using Centricon-30 microconcentrator columns (Amicon). Equal amounts of Cy3- and Cy5-labeled cDNA were combined in 3x SSC with 0.1% SDS.

Amino-allyl post-labeling

Total RNA (15–25μg) was spiked with *A. thaliana* chloroplast mRNA, as above, and incubated with 5μg oligo-$dT_{(20)}N$ primer for 10 min at 65°C, then chilled on wet ice. Remaining reagents were then added for final reaction conditions of 1x first-strand synthesis buffer, 10mM DTT, 0.5mM each dATP, dCTP and dGTP, 0.3mM dTTP, 0.2mM amino-allyl-dUTP, and 1000U Superscript II reverse transcriptase (Invitrogen). Reverse transcription reactions were run 2.5 hrs at 42°C, then inactivated by buffering with 0.5M EDTA and incubating 5 min at 95°C.

RNA was eliminated by alkaline hydrolysis in 0.2N NaOH (incubated 15 min at 65 °C, then neutralized by equimolar HCl and buffered with Tris-HCl), followed by RNase digestion (xU Ambion RNase cocktail, 30 min at 37°C). cDNA was filter-purified using QiaQuick PCR Purification columns (Qiagen), replacing Qiagen buffers PE and EB were replaced by 75% ethanol and distilled water, respectively. cDNAs were dried by vacuum centrifugation and stored at -20 °C.

For CyDye post-labeling, cDNAs were redissolved in 10 μl 0.1M NaHCO3 (pH9.0) containing an individual aliquot of previously dried amine-reactive Cy3 or Cy5 (Amersham Biosciences), then incubated 1.5-2 hrs at room temperature in full darkness. The labeling reaction was quenched by addition of excess hydroxylamine (4.5 μl at 4M) and 15 min at room temperature in full darkness. Following addition of 35 μl 100mM NaOAc (pH 5.2) and 50 μl nuclease-free water, labeled cDNA was purified using QiaQuick PCR Purification columns (Qiagen) according to standard protocols. cDNA concentrations were determined spectrophotometrically (A_{260}, A_{280}), then equal quantities of paired Cy3- and Cy5-labeled cDNAs were combined and dried by vacuum centrifugation.

Microarray hybridization

Processed microarrays were pre-hybridized 2-4 hrs at 65°C in 5x SSC with 0.1% SDS, 10 mg/ml bovine serum albumin (or casein), and 0.1mg/ml sonicated salmon sperm DNA. Shortly prior to hybridization, arrays were removed from pre-hybridization buffer and washed briefly in water and isopropanol, then air-dried.

Labeled cDNA was redissolved in hybridization buffer (3x SSC with 0.1% SDS) containing 0.4 μg/μl each polyA blocker (Sigma) and yeast tRNA (Invitrogen), and 0.8 μg/μl sonicated salmon sperm DNA (Fisher Scientific). This mixture was denatured by heating 2 min at 95°C, then quickly pipetted onto the microarray surface and covered with an appropriate cover slip. Arrays were hybridized (14-18 hrs at 65°C) in sealed hybridization chambers containing a reservoir of 2x SSC. Hybridized arrays were washed 5 min in 2x SSC + 0.1% SDS, 3 min in 0.2x SSC, and 3 min in 0.1x SSC. Slides were briefly rinsed in distilled water and isopropanol, then air-dried and stored in darkness with dessication prior to scanning.

2.3 Results

cDNA probe sets

Embryonic heart cDNA library

At the time this work began, all 5102 clones in the embryonic heart cDNA library had been sequenced. However, due to the relative paucity of other publicly available sequence data, many clones were not assigned gene identities. Thus, each clone was assigned a priority ranking (1-3) based on the redundancy of that sequence within the library and the level of confidence in the identity assigned to that sequence. Aliquots of all priority 1 (i.e. known genes; 189 clones) and most priority 2 (i.e. ambiguous identities, unique sequences; 1155 of 1257 clones) clones were transferred into 96-well plates entitled MAZF 01-14 (MicroArray ZebraFish).

Over the next several months, as the NCBI UniGene and TIGR TC EST clustering databases developed, priority 2 and 3 clones lacking strong homology to known genes

were assigned UniGene and/or TC identifiers. Based on this information, 378 redundant clones were removed from replicates of MAZF 01-14 (new plates called MAZF 31-44). Embryonic heart ESTs representing 759 additional genes or EST clusters were aliquoted into plates MAZF 45-52. This work was primarily the responsibility of Dr. Matthew Grow. Complete clone lists for all MAZF plates can be found in supplemental material (attached disk).

Over time, the integrity of the embryonic heart library declined. Obtaining high quality PCR products became difficult, then impossible. Phage stocks also lost the ability to reinfect bacterial cells. While these issues presumably arose from some error in storage or handling of the original phage stocks, the precise origin of the problem was never determined. Nonetheless, it was necessary to discontinue work with this library.

Adult heart cDNA library

Aliquots of a zebrafish adult heart cDNA library transformed into DH5α *E. coli* cells were obtained from Dr. Ashok Srinivasan. The average size of cloned inserts in this library was ~1kb, and >90% of all clones contained significant inserts [A. Srinivasan, pers. comm.]. To estimate gene representation within the library, a small aliquot was plated onto LB-Agar and 96 colonies were randomly selected for plasmid DNA isolation and sequencing. Approximately 50% of these clones represented known genes, and 35% were redundantly represented [M. Grow, pers. comm.].

Additional aliquots were used to generate a gridded library. Individual colonies (76,800) were robotically picked from LB-Agar plates and inoculated into triplicate liquid cultures. One set of cultures (R2) was held at 4°C for frequent use. Other culture plates contained HMF and were stored at -80°C; one set was designated the 'master' (M) and was retained for archival purposes, while the other (R1) was used to generate replacement frequent-use replicates (R3, R4) on an approximately annual basis. At undefined points during the colony picking process, 7 of 96 pins in the robotic head malfunctioned, resulting in ~7% failure of bacterial growth.

Additional genes of interest

Gene representation of cDNA libraries was supplemented with clones for genes with known toxicological or developmental importance. Fragments of 11 nuclear receptor and 3 cytochrome P450 (CYP) genes were PCR-amplified from whole adult zebrafish cDNA using primers designed against publicly available gene sequences (Table 2.2). These fragments were cloned into the pGEM-T Easy vector (Promega), and the identities of all clones were confirmed by sequencing. Attempts to clone other nuclear receptors, including retinoic acid, retinoid, and peroxisome proliferator activated receptors, were abandoned after two primer pairs per gene failed to amplify the desired product.

We obtained additional clones from a number of other researchers. Members of Dr. John Stegeman's group (Woods Hole Oceanographic Institution) contributed nuclear receptor and CYP gene fragments that had been generated through homologous cloning (Table 2.3). Dr. Frederick Goetz (Marine Biological Laboratory) provided clones for several genes involved in regulation of cell-cycle and apoptosis (Table 2.4). Multiple researchers from the Cardiovascular Research Center (Massachusetts General Hospital) contributed a total of 68 clones for genes with known roles in development and function of the heart, blood, and vasculature (see supplemental material).

Housekeeping genes and controls

Initially, fragments of 12 mitochondrial genes (10 protein-coding genes, 12s ribosomal RNA, and the D loop region) were PCR-amplified from genomic DNA [M. Grow]. However, this process resulted in low-quality PCR products that did not produce significant hybridization signal when included on AH001 microarrays (Figure 2.1). PCR primers designed against transcribed sequences (Table 2.5) successfully amplified 12s and 16s ribosomal RNAs, the regulatory D loop region, and all protein-coding genes except ATP synthase subunit 8 (Figure 2.2). These fragments were not cloned, but rather, amplified from whole embryo cDNA as needed.

Seven housekeeping genes were also targeted for inclusion on microarrays based on previously published use as controls in gene expression studies. In this case, amplification from genomic DNA was more successful. Hybridization signals from

features on AH001 arrays were far from strong, but fluorescence intensities routinely exceeded detection thresholds. PCR primers were redesigned to eliminate primer-dimers and other artifacts that might reduce PCR efficiency (Table 2.6). In 6 of 7 cases, new primers generated strong, unique bands of the expected size (Figure 2.2); no further attempt was made to obtain a phospholipase A2 fragment.

Arabidopsis thaliana chloroplast genes were added to zebrafish cDNA probe sets to provide negative controls (SpotReport® Array Validation System kits, Stratagene). Initially, only three chloroplast genes (*A.th*1-3, Table) were available as components of the SpotReport®-3 system. Later work made use of the expanded SpotReport®-10 kit (Table 2.7).

Proof of concept – MAZF001

A limited number of MAZF001 arrays were produced for purposes of proof-of-concept and protocol optimization. cDNA probes were generated from 1344 priority 1 and 2 embryonic heart ESTs (plates MAZF 01-14). After a single attempt at amplification, 1264 clones (94%) were successfully amplified. Only 32 of 80 initially failed clones could not be amplified in a second round of PCRs. Thus, the final PCR success rate was >97.5% (1312/1344). Of these 1312 clones, 1184 (>90%) yielded unique PCR products. Dr. Matthew Grow provided PCR products for housekeeping genes (amplified from genomic DNA) and a small number of known cardiovascular genes (plates MAZF 16 & 18). All PCR products underwent isopropanol precipitation and were reconstituted in 3x SSC + 0.1% sarkosyl. Purified probes were arrayed in duplicate by the staff at the Harvard Center for Genomics Research using a GMS 417 arrayer.

Figure 2.3 shows a representative example of hybridizations with MAZF001 microarrays. Features were round and regular in shape, and hybridization signal was evenly distributed within features. In addition, duplicate features yielded nearly identical signal intensities and expression ratios (Figure 2.4). These data constituted a successful proof of concept. Encouraged by the high quality of these arrays, we began work to produce a large number of higher-density microarrays for experimental work.

Heart-Tox Chips (MAZF002 & 003)

MAZF002 and MAZF003 were produced using the manufacture protocols validated by MAZF001. The Heart-Tox probe set consisted of an expanded set of 2112 embryonic heart clones (MAZF 31-52), plus mitochondrial and housekeeping genes (amplified from genomic DNA), 14 nuclear receptors and CYPs, and 51 cardiovascular genes. In all, 2243 probes were spotted in duplicate to generate 4486-feature microarrays (see supplemental material for complete print list). An error in the arraying protocol resulted in multiple probes being printed in the same position. Additionally, failure of the arrayer's digital communication port prevented completion of the print run.

The MAZF002 probes were temporarily stored at -20°C and used for print lot MAZF003. In this case, a custom-built split-pin arrayer system (Harvard Center for Genomics Research) was used to print arrays. Buffer autofluorescence on selected slides from this print lot showed a gradual failure of DNA deposition in the second half of the print run, a situation indicative of clogged arrayer pins. As clogged pins carrying spotting buffer and DNA are likely to cause cross-contamination of probes, the Heart-Tox probe set was discarded. Due to the technical difficulties described, MAZF002-003 arrays were used exclusively for protocol optimization.

PCR product purification

In scaling up from MAZF001 to MAZF002, some disadvantages of using isopropanol precipitation for the preparation of cDNA probes became apparent. Precipitation yields were extremely variable, and often very low. Furthermore, on a high-throughput basis, precipitations were both time- and labor-intensive; cold incubation (>30 min) and centrifugation (45 min) time became limiting, particularly with centrifuge capacity limited to 4 plates.

Thus, the performance of isopropanol precipitations and Millipore Multiscreen PCR filter plates were compared using replicate PCRs from 10 randomly selected adult heart cDNA clones (Figure 2.5). Addition of NaCl to PCRs prior to purification significantly increased the efficiency of standard isopropanol precipitations. However, this also added another manual step to the purification protocol. Mean percent yields from four protocols

utilizing Multiscreen PCR 96-well filter plates were similar to those from isopropanol precipitations with added salt. The Biobot vacuum manifold (in-house) and the Millipore vacuum manifold (borrowed from Harvard Center for Genomics Research) performed comparably. Likewise, there was no clearly superior method for resuspending purified DNA; yields from repetitive pipetting and agitation were similar.

Agarose gel electrophoresis showed significant shifts in the observed size of many PCR products after filter purification (Figure 2.5). Addition of 10x PCR buffer (1:10 vol:vol) to purified PCR products resolved such differences (data not shown), suggesting that apparent size shifts were artifacts of partial denaturation of PCR products in low ionic strength (i.e., salt-free) conditions. As such size shifts were never observed after isopropanol precipitations, this data indicated that Multiscreen filter plates provided a more complete removal of PCR reagents.

Target cDNA Preparation

Most early protocols for microarray hybridization called for the use of purified mRNA as template for reverse transcription in the presence of Cy-dye-conjugated nucleotides. The need to isolate mRNA was a significant hindrance, as the small size of zebrafish embryos limits tissue availability and mRNA purification by oligo-dT affinity is rather inefficient. In addition, so-called direct labeling often produced strong systematic Cy3-bias. For example, in a dye-swapping experiment comparing RNA from embryos exposed to two concentrations of TCDD, Cy3 signal overwhelmed Cy5 signal regardless of which dose group was Cy3-labeled (Figure 2.6). While this phenomenon was not observed in all hybridizations, sporadic occurrences rendered entire experiments (and numerous arrays) useless.

Amino-allyl post-labeling of cDNA generated from total RNA consistently provided comparable or stronger hybridization signal than did direct-labeling of mRNA (representative comparison shown in Figure 2.7). Furthermore, homotypic hybridizations using two different print lots and amino-allyl post-labeled cDNA from either whole adult zebrafish or adult heart tissue revealed limited systematic bias or variance (Figure 2.8). The unadjusted slopes of best-fit lines for Cy3 versus Cy5 fluorescence intensities were

1.14 (R^2=0.98, MAZF001) and 1.19 (R^2=0.89, AH001), indicating very slight Cy3 bias. Overall variability in measured Cy3:Cy5 ratios was limited; for both hybridizations, 1.8-fold deviation from the median constituted the 99.7% confidence interval (3 standard deviations). Plotting log-transformed C3:Cy5 fluorescence intensity ratios against total fluorescence revealed that both bias and variance were most pronounced at low fluorescence intensities (Figure 2.9).

AH001 & AH001A

AH001 and AH001A arrays were the results of two replicate print runs separated by approximately 12 weeks. The probe set consisted primarily of 4896 randomly selected, uncharacterized adult heart cDNA library clones (see supplemental material for complete list). PCR products from 12 out of every 96 reactions (one row per reaction plate) were visualized by gel electrophoresis. Based on this sampling, the PCR success rate for adult heart clones was estimated to be >90%. All cytochrome P450 and nuclear receptor clones were successfully amplified. As previously noted, PCR products for mitochondrial, housekeeping, and known cardiovascular genes were later found to be of relatively poor quality (Figure 2.1). All PCR products were filter purified, dried, and reconstituted in SSC spotting buffer. A GMS 417 arrayer was used to produce two print lots of 42 slides each. AH001 and AH001A arrays were used for further methodological development (below), as well as experimental work described in Chapter 3.

Microarray Quality Control

Buffer autofluorescence

Scanning slides for salt autofluorescence immediately following printing allowed general aspects of feature morphology to be assessed. However, salt fluorescence was strongly influenced by humidity and drying time (Figure 2.10). Furthermore, autofluorescence indicated significantly different feature morphology than did PicoGreen staining of another slide from the same print run (Figure 2.10), suggesting that buffer

autofluorescence was not particularly informative about the quantity or distribution of cDNA probes.

PicoGreen Staining

In order to test the ability of PicoGreen staining to quantitate DNA in arrayed features, we created a chip with four sub-arrays containing different amounts of the same 24 clones. The quantity of DNA per feature was varied by increasing the number of pin strikes used to deposit DNA in each sub-array. The entire array was stained under one cover slip, then photographed and analyzed using the same magnification, exposure, brightness, and contrast settings.

This method allowed determination of cDNA localization within features, as well as relative DNA probe quantities (Figure 2.11). Background-subtracted feature fluorescence increased linearly between 1 and 3 pin strikes (Two-factor ANOVA, p-value <0.001), then decreased slightly between 3 and 4 pin strikes.

Syto22 Staining

To assess the performance of Syto22, an AH001A array was incubated with Syto22 solution ~1 hr in darkness, then rinsed and air-dried. Fluorescence scanning of Syto22 staining provided results qualitatively similar to those obtained using PicoGreen staining of another slide from that print lot (data not shown). Syto22 staining was representative of maximum cDNA hybridization signal, and indicated even, round feature morphology on AH001A arrays (Figure 2.12).

Logistical constraints

As PicoGreen and Syto22 provided comparable results, the selection of one protocol for standard quality control purposes was based largely on logistical constraints. At sufficiently high magnification to allow visual detection and high resolution photography of PicoGreen staining, arrays had to be viewed and photographed in multiple portions. On the other hand, Syto22 required an incubation period three times as long as that for PicoGreen, and produced fluorescence that was too weak to be visualized by fluorescence

microscopy; laser-scanning Syto22-stained slides required the use of off-site equipment. Thus, as neither method was intended to be absolutely quantitative, PicoGreen staining was used as the primary quality control checkpoint for future array lots.

Slide processing

Corning provided two alternative protocols for processing arrays printed on GAPS slides. Rehydration of arrays and subsequent succinate blocking were suggested as means to reduce salt fluorescence within features and non-specific background fluorescence elsewhere. However, the process of rehydrating slides prior to DNA immobilization often caused excess DNA to flow beyond the confines of the spotted feature; the direction of DNA flow was dependent on the way that slides were flipped face-up after being held face-down over steam (i.e., flipping end-to-end caused vertical smears). Resulting DNA smears, detected by PicoGreen staining (not shown) or hybridization (Figure 2.1), gave features a tailed morphology and resulted in overlap between features.

Rehydrating slides by placing them face-up in an enclosed steam bath for 5 min eliminated DNA smearing (Figure 2.13). This modified rehydration and succinate blocking protocol produced low background fluorescence when used in combination with either albumin or casein blocking at pre-hybridization and hybridization steps (Figure 2.13). However, comparable results were obtained without succinate blocking (Figure 2.13), which required the use of expensive, highly toxic n-methyl-pyrrilidinone.

Prompted by this result, we tested an alternative processing protocol that included only the DNA immobilization and denaturation steps from the first protocol. UV cross-linking 100 slides took approximately 30 min, and subsequent washing steps required less than 10 min per batch of 50 slides. Thus, one person could process 100 slides in under one hour, less than half the time required to rehydrate and succinate block that number of slides. This protocol did not sacrifice hybridization quality; levels of background fluorescence were comparable to that seen on rehydrated, succinate-blocked arrays (Figure 2.13). This abbreviated protocol was used for all future array processing.

AH002A & B

Dr. Matthew Grow (Indiana University School of Medicine) provided PCR products from 8448 adult heart cDNA clones. Based on visual inspection of 96 out of every 384 reactions, PCR success was estimated to be >90%. PCR products were filter purified, then dried by vacuum centrifugation and shipped to Massachusetts General Hospital on dry ice. Upon arrival, 13 of 22 plates contained liquid of an unknown origin. Re-drying these samples left an unidentified residue that altered the consistency of re-dissolved PCR products. These probes were discarded; the final AH002 probe set consisted of the remaining 3456 adult heart cDNAs, plus all genes of interest and housekeeping genes (see supplemental material, attached disk). These PCR products were dissolved in 50% DMSO, emperically determined to be the optimal printing buffer for use with the GeneMachines OmniGrid arrayer (data not shown). AH002A and AH002B print runs were separated by 4 weeks and generated 100 slides each. PicoGreen staining of representative arrays indicated even distribution of DNA within regular, round features (data not shown).

Data analysis tools

Two Perl scripts were written to facilitate basic microarray data manipulation; executable scripts are included in supplemental material (attached disk). The first script, ArrayListModifier.pl, converted print lists output by the GMS 417 arrayer into a format readable by Axon GenePix 3.0 software used for image analysis. GenePix allows the user to manually flag features as *good*, *bad*, or *not found*; Flagger.pl was designed to automate this process and eliminate subjective judgements on hybridization image data. Taking an un-flagged GenePix results file in tab-delimited text format as input, Flagger.pl excluded from further analysis any feature with a low signal:noise ratio, significant signal saturation, or excessive spatial variation in ratio measurement (Figure 2.14). Expression ratios for features meeting all criteria were then normalized to the median ratio of all such features on that slide. Flagged, normalized data were output in tab-delimited text format appropriate for use by text editors or Microsoft Excel.

The functions performed by Flagger.pl were later automated and made more user-friendly by conversion to Visual Basic (VBA) macros with a graphical user interface in Microsoft Excel (Figure 2.15, and supplemental material). This package enabled application of user-determined threshold values for data flagging, as well as grouping and statistical analysis of data from multiple hybridizations (batch processing and grouping functions). Output files could be formatted for text editors, Microsoft Excel, or common microarray analysis software, such as Cluster (http://rana.lbl.gov/EisenSoftware.htm).

2.4 Discussion

Microarray production

In all, we have generated seven full-scale cDNA microarray print lots, as well as several small test batches. Initial MAZF001 arrays were of superb quality; feature morphology was excellent, hybridization signal was generally strong, and the observed reproducibility of results from replicate features is rare among spotted arrays [140]. However, protocols used for production of MAZF001 arrays did not translate well into a high-throughput work-flow.

Efficiency and consistency are of the utmost importance when designing protocols or instrumentation for use in high-throughput work of any kind. The need to balance performance and efficiency were constant considerations in interpreting results of methodological comparisons and selecting protocols for array synthesis. For example, Multiscreen PCR-96 plates (Millipore) were favored over isopropanol precipitations based largely on their adaptability to high-throughput application. Multiscreen plates are affordable, utilize a simple, rapid protocol that is easily automated, and provided consistently high quality results. In contrast, the QiaQuick PCR Purification system (Qiagen), commonly used for cDNA purification, was eliminated from consideration because of both protocol complexity and expense.

Dramatic differences in arrayer efficiency also influenced protocol selections. Array quality was similar when either the GMS 417 or the OmniGrid was used. However, the OmniGrid is capable of printing a given number of features onto 100 slides in a quarter of

the time that the GMS 417 could print that same number of features onto only 42 slides. As we had firsthand experience with the ability of high-salt buffers to severely clog quill-tip, or split, arrayer pins, the decision to switch to using the OmniGrid, in turn, prompted a re-evaluation of printing buffers.

Methodological comparisons utilizing MAZF001 arrays, as well as technical difficulties encountered in production of MAZF002 and MAZF003 print lots, resulted in the development of a robust work-flow for high through-put synthesis, quality assessment, and use of cDNA microarrays (Figure 2.16). These protocols, with some deviations, were used to produce nearly 300 zebrafish adult heart cDNA microarrays for use in experimental work.

Adult heart arrays

The use of an uncharacterized, redundant clone set (i.e., the adult heart cDNA library) for microarray construction is a unique aspect of this work. The rationale behind this strategy was two-fold. Firstly, avoiding the expense and time required to sequence a significant portion of the cDNA library significantly accelerated the completion of cDNA microarrays. Omitting library subtraction or normalization obviously saved additional time and labor. More importantly, it avoided a common pitfall of these methods, namely elimination of rare transcripts.

Of course, redundancy in the library restricted microarray gene representation. Taking into account estimates of PCR failure rates (≤10%) and redundancy within the adult heart cDNA library (~35%), the 4896 adult heart clones on AH001 arrays are likely to represent approximately 2800 unique cardiovascular genes. Likewise, the 3456 clones arrayed on AH002A/B probably correspond to approximately 2000 unique genes. To date, the Cardiac Gene Expression Knowledgebase has documented expression in human heart tissue of transcripts mapping to 7056 unique loci in the human genome [141]. Assuming similar gene complement and transcriptional regulation in zebrafish, clones found on adult heart microarrays may encompass 30-40% of cardiac transcripts. At this level of coverage, one would expect all major pathways to be represented by at least one arrayed clone. Thus, while these arrays will not provide *complete* transcriptional profiles

(few arrays do), they should enable meaningful functional profiling of varied conditions. Furthermore, the large proportion of randomly sequenced clones with no correlate in public sequence databases suggests significant opportunities for gene discovery.

Microarray quality control

The ability to evaluate array quality prior to performing expensive and labor-intensive hybridizations is crucial. This point was made abundantly clear by technical problems with MAZF002 and MAZF003 print lots. We have assessed the performance of two double-stranded DNA-binding dyes, PicoGreen and Syto22 (Molecular Probes), and developed a method using PicoGreen staining to rapidly evaluate newly printed slides. This method provides relative quantitation of ≤4-fold differences in the amount of printed cDNA; the detection range for this method may be greater than measured here, as repetitive deposition likely resulted in excessive build-up of salts that would inhibit PicoGreen fluorescence [142]. It should be noted that the selection of PicoGreen was based largely on logistical factors, namely the availability of a microarray scanner. Were an appropriate laser scanner readily available, Syto22 would have been the method of choice, as images of Syto22-stained slides could be captured and analyzed utilizing the workflow applied to hybridizations.

However, both PicoGreen and Syto22 suffer from two major limitations – (1) these dyes detect only double-stranded DNA, and (2) stained slides cannot be used for hybridizations. dsDNA specificity was not problematic when SSC spotting buffers were used, but limited the usefulness of these dyes for assessing probes printed in 50% DMSO. This problem is easily remedied by using similar dyes with affinity for ssDNA, such as SYBR Green II [143]. Thus, the more pressing issue is the fact that only a small number of slides from each print lot can be examined. As slide-to-slide variability can be a confounding factor in microarray experiments, the ideal quality control protocol would allow assessment of each individual array prior to use. Development of such a method, involving probe synthesis using PCR primers conjugated to a fluorophore with absorption/emission spectra distinct from Cy dyes, has only recently begun [144, 145].

Dye bias and labeling protocols

Cy3 dye bias observed in direct incorporation experiments is unsurprising given the nature of CyDye-conjugated nucleotides. CyDye-conjugation interferes with base pairing of modified nucleotides, and the larger size of Cy5 often results in ≤10-fold less efficient incorporation of Cy5-conjugated dCTP [146]. The degree of bias may vary sequence specifically [147, 148].

A common solution to the dye bias problem is the use of dye-swapping – each pair of samples is used for two hybridizations with reversed dye labeling [147]. Under this scheme, genes that are subject to dye bias will show the same expression ratio regardless of labeling direction; such genes are eliminated from further analyses. This solution has two drawbacks. Firstly, dye swapping requires two hybridizations for every comparison. Secondly, while eliminating biased data from analyses improves the quality of the dataset, it also restricts the gene representation.

Amino-allyl post-labeling provides a better solution. In amino-allyl-modified nucleotides, the amino-allyl moiety is conjugated to ribose and, thus, does not interfere with base pairing. Amine-reactive Cy3 and Cy5 are added after cDNA synthesis is complete (thus, the term post-labeling). The two dyes do not differ in their affinity for amino-allyl, and steric hindrance is extremely limited when conjugating to the outside of the DNA backbone. Thus, this step is not subject to significant dye bias. Additionally, amino-allyl is smaller than cyanine dyes, resulting in overall greater incorporation efficiency. Accordingly, we observed equal or stronger signal from amino-allyl labeled cDNA probes. The slight Cy3 bias that was observed is probably due to the fact that Cy5 is inherently a slightly weaker, more photo-labile fluor than Cy3.

Experimental design

A common approach in microarray work has been to print replicate features for each gene, then average all data from replicate features on pairs of dye-swapped hybridizations into a single gene expression ratio. In the current case, nearly identical data from replicate features provided no extra information. Similarly, based on observations of extremely limited dye bias and systematic variance when using amino-allyl post-labeling,

dye-swapping seemed superfluous. Thus, we adopted an experimental design strategy to maximize gene representation and biological replication, rather than technical pseudo-replication. Adult heart arrays contained only one feature per clone to allow more clones to be included on each array, and more arrays to be produced from a single probe set. Experiments were designed to include three or more independent biological replicates of sufficient size to provide for at least one hybridization (see Chapter 3). In accordance with this approach, data analysis tools were designed for high-throughput filtering and normalization of data, as well as basic statistical analysis of replicate hybridizations.

Conclusions

Ultimately, there is no single optimal protocol for the synthesis and use of cDNA microarrays. However, certain general guidelines can be drawn from the data at hand. While isopropanol precipitation is inexpensive and effective on small scales, filter purification of PCR products is likely to be more efficient on a high- throughput basis. Selection of a printing buffer should take into account arraying technology, as well as chemical properties of the buffer; high-salt buffers are best suited to GMS (now Affymetrix) ring-and-pin set-ups, while DMSO mixtures work well with split pins. Most importantly, routine quality control is absolutely necessary to detect technical problems early, avoid wasteful use of resources, and ensure high quality gene expression data.

We have produced seven full-scale print lots of zebrafish cardiovascular-specific microarrays. MAZF001 was completely dedicated to methodological development. MAZF002 and 003 print runs were fraught with technical difficulties; although frustrating, each difficulty highlighted important technical weaknesses in our protocols and emphasized the need for vigilant quality control. Adult heart arrays, AH001/A and AH002A/B, were constructed with the benefit of lessons learned from three previous print runs. These high-quality arrays have been a significant resource to the zebrafish toxicology community, enabling the gene expression profiling work presented in this thesis (Chapter 3), as well as several on-going collaborative projects (see Chapter 5).

Table 2.1 Names, sequences, and melting temperatures of primers used for PCR amplification of clones from embryonic and adult heart cDNA libraries. ZapExPCR primers bind regions flanking the insertion site of the λ-ZAP bacteriophage vector. SP6, T3 and T7 primer sequences are derived from corresponding promoter elements found on the majority of plasmid vectors.

Primer Name	Primer Sequence	Length (nt)	T_m (°C)
ZapExPCR-f	GCCAAGCTCGAAATTAACCCTCACTAAAGGG	31	68.2
ZapExPCR-r	CCAGTGAATTGTAATACGACTCACTATAGGGCG	33	68.2
SP6	ATTTAGGTGACACTATAG	18	46.9
T3	ATTAACCCTCACTAAAGGGA	20	53.2
T7	TAATACGACTCACTATAGGG	20	53.2

Table 2.2 PCR primer sequences and fragment sizes for nuclear receptors and cytochrome P450s included on cDNA microarrays. Primers were designed using publicly available sequence data. All PCR products were cloned into the pGEM-T Easy vector (Promega).

Gene name	GenBank accession #	Product size (bp)	PCR primer sequences
aryl hydrocarbon receptor 1 (AHR1)	NM_131028	497	ATGTCATTCATCAGAGTGTG ACCACTATTACAGAGCTCTGC
aryl hydrocarbon receptor 2 (AHR2)	NM_131264	1051	CCTCAGGGAGTCCCCACATC GCTTTCCTCAGAGTTGCCAC
AHR nuclear translocator (ARNT)	Y08434	700	AGGCGGCGGATGGTTTCTTG TCGGGATGGCAGAACTCCAG
estrogen receptor α (ER-α)	AF349412	763	GTAAAGATCGCGGAGGGCGTTC AGCAGGAGCTGGGCCTGGCG
estrogen receptor β (ER-β)	NM_174862	752	GGAGCGCTGCAGTTATCGAG GGATGGACTGTTGTTGTGAG
thyroid hormone receptor A (thra)	NM_131396	1108	GTGTCAGAGTGGGAACTCATTCG GTCTGCAGTGCTGGTGGGTTG
thyroid hormone receptor B (thrb)	NM_131340	1084	GTGGACATTGAAGCCTTCAGTC TCGGTCTAGGTACTGTAAGTGC
peroxisome proliferator activated receptor α (PPAR-a)	U93473	95	CTTCAGGCGGACGATTCGGCTC CGACAGTATTGGCACTTGTTTCG
retinoid receptor α (rxra)	NM_131217	925	GAAAGACCTGACGTACACTTG CGCTGGGGTTTATTTACATGC
retinoid receptor δ (rxrd)	NM_131238	1045	TCTTCGGGGAAGCATTATGGC TGCAGTCACAGTTATCTCCAG
retinoid receptor ε (rxre)	NM_131275	1015	CTGTGAGGAAGGACCTTAGCTAC CTGCGATACCCTGGTGCAAGC
cytochrome P450 1A (CYP1A)	AB078927, AF210727	598	TTGACACTATCAGTACGGCTC TTCTGGATCTAGAACACAGGC
ovarian aromatase (CYP19a)	NM_131154	1072	GCAGTGCATCGGGATGCATGAGC GCTGCGACAGGTTGTTGGTTTGC
brain aromatase (CYP19b)	NM_131642	~800	ATGATGGAAGCCTGAGGACGGC GTCTGTTGAGACGTCAACCACG

Table 2.3 Clone information for fragments of cytochromes P450, nuclear receptors, and related genes contributed by various researchers. At this time, all sequences except a 330bp fragment of CYP1B1 are unpublished.

Gene Name	GenBank accession #	Fragment size (bp)	Cloning Vector	Contributed by
AHR repressor (AHRR)	none		pGEM-T Easy	B. Evans
pregnane X receptor (PXR)	none	1132	pGEM-T Easy	A. Bainy
cytochrome P450 1B (CYP1B)	none	~500	pGEM-T Easy	B. Woodin
cytochrome P450 1B1 (CYP1B1)	AF235139	330	pGEM-T Easy	C. Godard
cytochrome P450 2AA2 (CYP2AA2)	none	1497	pGEM-T Easy	A. Bainy
cytochrome P450 2AA1 (CYP2AA1)	none	~1500	PCR product	A. Bainy
cytochrome P450 51 (CYP51)	none	1250	pGEM-T Easy	A.M. Morrison

Table 2.4 Clones for genes involved in cell-cycle regulation and apoptosis, provided by Dr. Frederick Goetz (Marine Biological Laboratory).

Gene Name	GenBank accession #	Fragment size (bp)	Cloning Vector	Contributed by
cyclin A1 (cycA1)	AF268045	1639	pBK-CMV	F. Goetz
cyclin B1 (cycB1)	AF268043	1534	pBK-CMV	F. Goetz
cyclin D1 (cycD1)	AF365874	1999	pBK-CMV	F. Goetz
cyclin-dependent kinase 9 (cdk9)	AF268046	1768	pBK-CMV	F. Goetz
cell division control protein 2 (cdc2)	AF268044	1236	pBK-CMV	F. Goetz
tumor suppressor p53	AF365873	2199	pBK-CMV	F. Goetz
steroidogenic acute regulatory protein	NM_131663	1291	pBK-CMV	F. Goetz

Table 2.5 PCR primer sequences and product lengths for zebrafish mitochondrial genes included on AH002 cDNA microarrays.

Gene name	GenBank accession	PCR primer sequences	Fragment size (bp)
NADH dehydrogenase subunit 1	NC_002333	TCAACGCTGGCAGAAACAAAC TGGTCGTATCGGAATCGTGG	290
NADH dehydrogenase subunit 2	NC_002333	TAGCACAACAACACCACCCACG GCTGTGGCTGGTAGGTCTTGTTTC	706
NADH dehydrogenase subunit 3	NC_002333	CGACCTTATCATTGGTCTTAGC GGCTCATTCGTAGGCTAGTC	290
NADH dehydrogenase subunit 4	NC_002333	ACCCGATGAGGTAATCAAGC TCAAGTTTGGTAGAGGTGGAAG	727
NADH dehydrogenase subunit 4L	NC_002333	CGCACTTTAGTCTTAACGCAGC TATGTGGTCAGATCCGTGGG	260
NADH dehydrogenase subunit 5	NC_002333	TTGGCTGATGATTTGGGCGGAC TGTGTCGGGGGCTTCCTAAACAG	852
NADH dehydrogenase subunit 6	NC_002333	AGCCGAGCCTTTTCCTGAAG GCACGAAGCACACCATAACTAAGAC	291
cytochrome c oxidase 1	NC_002333	CCAGGATTCGGCATTATCTCCC CTTCTCGTTTGGCGGTAAAGG	700
cytochrome c oxidase 2	NC_002333	AGGATTCCAAGACGCAGCATC TTAGCCCCGCAGATTTCAGAG	588
cytochrome c oxidase 3	NC_002333	CCAAGCCCATGACCACTAACTG CGACGAAGTGTCAATATCAAGCG	700
D Loop	NC_002333	CCTGGTATCTGGTTCAAATCTCACG TATTGGCTGTACGTTCTCGGGC	441
12s ribosomal RNA	NC_002333	AAACTCGTGCCAGCAACC ACTTTTCCCCCCTTGTCTG	676
16s ribosomal RNA	NC_002333	GCACAAGTGTAAGCCAAGTTG TTTCGGGAAGAGGTTTTAGC	937
cytochrome b	NC_002333	CATCTGTTGTGCATATTTGCCG AGCATGTCTGCTACCAGTGTTCAG	808
ATP synthase subunit 8	NC_002333	TGCCTCAGCTTAATCCAAAAC TGTGCTCTTTAGCATCAACTTG	135
ATP synthase subunit 6	NC_002333	ACCAACTTATGACCCCACTAAAC AAGAAAAGGACGGAGGCAG	430

Table 2.6 Highly expressed housekeeping genes included on microarrays as biological negative controls. Primers for G3PDH, ubiquitin, and CAB45 were based on TIGR TC cluster consensus sequences with strong homology to the gene of interest.

Gene name	GenBank or TIGR accession #	PCR primer sequences	Fragment size (bp)
β-actin	NM_131031	TGAGCACGGTATTGTGACCAACTG GCAAGAGAGGTGATTTCCTTCTGC	750
elongation factor 1-α (EF1-α)	L23807	TCTACAAATGCGGTGGAATCG CAACCATACCAGGCTTGAGGAC	750
glyceraldehyde 3-phosphate dehydrogenase (G3PDH)	TC84783	CGAACAGAGGCTTCTCACAAACG CAGCGTCAAAGATGGATGAACG	932
ubiquitin	TC94110	CATCTAAGAGCTGGTGGTGGATTG AGCACAGACAGCCTCATGTGTGAC	555
$Ca2^{+}$-binding protein 45 (CAB45)	TC96362	GATTCTTGCGGTTATCGGTCTG AAACTTCACACGGTATTCGTCCC	756
ornithine decarboxylase	NM_131801	CTGAGTGTGAAGTTTGGAGCGAC CATCGGGCTTGGGTTTCTTG	550
phospholipase A2	NM_131295	TTTGGGTGTGAAGGAGACGACC ACTGAGCGAAAGGGAAACCG	970

Table 2.7 *Arabidopsis thaliana* genes used as negative controls on zebrafish microarrays. All PCR products were purchased from Stratagene as components of the SpotReport®-3 (*A.th*1-3) and SpotReport®-10 (all genes) Array Validation System kits. Information shown here was taken from the SpotReport®-10 Array Validation System instruction manual (Stratagene catalog #252010).

Clone ID	Gene name	Stratagene product #	GenBank accession #	Fragment size (bp)
A.th 1	photosystem I chlorophyll a/b-binding protein (Cab)	252101	X56062	500
A.th 2	RUBISCO activase (RCA)	252102	X14212	513
A.th 3	ribulose-1,5-bisphosphate carboxylase/ oxygenase, large subunit (rbcL)	252103	U91966	521
A.th 4	lipid transfer protein 4 (LTP4)	252104	AF159801	527
A.th 5	lipid transfer protein 6 (LTP6)	252105	AF159803	477
A.th 6	papain-type cysteine endopeptidase (XCP2)	252106	AF191028	507
A.th 7	root cap 1 (RCP1)	252107	AF168390	533
A.th 8	NAC1	252108	AF198054	457
A.th 9	triosphosphate isomerase (TIM)	252109	AF247559	498
A.th 10	ribulose-5-phosphate kinase (PRKase)	252110	X58149	497

Figure 2.1 Scans of two AH001 microarrays hybridized with whole embryo cDNAs showing general lack of hybridization signal from regions containing genomic PCR products for mitochondrial, housekeeping, and many known cardiovascular clones (white rectangles). Smearing of cDNA probes due to movement of slides prior to UV cross-linking is also apparent, to various degrees, on both slides.

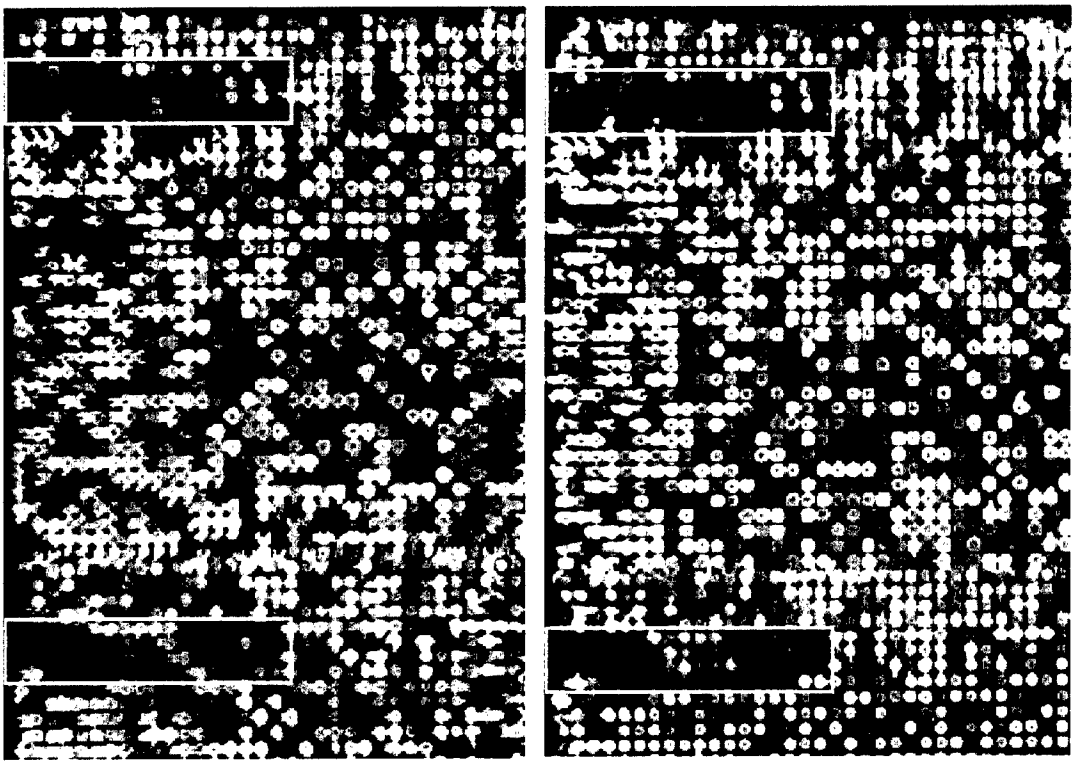
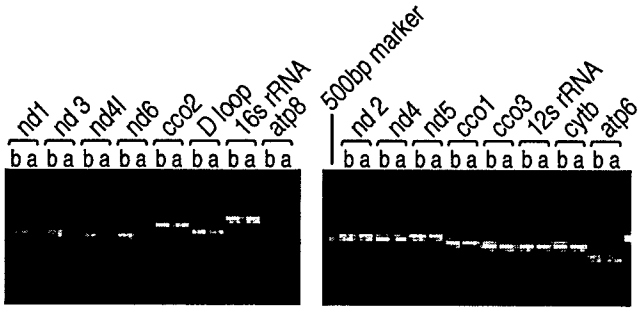


Figure 2.2 PCR products for mitochondrial (a) and housekeeping (b) genes, amplified using primers described in Tables 2.5 and 2.6. Each PCR product is shown before and after purification using Millipore Multiscreen PCR-96 (labeled b and a above the appropriate well). Several PCR products were not visible by gel electrophoresis after purification; the presence of these products was confirmed spectrophotometrically.

(a)


nd1
nd 3
nd4l
nd6
cco2
D loop
16s rRNA
atp8
500bp marker
nd 2
nd4
nd5
cco1
cco3
12s rRNA
cytb
atp6
b a b a b a b a b a b a b a b a
b a b a b a b a b a b a b a b a b a

(b)

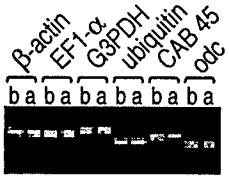
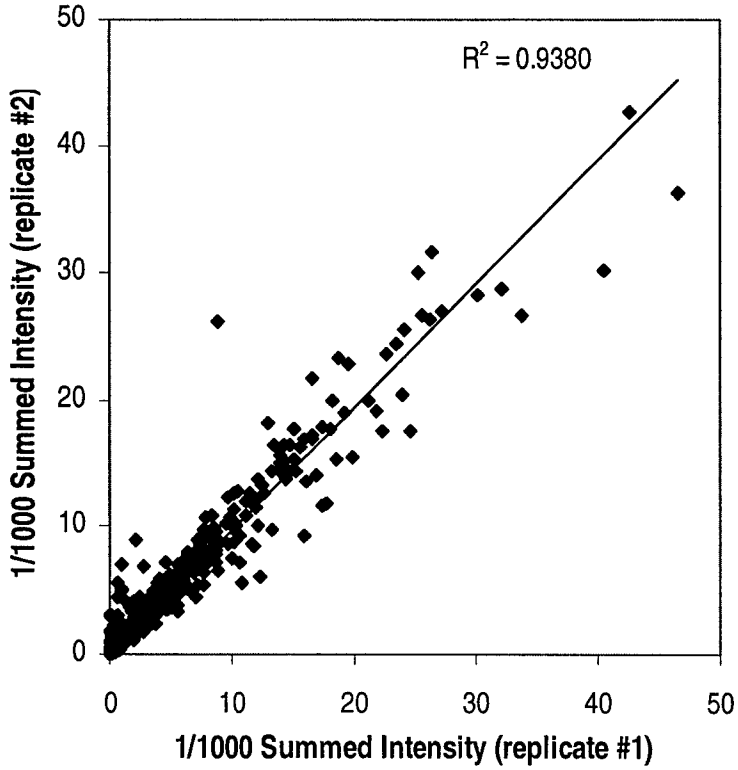

β-actin
EF1-α
G3PDH
ubiquitin
CAB 45
odc
b a b a b a b a b a b a

Figure 2.3 MAZF001 microarray hybridized with 1 μg direct-labeled cDNA from zebrafish adult heart tissue (Cy5, red) and from whole adult zebrafish minus heart tissue (Cy3, green). All probes were spotted in duplicate; panels 2 and 4 are exact replicates of panels 1 and 3.

1
2
3
4

Figure 2.4 Comparison of results from duplicate features on the MAZF001 hybridization shown in Figure 2.3. Replicate measurements of both total fluorescence intensities (top) and relative expression ratios (bottom) were tightly correlated.


$R^2 = 0.9380$
1/1000 Summed Intensity (replicate #2)
1/1000 Summed Intensity (replicate #1)
0
10
20
30
40
50

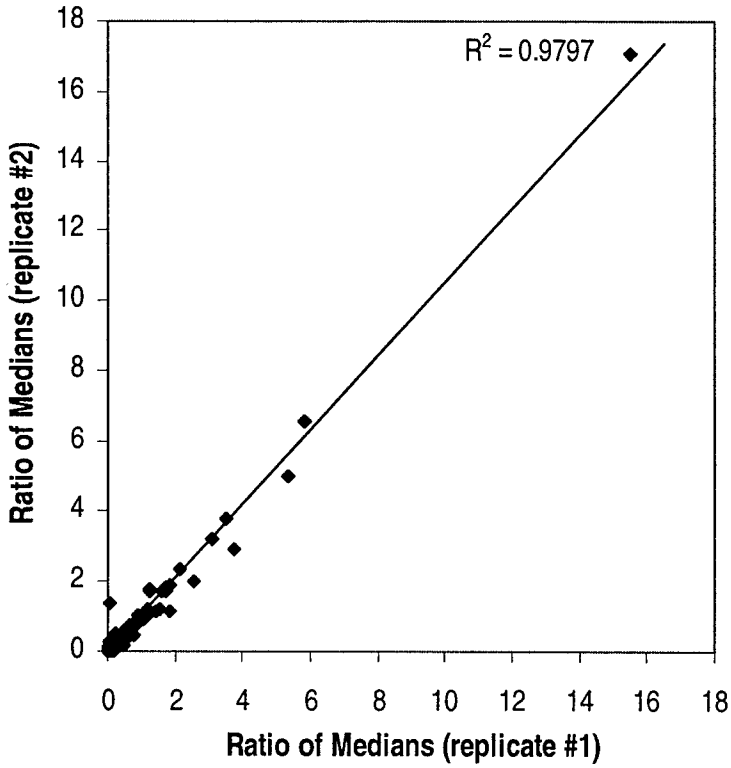

$R^2 = 0.9797$
Ratio of Medians (replicate #2)
Ratio of Medians (replicate #1)
0
2
4
6
8
10
12
14
16
18

Figure 2.5 Comparison of yields from isopropanol precipitations and filter purifications of ten PCR products. Two isopropanol precipitation protocols (black bars), with and without addition of 200nM NaCl, were compared to filter purification using Millipore Multiscreen PCR-96 plates (striped bars). Two vacuum manifolds (BioBot and Millipore) and two DNA resuspension methods (repetitive pipetting or gentle agitation) were compared. Gel electrophoresis of PCR products before (b) and after (a) purification is shown at left. Spot densitometry measurements were used to determine percent yields for each PCR product; mean yields for each method are shown at right.

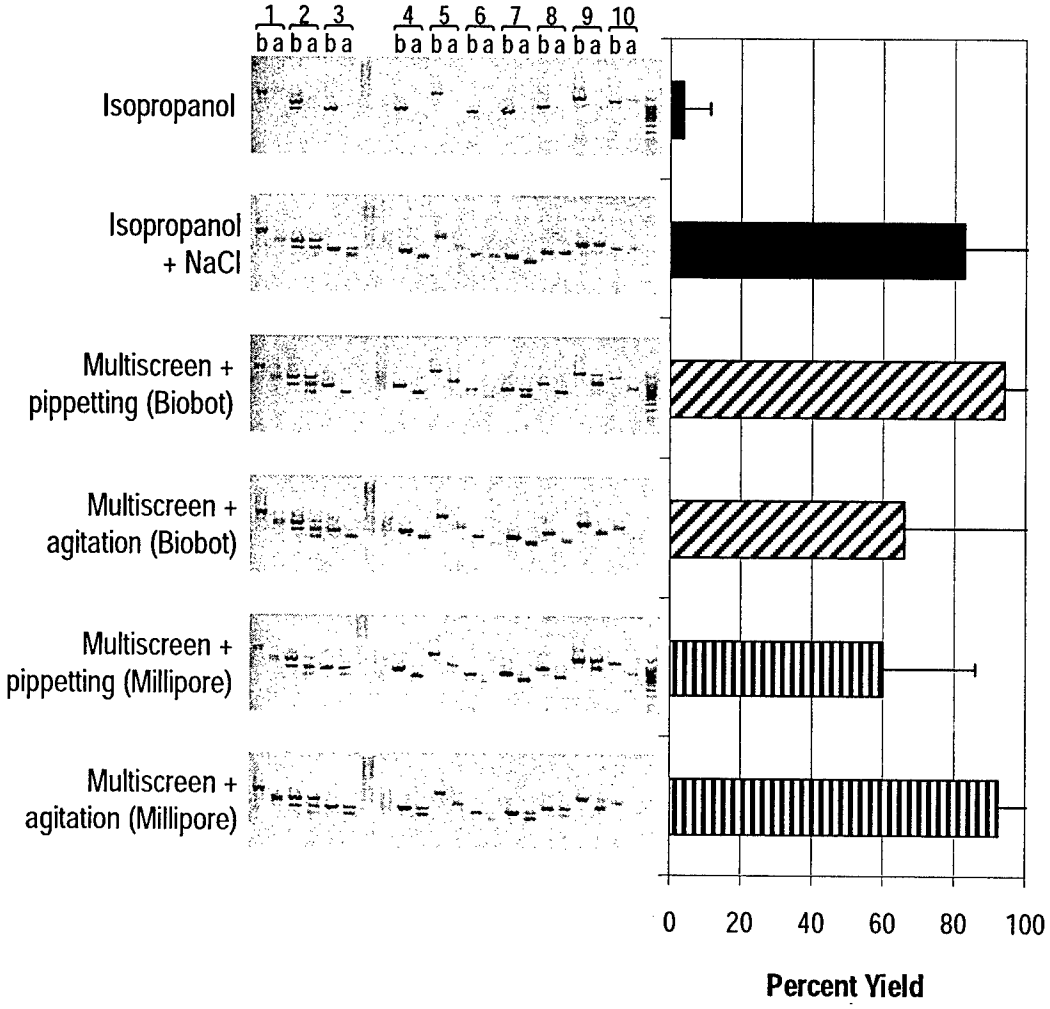
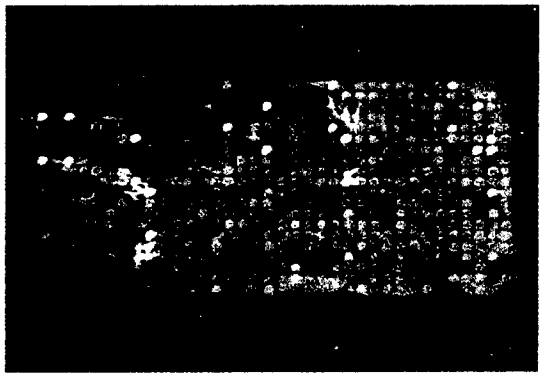

1 2 3 4 5 6 7 8 9 10
b a b a b a b a b a b a b a b a b a b a
Isopropanol
Isopropanol + NaCl
Multiscreen + pippetting (Biobot)
Multiscreen + agitation (Biobot)
Multiscreen + pippetting (Millipore)
Multiscreen + agitation (Millipore)
0
20
40
60
80
100
Percent Yield

Figure 2.6 Dye-swapped pair of AH001 hybridizations comparing cDNA from 72 hpf larval zebrafish exposed to either 1.7 ng/g or 1.1 ng/g TCDD. Cy3 signal overwhelmed Cy5 signal regardless of which cDNA sample was Cy3-labeled.

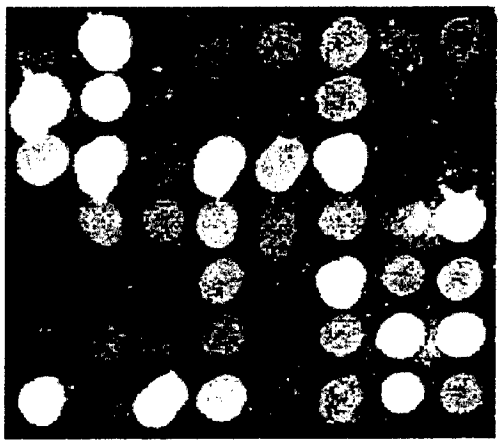
(a)



(b)

Figure 2.7 Comparison of fluorescent signal generated by homotypic hybridizations with (a) 500 ng amino-allyl post-labeled cDNA from adult heart total RNA or (b) 1 μg direct-labeled cDNA from adult heart mRNA. Hybridizations were performed on slides 21 and 22 from print lot MAZF001.

(a) 500 ng amino-allyl post-labeled cDNA



(b) 1 μg direct-labeled cDNA

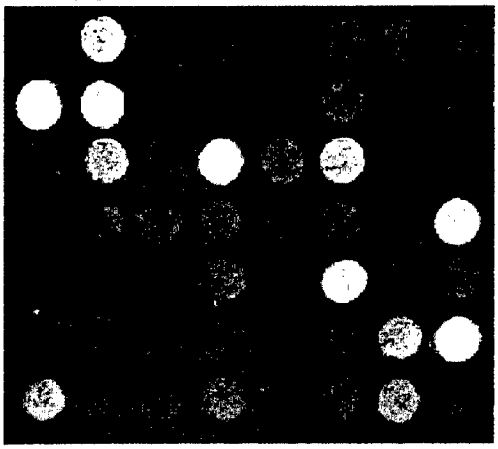
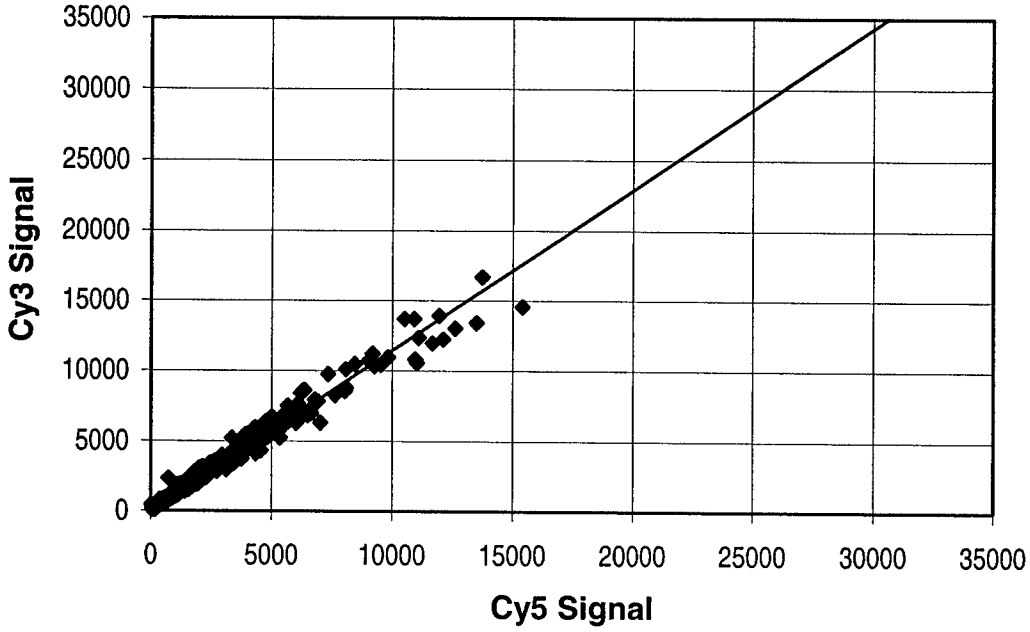


Figure 2.8 Fluorescence intensity scatter plots for homotypic hybridizations using amino-allyl post-labeling of 1 μg cDNA from whole adult zebrafish (a, MAZF001) or 365 ng cDNA from adult heart tissue (b, AH001). Background-subtracted median fluorescence intensities from Cy5 (635nm) and Cy3 (532nm) channels were compared. Linear regression have been fitted to each data set.

(a) MAZF001, 500 ng cDNA from whole adult zebrafish



(b) AH001, 365 ng cDNA from adult heart tissue

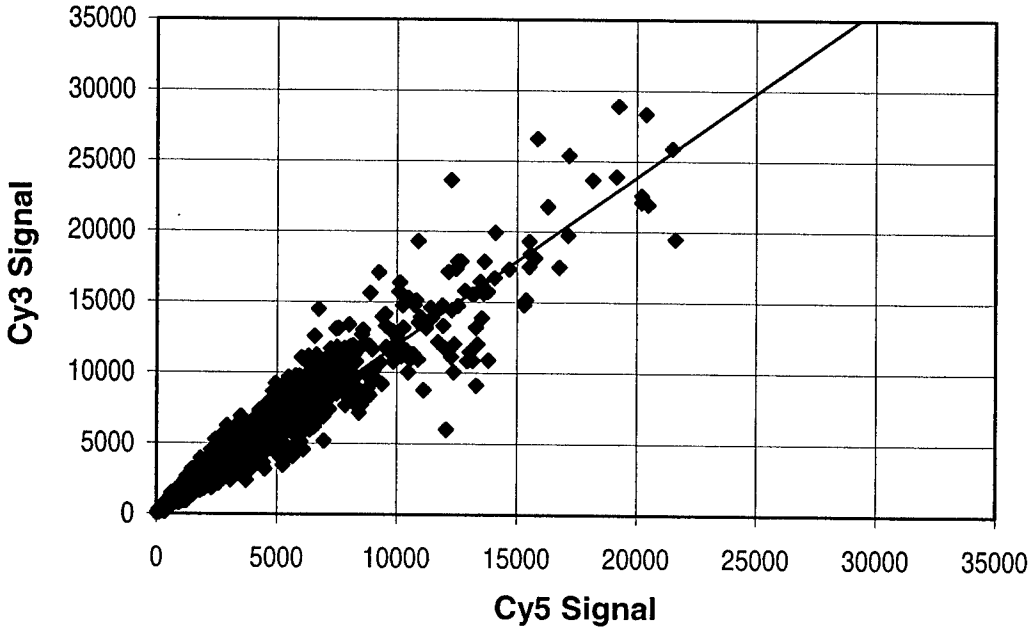
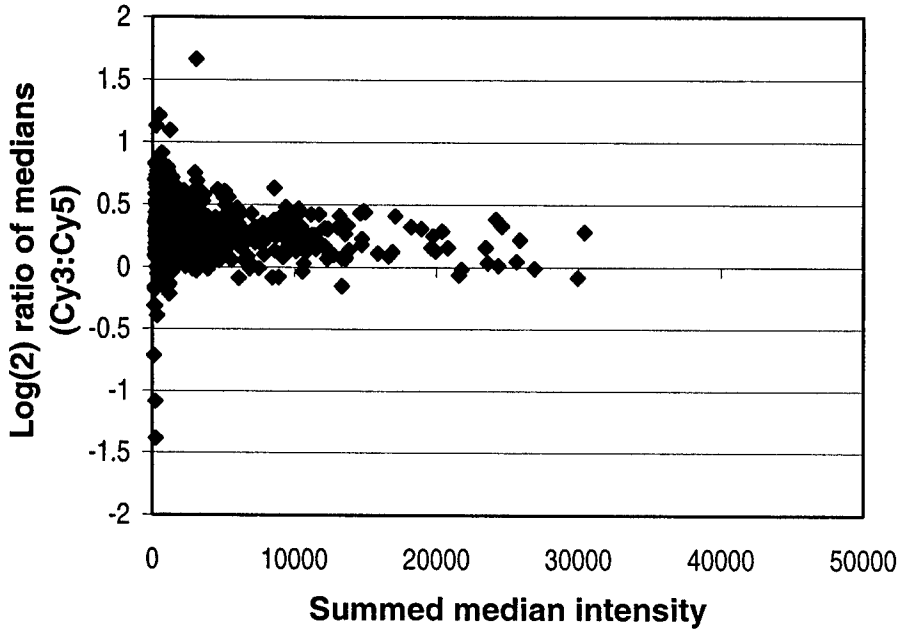


Figure 2.9 Ratio-intensity plots for homotypic hybridizations. One outlier is not shown in graph a, two are omitted from graph b.

(a) MAZF001, 500 ng cDNA from whole adult zebrafish



(b) AH001, 365 ng cDNA from adult heart tissue

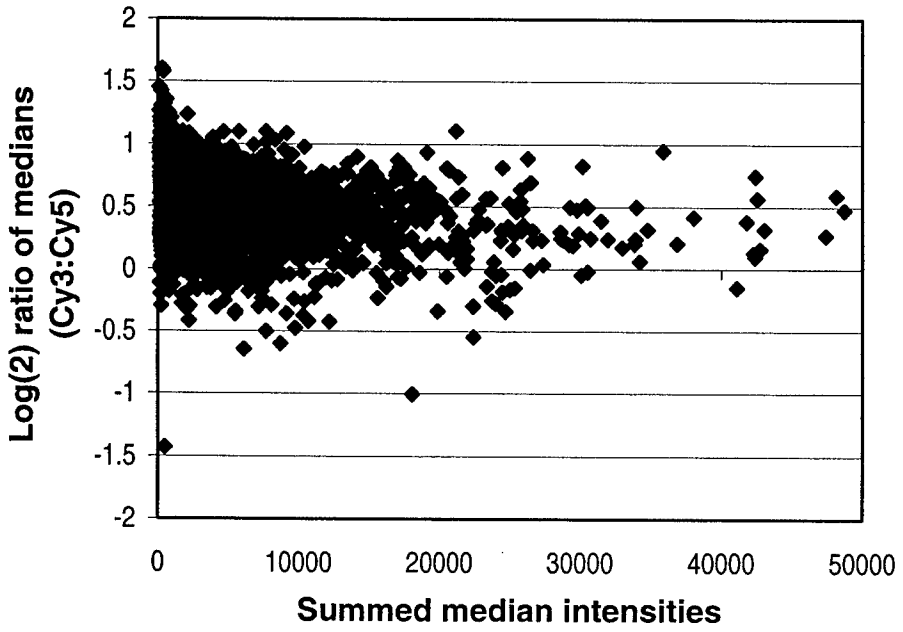


Figure 2.10 Effect of relative humidity during arraying on feature morphology, as indicated by salt autofluorescence (grey bars) or PicoGreen staining (black bars). During the course of the AH001 print run, relative humidity was recorded each time a new set of three probe plates was placed into the arrayer, approximately every 1.5 hrs. At the end of the print run, slide #42 was scanned for salt autofluorescence. Slide #7 was stained with PicoGreen. Slides were divided into regions corresponding to sets of three probe plates, and feature morphology within each region was ranked on a scale of 1-5 on the bases of fluorescence intensity, size, and regularity of shape.

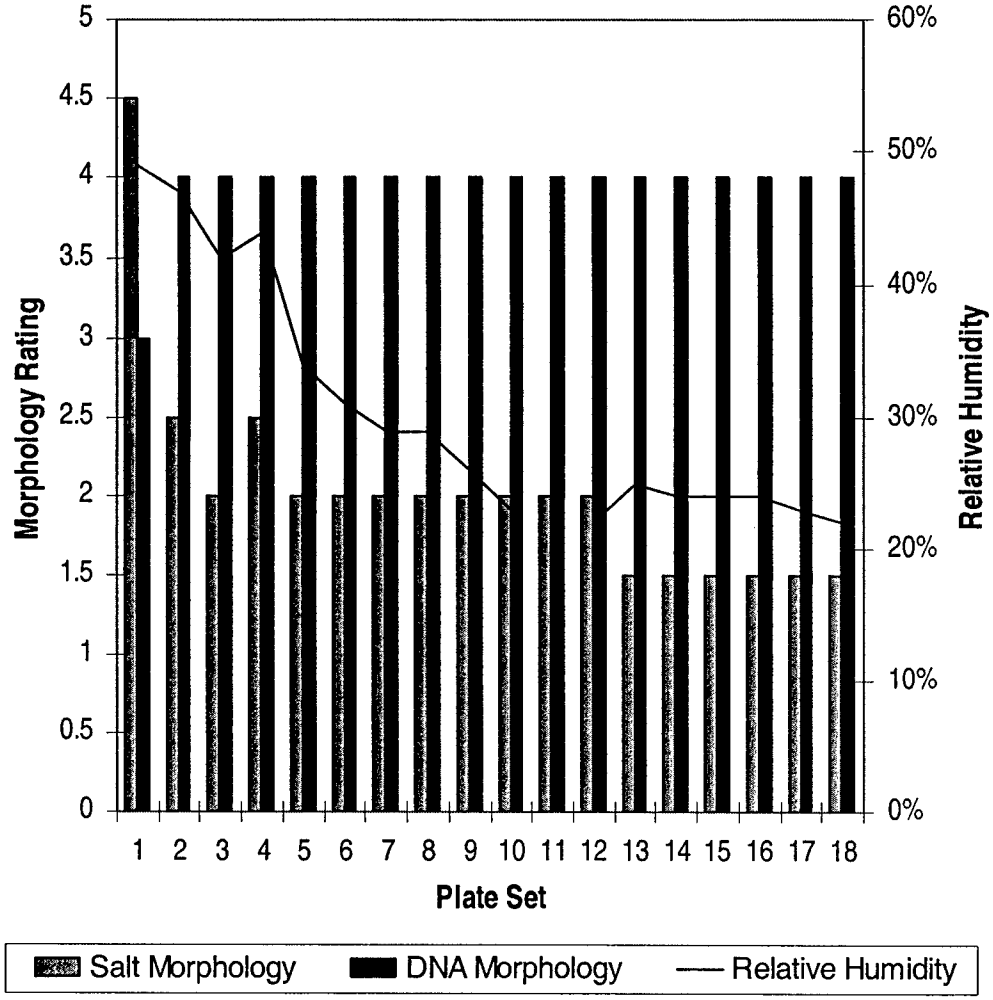

5
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
Morphology Rating
60%
50%
40%
30%
20%
10%
0%
Relative Humidity
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18
Plate Set
Salt Morphology
DNA Morphology
Relative Humidity

Figure 2.11 Quantitation of arrayed DNA by PicoGreen staining. The quantity of DNA in each feature was varied by increasing the number of pin strikes used to deposit probes. Fluorescence from all features generated with a given number of pin strikes was summed.

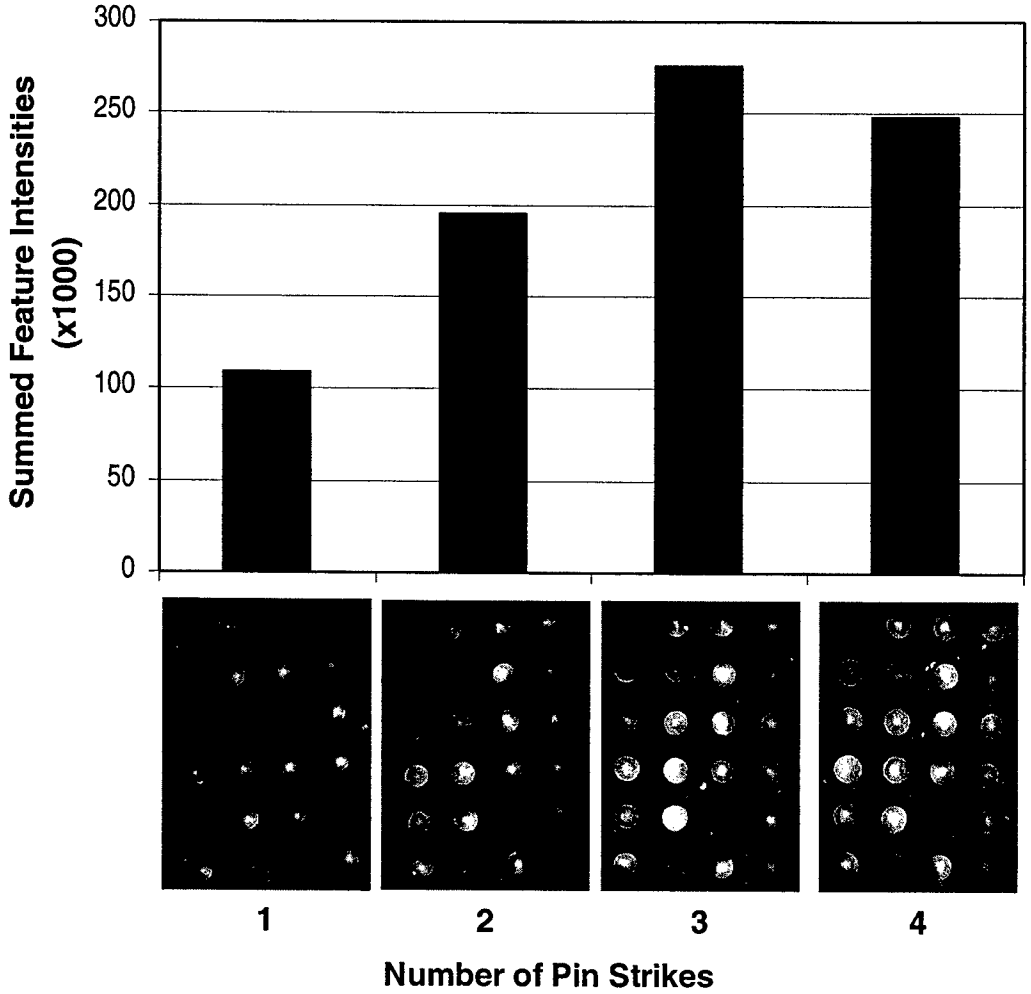

Summed Feature Intensities
(x1000)
300
250
200
150
100
50
0
1
2
3
4
Number of Pin Strikes

Figure 2.12 Laser-excited fluorescence scan of a Syto22-stained AH001A microarray (top), and comparison of feature morphology observed by Syto22-staining to that observed on two randomly selected AH001A hybridizations (bottom).

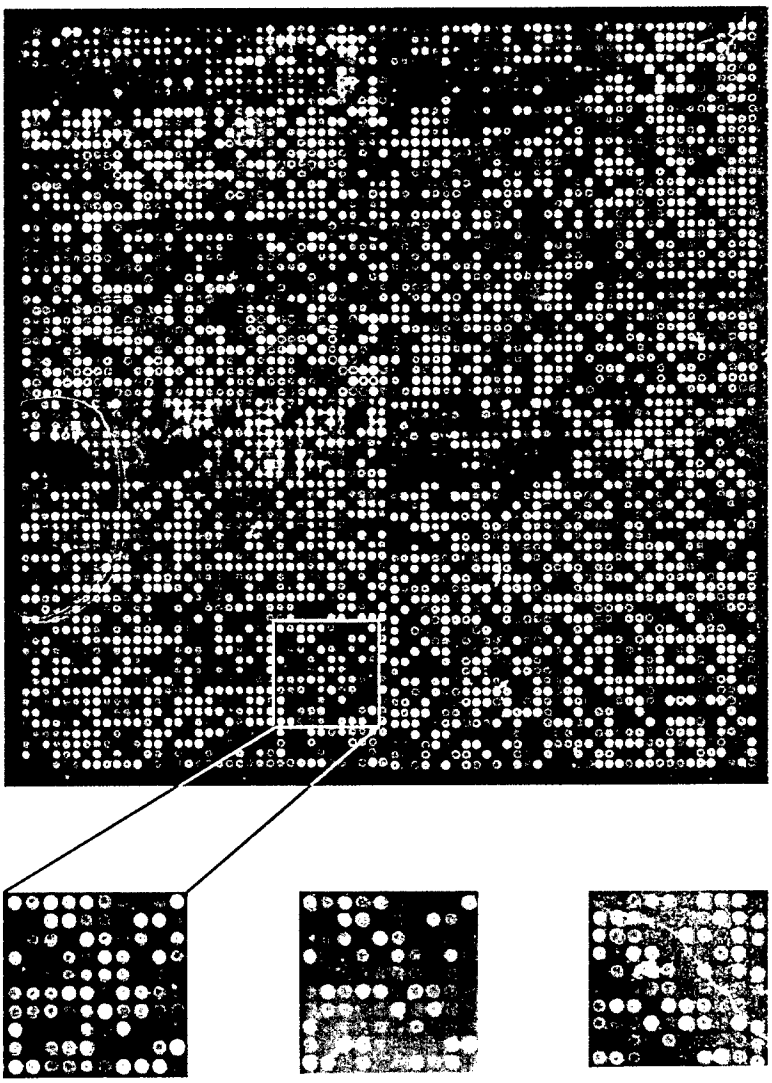
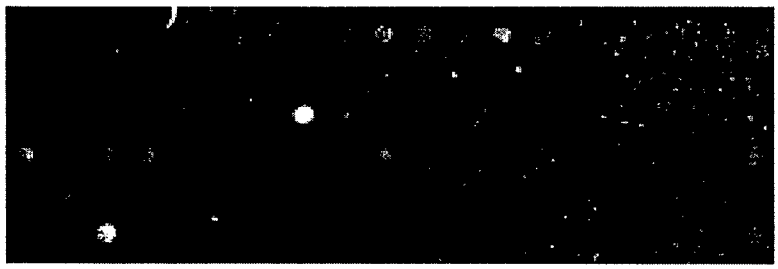
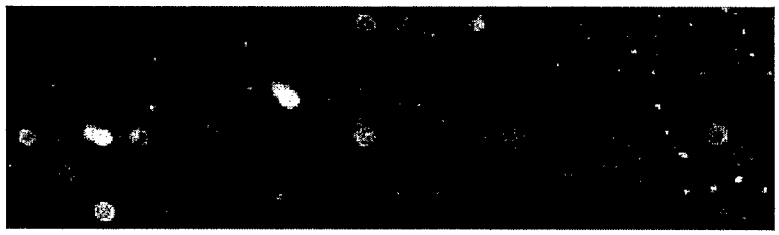


Figure 2.13 Comparison of background fluorescence observed when arrays were processed and pre-hybridized according to four alternative protocols.

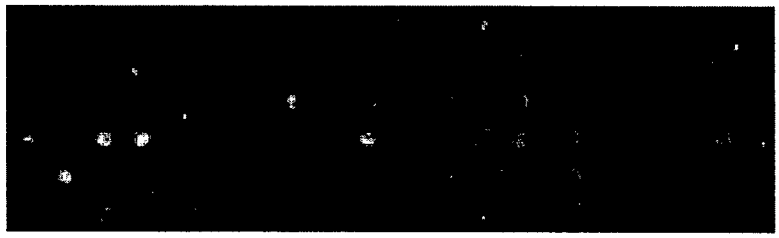
(a)



(b)



(c)



(d)

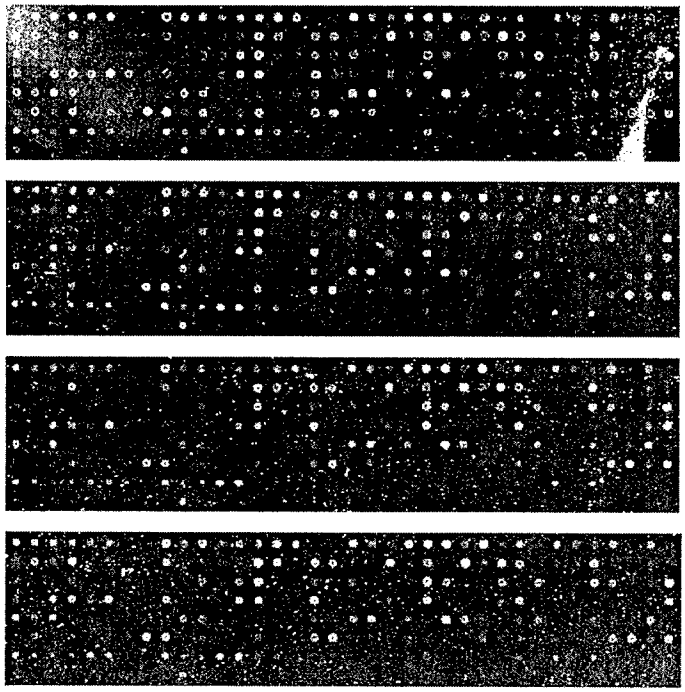


Figure 2.14 Schematic representation of the data flagging process implemented by Flagger.pl. Each feature was evaluated on three criteria, and only submitted to further analysis if passed by all three.

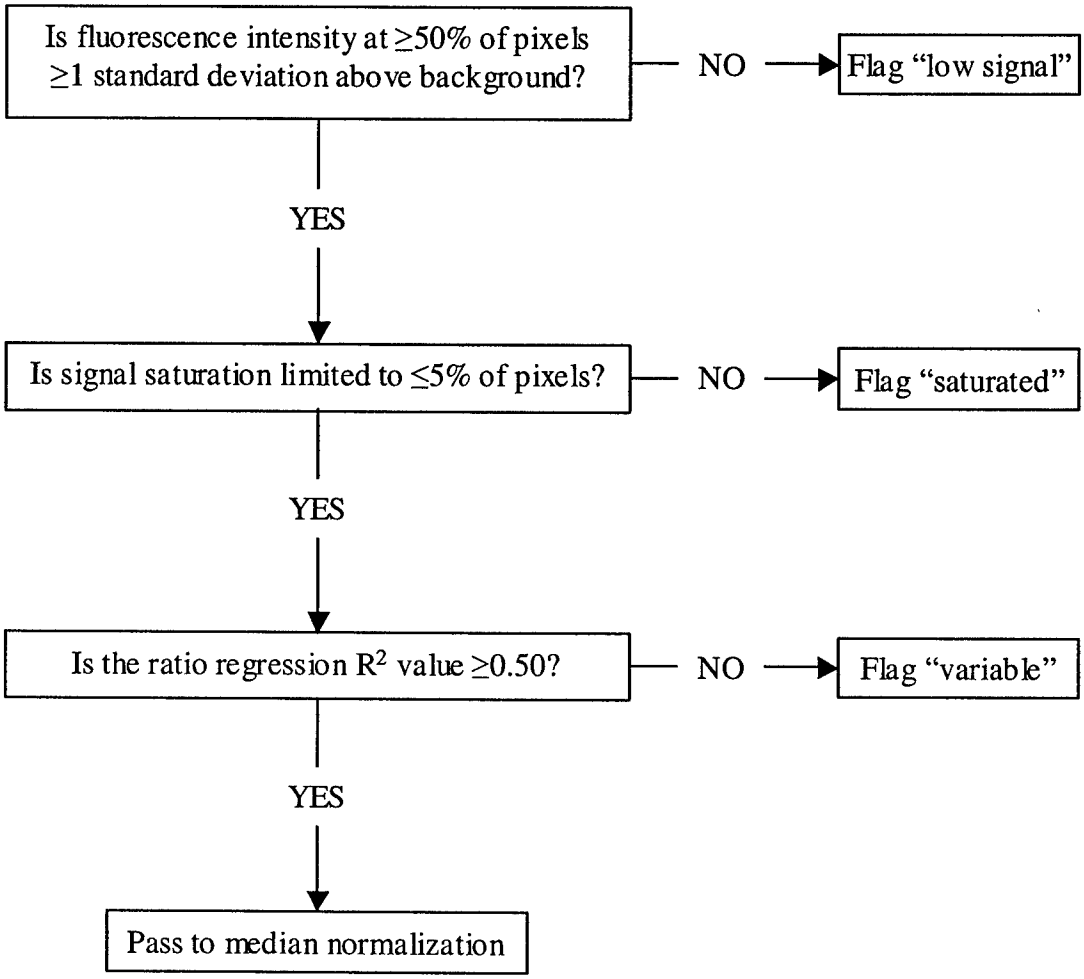

Is fluorescence intensity at ≥50% of pixels ≥1 standard deviation above background?
NO
Flag "low signal"
YES
Is signal saturation limited to ≤5% of pixels?
NO
Flag "saturated"
YES
Is the ratio regression R^2 value ≥0.50?
NO
Flag "variable"
YES
Pass to median normalization

Figure 2.15 Screen-dump of graphical user interface for microarray data analysis macros. VBA macros were developed by Heather Handley; integrated automation and graphical user interface by Chih Long Liu.

Version 1.01

Manual operation

Format GPR

Auto Flag GPR

Normalize GPR

Parameters – Autoflagging the GPR file

Remove spots <	85	uM diameter
Remove spots >	25	% saturating pixles in CH 2
Remove spots >	25	% saturating pixles in CH 1
Remove spots <	50	% pixel intensities in both channels are 1 SD above background

Process Files

Output file options

- ☐ Generate Stat files for replicates
- ☐ Generate pre-cluster files
- ☐ Generate formatted GPR (GPR-XLS) files

Generate File List

Note: for batch operation, this will assume that auto-flagging uses the same parameters for each file.

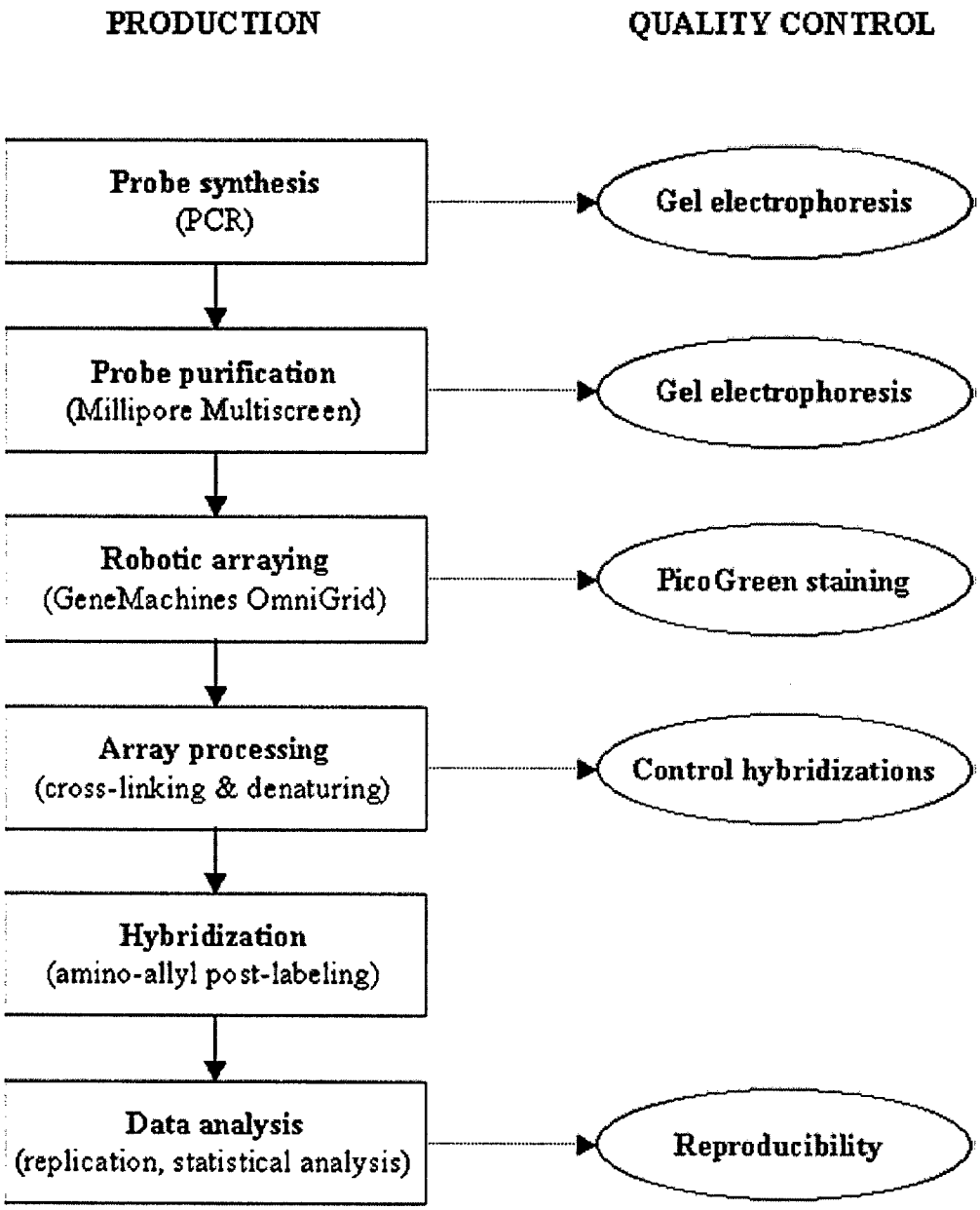
Please consult the instructions before running this macro for the first time.

Output Options	Output type
1	Ratio of Medians
2	Ratio of Means
3	Median of Ratios
4	Mean of Ratios

Output files generated:

Output Option	Group-Stat	Group-Cluster	Enter filename(s) here:	GPR-XLS	Stats	Cluster

Figure 2.16 Schematic diagram of microarray synthesis and hybridization workflow, showing techniques chosen for each production step and appropriate quality control measures.


PRODUCTION
QUALITY CONTROL
Probe synthesis
(PCR)
Gel electrophoresis
Probe purification
(Millipore Multiscreen)
Gel electrophoresis
Robotic arraying
(GeneMachines OmniGrid)
PicoGreen staining
Array processing
(cross-linking & denaturing)
Control hybridizations
Hybridization
(amino-allyl post-labeling)
Data analysis
(replication, statistical analysis)
Reproducibility

CHAPTER 3

The gene expression profile of 2,3,7,8 tetrachlorodibenzo-*p*-dioxin in zebrafish embryos is consistent with dilated cardiomyopathy

Abstract

2,3,7,8-tetrachlorodibenzo-*p*-dioxin (TCDD) is a ubiquitous environmental contaminant and a potent cardiovascular teratogen. Fish and birds exposed to TCDD during early embryogenesis develop severe edema and hemorrhage typical of congestive heart failure. There are indications that TCDD-induced dilative cardiomyopathy is the underlying cause of overt toxicity. However, determining specific cardiac impacts in fish has been difficult. I have used cDNA microarrays (Chapter 2) to establish the cardiovascular gene expression profile of 72 hpf zebrafish following early embryonic exposure to ~ED10 and ED100 doses of TCDD. Alterations in cardiovascular gene expression were limited; only 25 known genes and 19 ESTs were significantly differentially expressed $\geq$1.8-fold (p-values $\leq 5 \times 10^{-4}$), and only CYP1A and CYP1B1 were differentially regulated >4-fold. The dose-specificity of TCDD responses was highlighted, not only by the small number of genes significantly differentially expressed at both doses (7), but also by the ability of small deviations in achieved doses to account for the majority of variation between replicate hybridizations.

Microarray analyses indicated induction of three major functional classes of genes – xenobiotic detoxification enzymes, cardiac sarcomere structural proteins, and energy transfer genes. TCDD-modulated expression of selected genes in each category was further explored by RT-PCR. As expected, xenobiotic metabolism enzymes, including CYP1A, CYP1B1, and glutathione S transferase, were robustly and dose-dependently induced. Induction of mitochondrial electron transfer proteins was variable and modest, at or approaching limits of detection by either microarray analysis or RT-PCR. Most sarcomeric proteins appeared to be robustly induced, but RT-PCR indicated strong suppression of cardiac troponin T2. Despite this inconsistency, the current data suggest that TCDD causes dilated cardiomyopathy in zebrafish.

3.1 Introduction

2,3,7,8-Tetrachlorodibenzo-*p*-dioxin (TCDD) is a widespread anthropogenic contaminant in the marine environment, and a potent toxicant that disrupts cardiovascular development in teleost fish. Hallmark symptoms of TCDD embryotoxicity in fish include reduced heart size, circulatory failure, pericardial and yolk sac edema, hemorrhage, and early life stage mortality [10, 24, 25]. This suite of symptoms, similar to blue sac syndrome in salmonid fish, has been observed in over a dozen fish species exposed to TCDD and related pHAH [9-15]. In zebrafish, weakened cardiac contraction can be observed as early as ~50 hpf, followed by congestion and reduced perfusion of peripheral vascular beds, and finally, edema and hemorrhage [10, 22-24]. This progression is generally conserved across fish species [14, 25], and is reminiscent of congestive heart failure.

TCDD-induced dilated cardiomyopathy leading to congestive heart failure with edema and hemorrhage has been clearly demonstrated in avian embryos [32, 33]. Dilated cardiomyopathy appears to be the result of inhibited cardiomyocyte proliferation during a period of significant ventricular muscle growth and rearrangement [33, 35]. Corresponding cardiac remodeling processes in fish are poorly understood, and the exquisite sensitivity of fish to TCDD-induced edema has made pinpointing impacts on cardiac morphology difficult. However, similarity in the overt embryotoxicity of TCDD in fish, birds and mammals suggests that a common molecular mechanism is responsible.

TCDD toxicity is known to be largely dependent on the aryl hydrocarbon receptor (AHR) [54, 55, 57-59]. AHR is a basic-helix-loop-helix Per-ARNT-Sim family (bHLH-PAS) ligand-activated transcription factor with a broad affinity for aromatic hydrocarbons [36]. Binding of TCDD by cytosolic AHR causes activation, nuclear translocation, and dimerization with aryl hydrocarbon receptor nuclear translocator (ARNT). The AHR-ARNT complex acts via DNA sequence motifs, known variously as AHR-, dioxin-, or xenobiotic-response elements (AHRE, DRE or XRE), to modulate gene expression.

AHR is capable of regulating expression of numerous genes, and the primary toxicity-eliciting events are not certain. CYP1A induction by aromatic hydrocarbons is the most sensitive known response to AHR activation, and its potential role in TCDD toxicity has been subject to much investigation. Developmental expression and induction of CYP1A is strongly correlated, temporally, spatially, and dose-dependently, with symptoms of embryotoxicity [15, 26, 27, 68, 69]. Furthermore, blocking induction of CYP1A protein expression [54] or enzymatic activity [22] protects against TCDD-induced cardiovascular embryotoxicity. Thus, while all indications are that CYP1A is involved in TCDD embryotoxicity, the precise mechanism is unclear. Aberrant production of reactive oxygen has been suggested as a possible mode of action [64, 88], but this issue is still under investigation.

Recently, microarray-based gene expression profiling and serial analysis of gene expression (SAGE) has provided a list of several hundred TCDD-responsive genes [93-95, 123, 124]. Most of this work has focused on liver tissue and cultured hepatocytes [93, 95, 123, 124], but spleen and thymus tissues [95], and cultured lung epithelial cells [94] have also been interrogated. While certain general trends are emerging from these broad-scale studies, an abundance of disparities highlights the importance of dose and cell-type (and likely other biological factors) in shaping molecular responses to TCDD. As no comparable data are available regarding TCDD-modulated gene expression in either embryos or cardiovascular tissues, it is difficult to gauge the relevance of data from other systems to processes of cardiovascular embryotoxicity.

The goal of the current work was to use cardiovascular-specific cDNA microarrays (technical development described in Chapter 2) to identify genes whose expression is modulated by TCDD. Gene expression profiling of 72 hpf zebrafish embryos following early embryonic exposure to two doses of TCDD has revealed relatively limited alterations in cardiovascular gene expression; 21 known genes and 18 ESTs were significantly differentially expressed $\geq$1.8-fold (p-values $\leq 5 \times 10^{-4}$). The majority of known genes fall into three functional classes – xenobiotic detoxification enzymes, sarcomere structural proteins, and genes involved in cellular energetics. Selected genes

from each of these categories were further interrogated using real-time RT-PCR. Despite certain inconsistencies in the current data, the weight of evidence from this work suggests that TCDD causes dilated cardiomyopathy in zebrafish, as in birds.

3.2 Methods

Embryos and Chemicals

[^{3}H]TCDD (ChemSyn Laboratories) and unlabeled TCDD (ChemService) were obtained from providers in toluene solutions. For experimental use, toluene was evaporated and TCDD was reconstituted in DMSO, allowing ≥24hrs for complete dissolution prior to use. Stock solutions of [3H]TCDD alone and 50/50 tritiated/unlabeled TCDD were prepared at 0.2μM ,1μM, 2μM, 5μM, 10μM and 30μM.

Developmentally synchronous zebrafish embryos were obtained by performing crosses in mating tanks with removable barriers. Trios of two females and one male were held overnight in divided tanks. The following morning, barriers were removed and fertilized embryos were collected within 30min to ensure all embryos were within two cell cycles of each other.

At 2½-3 hours post fertilization (approximately 1000-cell stage), embryos were placed in glass petri dishes containing 0.05% (vol/vol) DMSO or appropriate TCDD stock solution in E3 egg water. Embryos were held in dosing solutions for 1.5 hrs at 28°C on an orbital shaker. Embryos were then removed from the dosing solutions and rinsed thoroughly with clean E3 before being transferred to clean plastic petri dishes. Embryos and larvae were maintained in clean E3 egg water at 28°C.

Approximately 24hrs after dosing, triplicate samples of 3 embryos per treatment group were removed to 20ml liquid scintillation vials, anaesthetized on ice and solubilized using 500μl Solvable reagent (Packard). Scintiverse II scintillation fluid (14.5mls) was added and samples were dark-adapted 2-24 hrs prior to liquid scintillation counting to establish accumulated embryo loads of [^{3}H]TCDD.

At 72 hpf, 50-100 larvae per treatment group were removed to 1.7ml tubes and excess egg water was aspirated. In initial experiments, 1ml RNALater was added and embryos

were stored at -20°C. Embryo samples for the 5nM microarray experiment and for follow-up experiments were flash frozen and stored at -80°C prior to RNA preparation.

DNA Microarray Hybridizations

AH001 and AH001A cardiovascular-specific cDNA microarrays (Chapter 2) were used for this work. These arrays consisted of 5,184 PCR products representing approximately 2,000 distinct zebrafish genes. With the exception of control genes, the sequence of arrayed clones was unknown prior to hybridization and subsequent analyses.

Total RNA was isolated from embryo homogenates using TriZol reagent (Invitrogen) according to manufacturer's protocol. For long-term storage, RNA pellets were kept in 70% ethanol at -80°C. After removal of ethanol, RNA was dissolved in water and stored frozen.

Amino-allyl modified cDNA was generated by reverse transcription in the presence of amino-allyl-dUTP. Total RNA (15–25μg) spiked with *A. thaliana* chloroplast mRNA (100, 250, and 500 ng of Cab, RCA, and rbcL RNA, respectively; SpotReport®-3 Array Validation Kit, Stratagene) was incubated with 5μg oligo-$dT_{(20)}N$ anchored primer for 10 min at 65°C, then chilled on wet ice. 5x First-strand buffer (4μl)and 0.1mM DTT (2μl) were added for a final reaction volume of 20μl. Remaining reagents were then added for final reaction conditions of 1x first-strand synthesis buffer, 10mM DTT, 0.5mM each dATP, dCTP and dGTP, 0.3mM dTTP, 0.2mM amino-allyl-dUTP, and 1000U Superscript II reverse transcriptase. Reverse transcription reactions were run 2.5 hrs at 42°C, then inactivated by buffering with 0.5M EDTA and incubating 5 min at 95°C.

RNA was eliminated by alkaline hydrolysis in 0.2N NaOH (incubated 15 min at 65 °C, then neutralized by equimolar HCl and buffered with Tris-HCl), followed by RNase digestion (xU Ambion RNase cocktail, 30 min at 37°C). cDNA purification and buffer exchange was accomplished by filter-purification according to standard protocols (QiaQuick PCR Purification Kit, Qiagen), except that Qiagen buffers PE and EB were replaced by 75% ethanol and distilled water, respectively. cDNAs were dried by vacuum centrifugation and stored at -20 °C.

For CyDye post-labeling, cDNAs were redissolved in 10 μl 0.1M NaHCO3 (pH9.0) containing an individual aliquot of previously dried amine-reactive Cy3 or Cy5 (Amersham Biosciences), then incubated 1.5-2 hrs at room temperature in full darkness. The labeling reaction was quenched by addition of excess hydroxylamine (4.5 μl at 4M) and 15 min at room temperature in full darkness. Following addition of 35 μl 100mM NaOAc (pH 5.2) and 50 μl nuclease-free water, labeled cDNA was purified using QiaQuick PCR Purification columns (Qiagen) according to standard protocols. cDNA concentrations were determined spectrophotometrically (A_{260}, A_{280}), then equal quantities of paired Cy3- and Cy5-labeled cDNAs were combined and dried by vacuum centrifugation.

Immediately prior to hybridization, labeled target cDNA was redissolved in hybridization buffer (3x SSC with 0.1% SDS) containing 0.4 μg/μl each polyA blocker (Sigma) and yeast tRNA (Invitrogen), and 0.8 μg/μl sonicated salmon sperm DNA (Fisher Scientific). This mixture was denatured by heating 2 min at 95°C, then quickly pipetted onto the microarray surface and covered with an appropriate cover slip. Arrays were hybridized (18 hrs at 65°C) in sealed hybridization chambers containing a reservoir of 2x SSC.

Following hybridization, slides were washed 2 min + 3 min in 2x SSC with 0.1% SDS, then 2 min + 1 min in each 1x and 0.1x SSC. Slides were dipped into distilled water, then isopropanol, then air-dried and stored in darkness with dessication prior to scanning.

Microarray Data Analysis

Array scanning and image analysis was performed using Axon GenePix 3.0 software. Axon results files were either imported to Microsoft Excel for basic statistical analyses (see Appendix A for VBA scripts), or submitted to the Rosetta Resolver database and analysis package (administered by the Harvard Center for Genome Research).

Real-time RT-PCR

Total RNA was prepared using TriZol reagent (Invitrogen), then incubated 30 min at 37°C with 2 Units DNA-*free* DNase I (Ambion) to remove genomic DNA contamination. DNase was inactivated by addition of 4.5 μl DNase Inactivation reagent; after ~5 min, DNase Inactivation reagent was pelleted by centrifugation. cDNA was generated from 2 μg DNase-treated total RNA according to standard reverse transcription protocols (Superscript II RT, Invitrogen)

PCR reactions consisted of 1x SYBR® Green PCR Master Mix (Applied Biosystems), 1 μl cDNA, and 400 nM each primer (Table 3.1). Initial enzyme activation (2 min at 50°C) and DNA denaturation (1min at 94°C) steps were followed by 40 two-step amplification cycles (15 sec at 94°C, 1 min at 60°C). The dissociation curve of each PCR product was determined after amplification (ABI Prism 7000).

For relative quantitation, SYBR® Green detection was accompanied by ROX passive detection and normalization. PCR efficiency was assumed to be 2, and the threshold-crossing cycle number (Ct) was determined at SYBR® Green fluorescence R=0.2. Relative expression ratios (R) were calculated according to the previously described [149] equation:

$$R = \frac{(E_{target})^{\Delta Ct_{target}(DMSO - TCDD)}}{(E_{ref})^{\Delta Ct_{ref}(DMSO - TCDD)}}$$

3.3 Results

TCDD cardiovascular embryotoxicity

The aim of this study was to identify alterations in gene expression correlated with specific cardiovascular impacts resulting from TCDD exposure. In order to determine appropriate doses for expression profiling, I assessed the TCDD sensitivity of the Tübingen long tail (TL) strain used for this work. Developmentally synchronous zebrafish embryos were exposed to either 0.05% DMSO (vehicle) or varying concentrations of [^{3}H]TCDD (0.1nM, 0.5nM, 1nM, 2nM, or 5nM) in egg water.

Embryotoxic endpoints, including pericardial and yolk sac edema, impaired peripheral circulation, and mortality, were assessed at ~80 and ~96 hpf.

Susceptibility of TL embryos to edema and mortality was similar to that documented for other zebrafish strains [10]. Early life stage mortality was insignificant at both time-points (data not shown). Less than 10% of DMSO-treated (control) embryos exhibited mild edema, and 0.1nM TCDD did not significantly enhance pericardial edema at either time. At 80 hpf, ≥0.5nM TCDD produced a dose-dependent increase in the frequency and severity of pericardial edema (Figure 3.1, arrows).

Experimental design and sources of variability

Samples for microarray analysis were obtained from two separate experiments. In the first, embryos from multiple clutches were pooled, then randomly divided into groups of ~100 embryos and exposed to either 0.05% DMSO or 0.5nM TCDD (Figure 3.2). In the second experiment, groups of ~100 embryos from each of 4 individual clutches were exposed to 5.0nM TCDD; ~400 embryos pooled from the same clutches comprised a single control group (Figure 3.2). Overall mean embryo [3H]TCDD burdens for the two experiments were 1.84±0.42 ng/g and 10.74±1.38 ng/g (

Table 3.2).

The amount of labeled cDNA used in microarray analyses varied with RNA availability and reverse transcription efficiency. All hybridizations for low-dose samples utilized ≤500ng cDNA (Figure 3.2a), while 3 of 4 initial high-dose hybridizations were performed with 900-1000ng cDNA (Figure 3.2b, top row). To assess the effect of cDNA quantity on gene expression results, excess cDNA from two 5.0nM TCDD samples was used for additional hybridizations with ≤500ng cDNA (Figure 3.2b, TCDD B-2 and C-2). Mean expression ratios calculated from the three high-dose hybridizations performed with ≤500ng cDNA were compared to results from the three hybridizations using ~1μg cDNA. The two data sets were closely correlated (Figure 3.3), and yielded significantly different results for only 66 clones (single-factor ANOVA, p-values ≤0.01). Thus, all high dose data were combined regardless of the quantity of target cDNA used

($N_{5.0nM} = 6$). These results also indicated that quantitative comparison between dose levels was valid, despite technical differences.

Data from all low-dose (0.5nM TCDD) replicate hybridizations were highly (>80%) cross-correlated, while correlation coefficients for 5.0nM TCDD replicate hybridizations ranged from 15% to 59% (Table 3.3). Regression analysis indicated that variation in correlation values was largely a function of slight differences in accumulated doses (Figure 3.4). Correlation between low-dose samples was strongly dependent on both differences in embryo TCDD levels ($R^2 = 0.93$) and on the control sample used ($R^2 = 1.0$, not shown). Dosage disparities accounted for a smaller proportion of the variation between high-dose replicate hybridizations ($R^2 = 0.67$), and control sample variation was null, as RNA from a single pool of DMSO-treated embryos was used for all hybridizations (Figure 3.2). Additional variation between high-dose replicate hybridizations may be attributable to the use of individual clutches, rather than pooled embryos, for replicate groups.

≥2-fold differential expression

A common method of analyzing DNA microarray data is to identify all genes whose expression is induced or suppressed by ≥2-fold; this approach has been used in most studies of TCDD-modulated gene expression [93, 95, 123]. In accord with this standard, 88 arrayed clones with at least one dose-specific mean expression ratio ≥2.0 or ≤0.5 were sequenced and assigned gene identities based on protein-level homology. The largest portion of clones (43%) was composed of ESTs with no significant similarity to known proteins (Figure 3.5); over half of these clones fell into four EST clusters, named TR001-TR004. In addition, seven known proteins were identified, including AHR2, CYP1A, ovarian aromatase (CYP19a), and cardiac troponin T2 (Figure 3.5). However, 18 clones (20.5%) were found to contain no insert, and thus, to be false positive results (Figure 3.5).

Statistical confidence and systematic variance

In order to reduce the false positive rate, the Rosetta Resolver software package was used to determine statistical confidence intervals for each microarray feature, then

calculate confidence-weighted mean expression ratios and p-values. An appropriate threshold statistic for filtering the complete dataset was determined empirically from results for control genes and known false positives. Five out of six negative controls had p-values $\geq 1x10^{-3}$; *A. thaliana* clone #2 had a p-value of $2x10^{-4}$ at 0.5nM TCDD (Table 3.4). Of all the "no insert" false positives, only one had a p-value $\leq 1x10^{-3}$ (AH042259 high-dose mean expression ratio = -1.86, p-value = $3.11x10^{-6}$). In contrast, confidence statistics for positive controls fell in a range of p-values $\leq 1.8*10^{-5}$ (Table 3.4). Thus, an intermediate statistical confidence threshold of p-values $\leq 5x10^{-4}$ was adopted.

A total of 496 clones were assigned at least one dose-specific p-value $\leq 5x10^{-4}$. No additional false positives were detected among the >250 clones for which high-quality DNA sequence data was obtained. Only 65 clones met the statistical significance criterion at both dose levels; 276 clones had p-values $\leq 5x10^{-4}$ at 0.5nM TCDD only, 155 at 5.0nM TCDD only (Figure 3.6). At both dose levels, 65-75% of all differentially expressed clones were induced by TCDD. However, the majority of changes at this statistical confidence level were relatively subtle, of magnitude (absolute value expression ratio) 1.3- to 1.6-fold (Figure 3.7).

From among the clones for which high-quality sequence data was available, 95 clones were assembled into 12 contigs corresponding to 11 known genes and one EST (Table 3.5). Known genes fell into two functional classes – mitochondrial genes, and sarcomeric proteins. Of the 13 protein-encoding genes in the mitochondrial genome of zebrafish [150], at least seven showed 1.2- to 1.5-fold induction by TCDD (Table 3.5a). Most mitochondrial genes were represented by at least four clones, all indicating similar magnitude up-regulation. Ambiguity in the number of impacted mitochondrial genes derived from inability to distinguish genes for NADH dehydrogenase subunits 4 and 4L based on available sequence data.

Effects on structural components of cardiac muscle sarcomeres were mixed (Table 3.5b). Four clones corresponding to cardiac troponin T2 were robustly up-regulated, particularly at the lower dose level. Observed induction of myosin was more modest, and cardiac α-actin expression was very slightly suppressed.

Approximately 4% (19 clones) of all clones with p-values $\leq 5x10^{-4}$ fell into a single EST cluster, TR004, which was robustly and dose-dependently induced (Table 3.5c). These 19 clones were all ≥85% identical at the nucleotide level, and jointly spanned ~1.8kb of sequence. TR004 sequences showed no significant similarity to known proteins, but very weak similarity to mammalian endogenous retroviral *ENV* genes.

The prevalence of low-magnitude changes, both generally and among mitochondrial genes, raised the issue of limits of detection. Thus, expected variation in measurement of null (i.e. no change) results was determined from two observations. Firstly, mean expression ratios for six negative control genes ranged between -1.61 and +1.38 (Table 3.4). Additionally, 1.8-fold change defined the 99.7% confidence interval (3 standard deviations) for two homotypic control hybridizations (Chapter 2). Thus, a conservative limit of detection of differential expression was established at ≥1.8-fold change. While nearly 400 clones had mean expression ratios ≥1.8 or ≤0.55, only 73 clones exhibited statistically significant (p-value $\leq 5x10^{-4}$) differential expression ≥1.8-fold at either dose level.

TCDD-induced differential gene expression

These 73 significantly differentially expressed clones corresponded to 25 known genes or ESTs similar to known proteins, and 19 ESTs with no homology to known proteins (Table 3.6). The identity of one clone was undetermined as a result of poor sequence quality. As in previous analyses, the majority of known genes fell into three major functional categories – xenobiotic detoxification enzymes, sarcomeric structural proteins, and genes involved in electron transfer and energetics (Table 3.6). The remaining seven genes represented diverse cellular processes, including steroid synthesis (20 β-hydroxysteroid dehydrogenase), and erythrocyte morphology and function (pyrimidine 5' nucleotidase), transcriptional regulation (cryptochrome 1a), and water transport (AH042420). The EST cluster TR004 was also highly represented in this analysis (Table 3.6f).

The same general trends noted previously were apparent in this analysis. Nearly 75% of all differentially expressed genes, including sarcomeric proteins and most

detoxification enzymes, were significantly induced at one or both doses (Figure 3.8). The excess of up-regulation responses was most pronounced at the higher dose level; only two genes – 12s and 16s ribosomal RNAs – were significantly suppressed ≥1.8-fold by exposure to 5.0nM TCDD (Table 3.6d).

The most common expression pattern (27 genes) was significant differential expression at only the lower dose; only seven genes were significantly differentially expressed at both doses (Figure 3.8). There was also a general trend toward lesser magnitude changes at the higher dose. Induction of both CYP1A and CYP1B1 was strongly and directly dose-dependent. In contrast, the mean (± standard deviation) magnitudes of change for all other differentially expressed genes were 2.15±0.85 and 1.73±0.69 at 0.5nm and 5.0nM TCDD, respectively (single-factor ANOVA, p-value <0.01).

Dose-dependent differences were only statistically significant (single-factor ANOVA, p-value <0.01) for 4 genes and 3 ESTs. 20 β-hydroxysteroid dehydrogenase was induced in a directly dose-dependent manner. Cytochrome C oxidase, NADH dehydrogenase, and ESTs AH045277 and AH046249 were more strongly induced at the lower dose. Suppression of ATP synthase also appeared to be inversely dose-dependent.

EST AH041068 showed a trend toward dose-dependent reversal of response direction (Table 3.6e). While induction by 5.0nM TCDD was not statistically significant, low- and high-dose mean expression ratios of -1.88 and +2.06, respectively, were significantly different from each other (single-factor ANOVA, p-value <0.01). Conversely, an EST similar to aquaporin 8 (AH042420) was significantly induced by 0.5nM TCDD, and showed a non-significant trend toward suppression by 5.0nM TCDD (Table 3.6d)

Follow-up by real-time RT-PCR

Real-time RT-PCR was used to further define dose-response curves for TCDD-responsive genes identified by microarray analyses. Two pools of synchronous embryos were split into six treatment groups – 0.05% DMSO or 0.5nM, 1.0nM, 2.5nM, 5.0nM or 15nM TCDD. Duplicate PCRs were run using aliquots of cDNA samples from each treatment group. To assess genomic DNA contamination, PCR was also run on samples

from which reverse transcriptase was withheld; significant genomic DNA amplification was never observed prior to PCR cycle 25 and did not interfere with cDNA quantification (data not shown).

ARNT2 and β-actin served as negative controls. Neither biological nor technical replicates were significant sources of variation for either gene (two-factor ANOVAs, p-values >0.05), and no significant change in ARNT2 or β-actin expression levels was observed at any dose level (Figure 3.9a). Thus, for all other genes, technical and biological replicates were combined (i.e. n=4), and ARNT2 expression ratios were used as references for normalization.

CYP1A expression ratios were determined from only one technical replicate (i.e. n=2, biological replicates), as accidental omission of PCR primers resulted in failure of reactions in replicate plate #2. CYP1A mRNA was significantly induced in each replicate, but the magnitude of induction varied significantly between the two samples (two-factor ANOVA, p-values <0.05). In both cases, CYP1A induction increased dose-dependently up to 1.0-2.5nM TCDD, then declined at ≥5.0nM TCDD (Figure 3.9b).

As microarray results suggested subtle and variable induction of mitochondrial proteins, RT-PCR was used to independently assess TCDD modulation of these genes. Only one subunit of NADH dehydrogenase was assayed, as microarray results were similar for all subunits. Mean expression ratios for all four mitochondrial genes generally ranged 1.2-1.5 (Figure 3.9c), but were never significantly different from null (single-factor ANOVAs, p-values >0.05). Similarly, the subtle down-regulation of cardiac α-actin suggested by microarray data could not be confirmed by RT-PCR (Figure 3.9e).

In stark contrast to microarray data, RT-PCR indicated strong suppression of cardiac troponin T2 (Figure 3.9d). This suppression was highly statistically significant at 1.0nM, 2.5nM, and 15nM TCDD (one-tailed paired T-tests, p-values ≤0.001). As induction of troponin C has been observed in other TCDD gene expression profiling work [93], TCDD regulation of this isoform was also assayed; no significant effect was observed (Figure 3.9e).

3.4 Discussion

Limits of detection

The current data were filtered according to dual criteria of magnitude of change and statistical confidence. Threshold values for each criterion were determined empirically, as there are no applicable theoretical guidelines. While there is some debate regarding what level of change is biologically relevant, this question is generally obscured by technical limitations on detection of differential expression. The current limit of detection was based on a small number of negative controls and two homotypic hybridizations. As control and experimental hybridizations made use of multiple array sets and cDNA sources, gene-specific systematic variance and/or bias was not assessed. It is quite possible that many changes of lesser magnitude than the current detection limit of 1.8-fold are both statistically and biologically relevant.

However, as confidence statistics calculated by Rosetta Resolver cannot be interpreted as standard p-values, the question of what constitutes a statistically significant result is not straightforward. Resolver p-values represent the probability that a modeled normal distribution for a given gene includes the null change value log(ratio) = 0, and thus, do not take into account variance in the measurement of null change. As a result, they may overestimate actual significance to an unknown degree. Indeed, application of p-value thresholds of 0.05 or 0.01 would have resulted in a high false positive rate.

Detection of differential expression at the high dose was restricted by greater variance between replicate hybridizations, presumably due to genetic variation between individual clutches. Such variability was somewhat surprising given the fact that all parental fish were members of a single inbred strain. There is no notable difference in sensitivity of TL clutches to TCDD embryotoxicity. However, the occurrence of phenotypes, such as long fins and lack of skin pigmentation, that result from known background polymorphisms varies considerably between clutches. While strictly anecdotal, these observations suggest the presence of significant genetic variation within the TL strain. Combining clutches, as was done in the low-dose experiment, artificially removes evidence of such variation, thereby overestimating the biological significance of certain

changes. This may, in part, account for the need to apply unusually stringent thresholds to avoid false positives, nearly all of which were detected at the lower dose. However, such stringent analyses are unlikely to severely impact detection of changes in gene expression involved in TCDD embryotoxicity; such changes should be as consistent among TL clutches as is TCDD susceptibility.

General trends in TCDD-modulated gene expression

At 44 genes or ESTs, the number of TCDD-regulated genes was somewhat lower than that observed in other broad-scale gene expression studies. Martinez and colleagues [94] identified 68 differentially expressed genes from a microarray of 2091 genes, similar to the estimated 2000 genes represented on our zebrafish cardiovascular arrays. Other studies using larger microarrays have documented differential expression of 0.9% [123] and 6.9% [93] of all investigated genes; the current results fall at the low end of this range. To some degree, the limited number of observed changes may be related to the restricted (i.e. cardiovascular-specific) focus of the microarrays used.

An emerging feature of TCDD-modulated gene expression is a pattern of widespread, subtle changes in gene expression, as opposed to strong pressure on specific pathways. In the current work, only two genes – CYP1A and CYP1B1 – were significantly induced >4-fold. Similar patterns have been noted in other TCDD gene expression profiling experiments [93, 123]. Signal compression in microarray analyses has been an issue of some concern, but does not appear to explain this trend. Our microarray-based measurements of CYP1A induction were similar to both current and previously published RT-PCR data [42]. Additionally, Martinez and coworkers [94] confirmed the magnitude of several subtle changes detected by microarray analyses. Thus, this and other gene expression profiling work is reaffirming the highly pleiotropic nature of impacts by TCDD, as well as the unique sensitivity of CYP1A to TCDD.

The relative importance of AHRs' roles as a transcriptional enhancer and suppressor is a matter of some debate. Whereas most studies have recorded nearly equal numbers of up- and down-regulated genes, Puga and coworkers [93] identified twice as many TCDD-suppressed genes as induced genes in hepatoma cells. In contrast, our work and that of

Frueh and colleagues [123] indicated that the number of inductions exceeded suppressions by nearly two-fold. The variation in outcomes of microarray studies suggests that the balance between induction and suppression varies depending on biological context. However, speculation on this topic is confounded by the fact that microarray-based detection of down-regulation is limited by the constitutive level of gene expression (i.e. suppression of low-level expression will not be readily detected).

Not surprisingly, dose level appears to be a primary factor in determining transcriptional responses to TCDD exposure. Only seven genes (<20% of changes) were significantly differentially expressed at both doses investigated here. Furthermore, relatively small differences in actual doses resulting from the same nominal exposure accounted for the majority of variation between hybridizations. Similarly, in comparing three dose levels spanning two orders of magnitude, Martinez and colleagues [94] found that more than half of all TCDD-regulated genes were differentially expressed at only one dose level. The limited number of genes that are responsive to TCDD over a large range of exposure levels may be of interest as potential biomarkers; certainly, this has been the case for CYP1A.

Interestingly, traditional dose-response curves (i.e. direct relationship between dose level and magnitude of change) were primarily limited to detoxification enzymes (CYP1A, CYP1B1, GSTπ). For most other genes, particularly sarcomeric proteins and several ESTs, greater differential expression was observed at the lower dose (0.5nM TCDD). As pooling clutches accomplishes essentially the same end as averaging data from multiple clutches, differences in experimental design do *not* account for the prevalence of reduced magnitude changes at the higher dose. Thus, the current data suggest that non-traditional dose-response relationships (i.e., stronger responses at lower doses) are a prevalent feature of TCDD-modulated gene expression. Martinez and coworkers [94] reached a similar conclusion in examining gene expression profiles in TCDD-exposed lung epithelial cells. Such low-dose specific responses likely represent adaptive responses to chemical stress, a category of responses that has been largely

ignored in favor of toxic mechanisms. However, the prevalence and importance of adaptive responses is beginning to gain recognition [151].

Xenobiotic responsive gene expression

Observed induction of xenobiotic metabolism enzymes largely met expectations. CYP1A induction has been observed in all vertebrates studied, and induction of glutathione S transferase (GST) is well documented in mammalian species [152, 153]. Accordingly, current data consistently indicated directly dose-dependent increases in expression of CYP1A and GSTπ. Microarray data also suggested robust dose-dependent induction of CYP1B1. In contrast to CYP1A, inducibility of CYP1B1 by AHR agonists does not appear to be universal phenomenon. However, CYP1B1 induction has observed in several systems [154-159].

Expression of xenobiotic detoxification genes other than the AHR gene battery was also altered in response to TCDD exposure. Induction of major vault protein was somewhat unexpected, and very intriguing, given mixed results with regard to induction of other multi-drug resistance proteins by TCDD and related chemicals [160, 161]. Glutathione peroxidase was unique among xenobiotic detoxification genes, in that low-level TCDD treatment strongly suppressed expression. TCDD treatment also reduces glutathione peroxidase expression in murine liver tissue [124]. However, this result may be better understood in the context of cardiovascular biology, as suppression of glutathione peroxidase has been observed in cases of dilated cardiomyopathy in mammals [162].

Perturbed cellular energetics

While the magnitude of induction of mitochondrial energy transfer proteins is uncertain, all analyses of current microarray data indicated some degree of up-regulation of NADH dehydrogenase and cytochrome C oxidase. RT-PCR data did not support a strong induction of these genes, but were consistent with a small (i.e. 25-50%) increase in mitochondrial gene expression. Thus, it seems likely that mitochondrial energy production processes in zebrafish embryos are subtly enhanced by TCDD. Stronger

alterations in expression of a Ca^{2+}-ATPase (SERCa), and ESTs similar to aconitase and adenylate kinase enzymes suggest additional disruption of downstream energy transfer processes. Similarly, induction of mitochondrial electron transfer proteins, including cytochrome c oxidase and cytochrome b, and various perturbations in other metabolic enzymes has been documented previously [123, 124].

Elevated mitochondrial gene expression might contribute to reactive oxygen-mediated processes of toxicity. Increased respiration-dependent reactive oxygen production by mitochondria has been observed in TCDD-treated mouse liver [163, 164]. In heart mitochondria, TCDD-induced reactive oxygen production has been linked to decoupling of respiration and downstream oxidative phosphorylation processes [165]. Such a situation could arise from concurrent induction of electron transfer enzymes and suppression of downstream metabolic enzymes, such as ATP synthase. However, any speculation with regard to ATP synthase is premature given rather contradictory results for this gene. Nonetheless, observations of altered expression of electron transfer proteins suggest a potential molecular mechanism for TCDD-induced mitochondrial reactive oxygen production, a problem that has heretofore been studied only on the level of enzymatic activity.

Perturbations in energy production and transfer are also in accord with current understanding of cardiomyopathy and heart failure. Heritable mutations in mitochondrial genes are associated with congenital cardiomyopathies [166]. Induction of mitochondrial electron transfer proteins, including NADH dehydrogenase, has been seen in both dilated and hypertrophic cardiomyopathies [162]. Decreased functionality of energy transfer enzymes, including adenylate kinase, is also typical of failing myocardium [167, 168].

Altered cardiovascular gene expression

Alterations in expression of cardiac sarcomere proteins is also indicative of TCDD-induced cardiomyopathy. Loss-of-function mutations in cardiac troponin T2, cardiac actin and cardiac myosin heavy chain are leading causes of human congenital cardiomyopathies [169]. Paradoxically, significant induction of sarcomeric proteins, including troponins and myosins, has been observed in both dilative and hypertrophic

cardiomyopathies in mammals [162, 170, 171]. Cardiac myosin levels are also increased in chick embryos with TCDD-induced dilated cardiomyopathy [32], and troponin C expression is elevated in TCDD-treated hepatoma cells [93].

Microarray data for multiple myosin isoforms and cardiac troponin T2 are in accord with other observations of induced expression of sarcomeric proteins in cardiomyopathy. In contrast, RT-PCR data indicated strong suppression of cardiac troponin T2, a condition more similar to loss-of-function mutations that cause cardiomyopathies. While conflicting data from the two methods is a significant technical concern (see below), it does not preclude the drawing of some tentative biological conclusions. Regardless of the direction of change, disruption of normal expression of cardiac troponin T2 and cardiac myosins would be indicative of cardiomyopathy.

Nonetheless, directly conflicting microarray and RT-PCR data regarding cardiac troponin T2 are troubling and difficult to justify. One explanation is that this gene suffers disproportionately from some systematic dye bias in microarray analyses. However, post-labeling protocols, such as that used in the current work, were specifically designed to avoid dye biases that are pervasive in direct labeling systems; the same modified nucleotide is used to generate both control and treated cDNAs, and subsequent dye coupling is not subject to significant steric hindrance. In the present case, a low rate of systematic bias was quantified by homotypic control hybridization, and taken into account by the use of a magnitude-of-change threshold. Nonetheless, it would be interesting to perform a dye-swap experiment to rule out dye bias as the source of this conflict.

It is also possible that differences in assayed sequences introduce discrepancies in end results. Whereas arrayed cDNA clones for cardiac troponin T2 were 600-900bp in length, PCR products for real-time RT-PCR analyses were constrained to ~100bp. However, priming sites for RT-PCR fell within arrayed sequences. Thus, how this difference would produce such drastically disparate results is unclear.

TCDD-regulated ESTs

Altered expression of numerous ESTs is an intriguing aspect of this work, as each EST presents an opportunity for novel gene discovery and insight into unexplored aspects of TCDD embryotoxicity. In many cases, additional sequence data would likely reveal these to be 3' untranslated regions of proteins that have been characterized in other species. However, it is worth noting that all annotations of the human genome include several thousand predicted genes with no homology to known proteins. Our knowledge of the vertebrate gene repertoire is far from complete.

Certain ESTs presented particularly interesting opportunities for further investigation. EST AH041068 demonstrated a unique dose-response relationship – suppression at low dose, induction at high dose, the origin of which would be fascinating to investigate. The EST cluster TR004 is another interesting case. The number of TR004 clones encountered suggests expression at a level comparable to cytochrome C oxidase. Furthermore, robust, dose-dependent induction suggests the possibility of involvement in TCDD toxicity. However, the identity of this EST could not be established from available sequence data.

Conclusions

Gene expression profiling of TCDD-exposed zebrafish embryos has provided a unique perspective on TCDD transcriptional modulation; the majority of other available data pertains to mammalian liver cells and tissue. Incorporation of the current data into comparisons of broad-scale gene expression data from multiple systems lends weight to several emerging general trends. For example, the vast majority of TCDD-induced changes detected in this and other studies are relatively subtle (i.e., <5-fold); the significance of this observation is uncertain. However, against a background of many small changes, induction of CYP1A stands in stark contrast. Thus, surveys of thousands of genes in multiple cell and tissue types are reaffirming the primacy of CYP1A induction among molecular effects of TCDD. Another important aspect of TCDD-modulated gene expression is the prevalence of low-dose specific, presumably adaptive, responses. Growing recognition of the importance of adaptive responses to TCDD and other chemical stressors may (hopefully) influence future approaches to risk assessment.

In addition to further elucidating general trends, this work has contributed to our understanding of TCDD cardiovascular embryotoxicity specifically. Altered expression of proteins that compose cardiac muscle sarcomeres was a consistent feature TCDD gene expression profiles in zebrafish embryos. In the case of cardiac troponin T2, microarray and RT-PCR data were (inexplicably) in direct disagreement. Thus, the nature of TCDD-induced alteration in cardiac troponin T expression requires clarification. However, taken as a whole, this work provides preliminary indications that early embryonic TCDD exposure causes dilated cardiomyopathy in zebrafish, as in birds.

This and other gene expression profiling work has provided hints of perturbed cellular energetics resulting from changes in expression of electron transfer proteins. In this case, most clones of mitochondrial genes indicated only subtle (1.3- to 1.5-fold) induction, but a few showed more robust responses. Unfortunately, such slight changes were well beneath the detection sensitivity of RT-PCR. However, induction of mitochondrial proteins might be important either as a step toward TCDD-induced mitochondrial reactive oxygen production or as a further indicator of cardiomyopathy. Certainly, the weight of evidence is sufficient to urge further investigation in this area. It will be particularly important to determine whether slight changes in mitochondrial gene expression are sufficient to generate detectable changes in respiration rates, and whether such changes are causally or secondarily related to TCDD toxicity.

Table 3.1 Primer sequences and PCR product information for real-time RT-PCR analyses. Primers for CYP1A and ARNT were provided by Dr. Mark Hahn.

Gene name	Sense (top) and anti-sense (bottom) primer sequences	Annealing temp. (°C)	Product size (bp)
β-actin	ATGGCTTCTGCTCTGTATGGCG TCCCCTGTTAGACAACTACCTCCC	52.6	75
cytochrome C oxidase 1	TGTAGGAATGGATGTAGACACCCGAG CCGTGGAGAGTGGCTAATCAGC	52.4	105
cytochrome b	CACACTTCTAAACAGCGAGGAATAGC TTGTCCAATGATGATGTAGGGGTG	53.3	135
NADH dehydrogenase 2	TCTCATTGGAGGGTGAAGCGG CAATCAGAGTAAGTTGCGGAGCG	52.5	125
ATP synthase 6	TATCCTCGTTGCCATACTTCTACCTTG ATAAGTTGGTTTGTGAATCGTCCAGTC	52.1	120
cardiac troponin T2	GAGAGACGGAGTGGAAAGAAACAG GAGAGCAGATTCATTGGCATTGTC	52.2	105
troponin C	AATCCCTGCCCTCATAACGC GTGTTCATCTGTCTGTCTGCTGC	53.3	95
cardiac α-actin	CTCCATCGTCCACAGAAAGTGC AAGGCATACGGGGGGTTAGTTG	50.9	75

Table 3.2 [^{3}H]TCDD levels (ng/g) in treated embryos, as determined by liquid scintillation counting. For each treatment group (e.g. TCDD 1, TCDD 2, etc.), three sub-samples of three embryos each were counted. Measurements in pmol TCDD/embryo were converted to ng TCDD/g embryo weight assuming 1mg embryos. Treatment group means (± standard deviation), as well as the overall dose-level means (± standard deviation) are shown.

	TCDD 1	TCDD 2	TCDD 3	MEAN
0.5nM TCDD:	1.37 ± 0.14	1.99 ± 0.14	2.17 ± 0.15	1.84 ± 0.42

	TCDD A	TCDD B	TCDD C	TCDD D	MEAN
5.0nM TCDD:	12.21 ± 0.46	9.19 ± 0.31	11.54 ± 0.36	10.00 ± 0.55	10.74 ± 1.38

Table 3.3 Pairwise correlation coefficients for replicate hybridizations for 0.5nM TCDD samples (a) and 5.0nM TCDD samples (b). Correlation analysis was performed on all features that were detected and flagged "good" on all hybridizations from that dosage group; the number of features used is indicated below each table. Self-self correlations are shaded in grey.

(a)

0.5nM TCDD	*1*	*2*	*3*
1	1		
2	0.82	1	
3	0.81	0.99	1

n=3772

(b)

5.0nM TCDD	*A*	*B-1*	*C-1*	*D*	*B-2*	*C-2*
A	1					
B-1	0.18	1				
C-1	0.61	0.43	1			
D	0.25	0.49	0.35	1		
B-2	0.15	0.58	0.45	0.59	1	
C-2	0.50	0.32	0.45	0.34	0.30	1

n = 3478

Table 3.4 Clone information and mean expression ratios for positive and negative control features. Fragments of Arabidopsis thaliana chloroplast genes were arrayed and spiked into hybridization samples in equal quantities to provide external, or technical, negative controls (a). Genes used as internal, or biological, controls (b, c) were selected based on established patterns of expression following TCDD exposure [42, 51]. Unique clone identifiers and, where available, corresponding GenBank accession numbers are provided.

Gene Name	Genbank Accession #	Clone ID	Mean Fold Change (p-value) 0.5nM TCDD	5.0nM TCDD
(a) *Arabidopsis thaliana* external negative controls				
photosystem I chlorophyll a/b-binding protein (CAB)	X56062	A.th_1	1.29 (0.02)	-1.07 (0.69)
RUBISCO activase (RCA)	X14212	A.th_2	-1.61 (0.0002)	-1.29 (0.0064)
ribulose-1,5-bisphosphate carboxylase/ oxygenase large subunit (RBCL)	U91966	A.th_3	1.44 (0.08)	-1.11 (0.46)
(b) Internal negative controls				
aryl hydrocarbon nuclear translocator (ARNT)		ARNT-1	1.16 (0.56)	1.21 (0.30)
b-actin	NM_131031	beta-actin	1.38 (0.17)	-1.37 (0.05)
ubiquitin		ubiquitin	-1.15 (0.31)	-1.12 (0.57)
(c) Internal positive controls				
aryl hydrocarbon receptor 2 (AHR2)	NM_131264	AhR2-5	1.91 (9.76e-13)	1.94 (1.80e-05)
cytochrome P450 1A (CYP1A)	AB078927, AF210727	CYP1A-1	28.63 (1.13e-09)	62.85 (0.0)

Table 3.5 Microarray expression data for redundantly represented genes determined to be differentially expressed at p-value $\leq 5 \times 10^{-4}$. These genes fell into three functional classes – mitochondrial genes (a), sarcomeric proteins (b), and ESTs (c). For each gene, the overall mean expression ratio, as well as the range of values for individual clones, is shown. The number of clones representing each gene is indicated (N).

GENE	N	0.5nM TCDD Mean	0.5nM TCDD Range	5.0nM TCDD Mean	5.0nM TCDD Range
(a) mitochondrial genes					
NADH dehydrogenase, subunit 1	4	1.34	1.21 - 1.60	1.21	1.15 - 1.30
NADH dehydrogenase, subunit 2	6	1.51	1.26 - 1.63	1.26	1.13 - 1.40
NADH dehydrogenase, subunit 4/4L	7	1.54	1.36 - 1.71	1.27	1.16 - 1.30
NADH dehydrogenase, subunit 5	5	1.33	1.09 - 1.41	1.46	1.11 - 1.59
Cytochrome C oxidase, subunit 1	20	1.45	1.23 - 1.69	1.23	1.05 - 1.48
Cytochrome b	11	1.45	1.31 - 1.66	1.18	1.02 - 1.27
ATP Synthase	2	1.26	1.18 - 1.33	1.36	1.31 - 1.41
(b) sarcomeric proteins					
Troponin T2	4	1.92	1.8 - 2.4	1.65	1.3 - 1.9
Cardiac α-actin	4	-1.23	-1.1 - -1.5	-1.3	-1.3 - -1.4
Cardiac myosin heavy chain β	10	1.6	-1.6 - +2.0	1.07	-1.1 - 1.2
Cardiac myosin light chain 2	3	1.59	1.4 - 1.8	1.24	1.1 - 1.4
(c) ESTs					
TR004	19	2.19	1.1 - 3.3	2.59	1.6 – 7.7

Table 3.6 Identities and mean expression ratios of genes whose expression was altered by embryonic TCDD exposure. Unique clone identifiers and, where available, corresponding GenBank or UniGene (for ESTs) accession numbers are provided. The strength of statistical support is indicated by asterisks to the right of fold-change values (no asterisk = not statistically significant, * = p-value $\leq 5 \times 10^{-4}$, ** = p-value $\leq 1 \times 10^{-7}$, *** = p-value $\leq 1 \times 10^{-10}$). Genes have been separated into the following functional classes: (a) genes involved in xenobiotic detoxification, (b) sarcomeric proteins, (c) enzymes responsible for electron transfer and cellular energetics, (d) genes with assorted known or predicted functions, (e) ESTs with no significant similarity to known proteins (undetermined = low quality sequence), and (f) EST cluster TR004.

(a) Xenobiotic detoxicification genes

Gene Name	Genbank Accession #	Clone ID	Mean Fold Change 0.5nM TCDD		5.0nM TCDD	
aryl hydrocarbon receptor 2 (AHR2)	NM_131264	AH040775	2.23		2.11	*
		AH042846	2.43	*	2.71	***
		AhR2-5	1.91	***	1.94	*
cytochrome P450 1A (CYP1A)	AB078927, AF210727	CYP1A-1	28.63	**	62.85	***
cytochrome P450 1B1 (CYP1B1)	AF235139	CYP1B1-1	ND		4.9	*
glutathione S transferase π (GSTπ)	none	AH045159	1.46		2.01	**
major vault protein		AH046177	1.61	*	1.85	***
EST, similar to phospholipid hydro-peroxide glutathione peroxidase	Dr.24921	AH041475	-2.02	***	-1.02	

(b) Sarcomeric proteins

Gene Name or Description	GenBank Accession #	Clone ID	Mean Fold Change 0.5nM TCDD		5.0nM TCDD	
cardiac troponin T2	AF282384	AH039524	2.24	*	1.87	
		AH041025	1.8	*	1.92	*
		AH041916	1.86	***	1.82	
		AH042942	1.76	***	1.92	*
		AH042942	1.76	***	1.92	*
		AH044533	2.11	*	1.58	*
		AH046261	2.35	***	1.33	
cardiac myosin light chain 2	AF114428	AH044350	1.83	***	1.22	
myosin light chain, similar to atrial forms		AH046137	1.86	*	1.18	
ventricular myosin heavy chain	AF114427	AH042793	1.91	*	1.16	
		AH045706	1.97	**	1.15	
		AH046321	1.93	**	1.1	
		AH046397	3.30	*	1.24	
myosin heavy chain, similar to cardiac forms		AH041834	1.92	*	1.44	
myosin heavy chain, similar to skeletal slow muscle forms		AH045853	1.93	*	-1.05	

(c) Electron transfer and energy production enzymes

Gene Name	GenBank Accession #	Clone ID	Mean Fold Change 0.5nM TCDD		5.0nM TCDD	
cytochrome c oxidase subunit I	NC_002333	AH040733	1.84	*	1.10	
		AH042406	2.39	***	1.17	
		AH044438	1.80	**	1.10	
		AH044694	2.25	***	1.23	
		AH046473	2.09	***	1.42	**
NADH dehydrogenase subunit 5	NC_002333	AH044990	1.93	***	1.36	*
Ca2+ ATPase (SERCa)		AH045786	1.90	***	1.23	
EST, similar to aconitase (aconitate hydratase, citrate hydrolyase)	Dr.2353	AH046765	2.17	*	1.82	
ATP synthase	NC_002333	AH038800	-1.89	*	-1.06	
		AH039011	-1.91	**	-1.22	
EST, similar to adenylate kinase		AH039190	-1.97	*	-1.50	

(d) Assorted genes with known or predicted functions

Gene Name	GenBank Accession #	Clone ID	Mean Fold Change 0.5nM TCDD		5.0nM TCDD	
20 β-hydroxysteroid dehydrogenase		AH038811	1.16		2.37	*
pyrimidine 5' nucleotidase, cytosolic		AH039334	1.41		1.88	*
mitochondrial 16s ribosomal RNA	NC_002333	AH038898	-1.46		-1.89	*
mitochondrial 12s ribosomal RNA	NC_002333	AH038991	-2.91	*	-1.75	***
ribosomal protein S8		AH041743	3.25		2.69	*
cryptochrome 1a	NM_131789	AH046188	1.55		1.84	***
EST, similar to aquaporin 8	Dr.916	AH042420	1.89	*	-3.18	

(e) ESTs with no significant similarity to known proteins

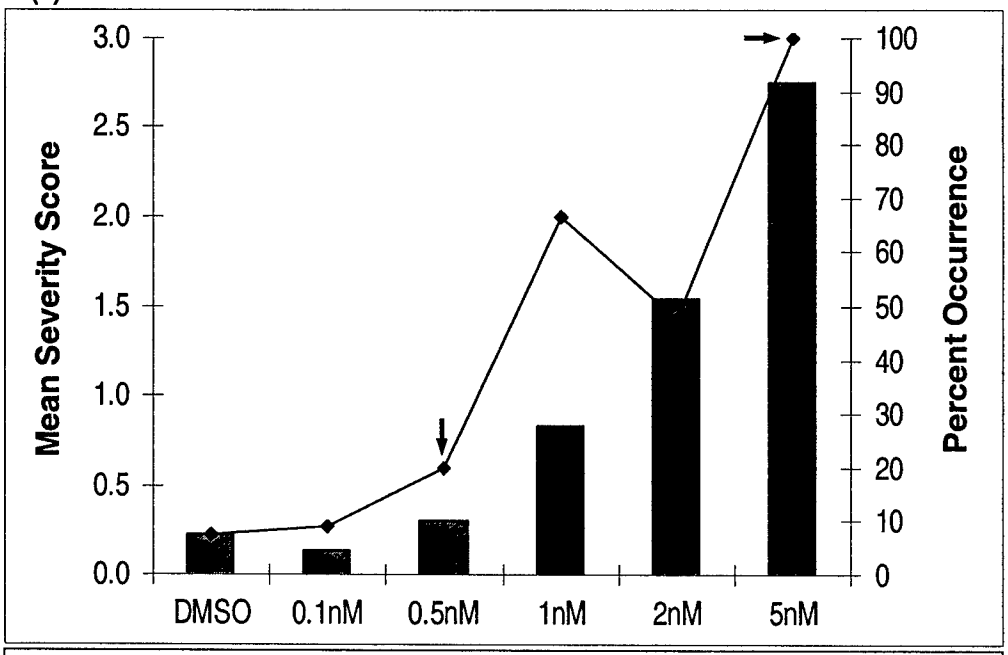
Gene Name	UniGene Accession #	Clone ID	Mean Fold Change 0.5nM TCDD		5.0nM TCDD	
EST	Dr.11104	AH045730	2.13	***	1.25	*
EST	Dr.22367	AH044506	1.83	*	1.10	
EST	Dr.23041	AH045277	1.83	**	1.21	
ESTs	Dr.24483	AH041557	2.41	***	3.93	
		AH046684	2.13	*	2.27	***
EST	Dr.25291	AH042159	-1.99	*	1.11	
EST	Dr.23558	AH045251	ND		1.92	*
EST		AH042885	1.89	*	ND	
EST		AH042622	2.16	*	ND	
EST		AH044370	8.39		1.99	*
EST		AH041068	-1.88	***	2.06	
EST		AH040900	-2.02	*	-1.14	
EST		AH038788	-2.46	*	-1.71	
EST		AH039003	-2.07	*	-1.61	
EST		AH046249	1.98	*	-1.2	
EST		AH046847	-2.41	***	-1.01	
EST		AH038815	-1.84	*	-1.04	
EST		AH042851	-1.8	*	-1.12	
undetermined		AH045830	1.86	*	1.90	*

(f) EST cluster TR004, with weak similarity to retroviral envelope proteins

	Mean Fold Change			
TR004 Clone ID	**0.5nM TCDD**		**5.0nM TCDD**	
AH044277	1.63	**	2.78	**
AH042756	1.87		2.22	***
AH042241	1.88		2.20	***
AH039458	1.89	***	2.14	*
AH044293	1.89	***	2.48	***
AH040801	2.04		3.10	***
AH043006	2.25		1.98	***
AH044418	2.44		2.83	**
AH046610	2.58	*	2.29	*
AH042961	2.60	***	2.66	**
AH046681	2.61	***	2.54	***
AH041814	2.77	***	2.20	***
AH046805	2.89	***	2.41	***
MEAN ± Std. Dev.	**2.26 ± 0.41**		**2.45 ± 0.32**	

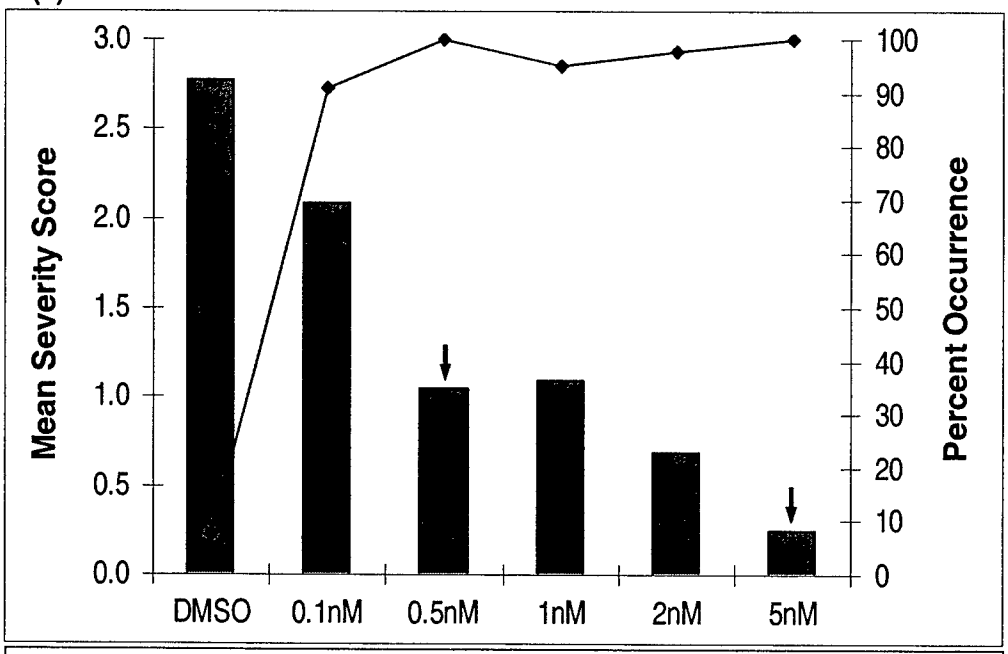
Figure 3.1 Dose-response curves for TCDD-induced pericardial edema (a) and impairment of caudal circulation (b), as observed at 80 hpf. Severity of impacts was scored on an individual basis according to a discrete ranking system, and mean severity scores were calculated from a sample size of 39-45 embryos per treatment group. Occurrence was determined as the percentage of all individuals exhibiting a given symptom at any severity level. TCDD concentrations selected for use in transcriptional profiling are indicated by black arrows.

(a) Pericardial edema



0 = No edema 1 = Mild edema 2 = Moderate edema 3 = Severe swelling

(b) Caudal circulation

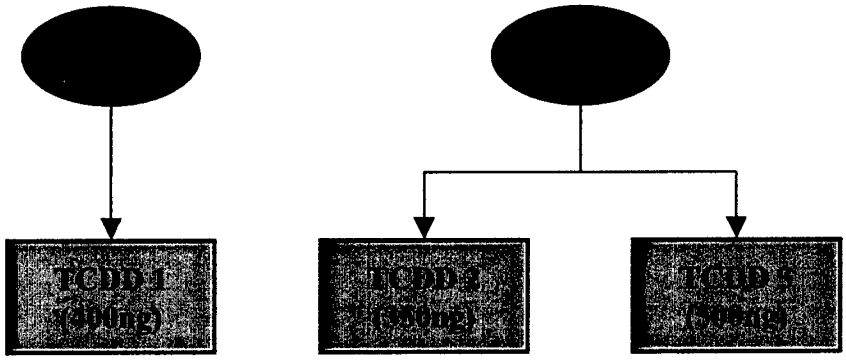


3 = Normal circulation 1 = Reduced flow rate and cell number
2 = Slight congestion in caudal vein 0 = Little or no circulating blood in tail

Mean Severity Rank —◆— Percent Occurrence

Figure 3.2 Schematic representation of experimental designs for TCDD exposure and microarray analyses. Replicate treatment groups are represented by ovals (DMSO) and rectangles (TCDD). Hybridization pairings are indicated by connecting arrows, with Cy-dye labeling shown in color (Cy5 = red, Cy3 = green). The quantity of labeled cDNA used in each hybridization is indicated in parentheses beneath the TCDD sample name.

a) 0.5nM TCDD experimental design



b) 5.0nM TCDD experimental design

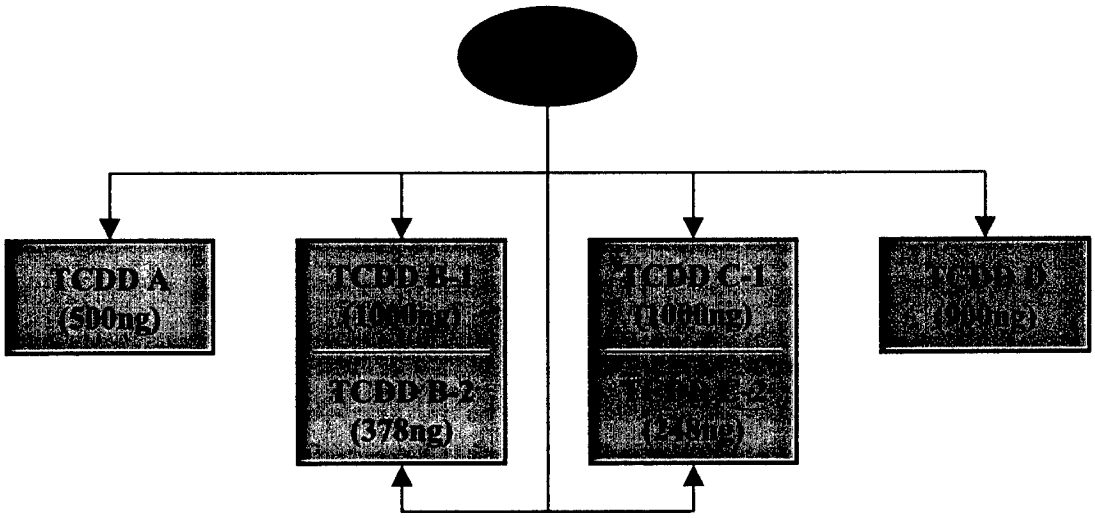


Figure 3.3 Lack of effect of cDNA quantity on microarray gene expression results, as determined by comparison of gene-specific mean expression ratios determined from three high-dose hybridizations using either ≤500ng cDNA (5.0nM TCDD A, B-2, C-2) or ~1μg cDNA (5.0nM TCDD B-1, C-1, D).

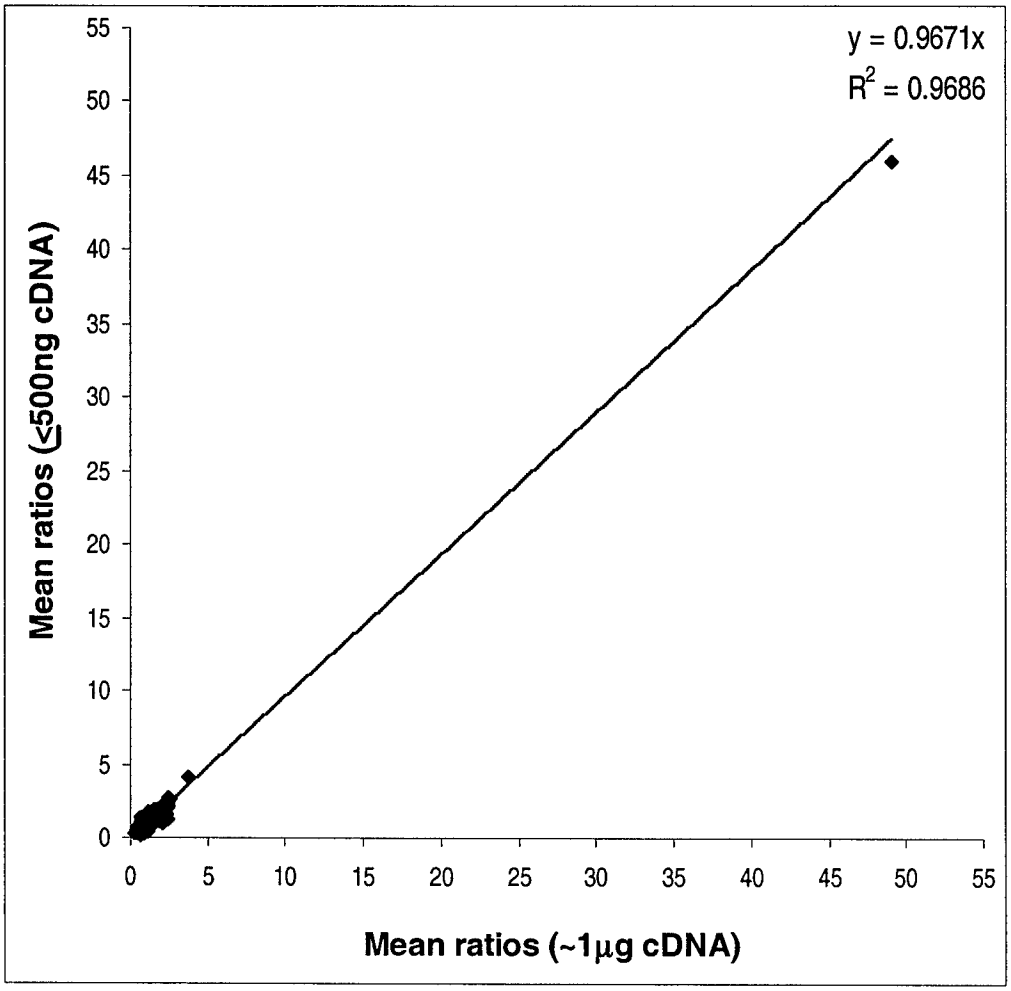

y = 0.9671x
R2 = 0.9686
Mean ratios (≤500ng cDNA)
Mean ratios (~1μg cDNA)
0
5
10
15
20
25
30
35
40
45
50
55

Figure 3.4 Inverse relationship between degree of difference in embryo TCDD levels and correlation between low-dose replicate hybridizations (red, squares) and high-dose replicates (blue, diamonds). Self-self correlation values (i.e. TCDD A vs. TCDD A) were excluded to avoid skewing results. Linear regression equations and R^2 values are shown. Data are taken from tables 3.2 and 3.3.

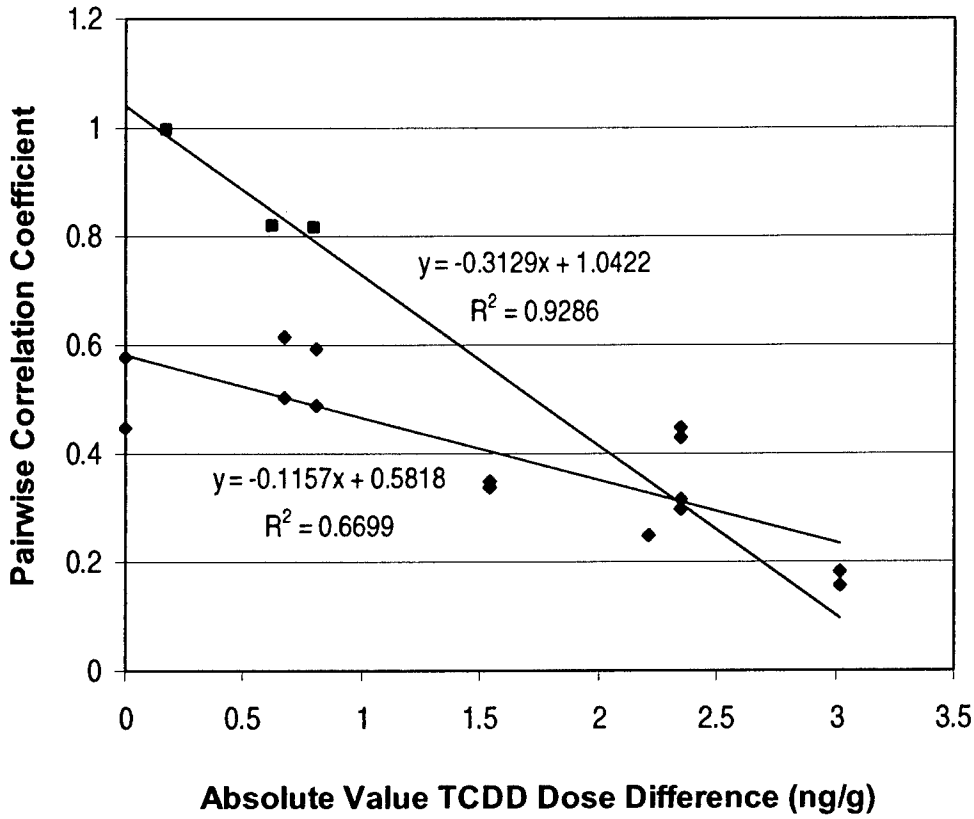

1.2
1
0.8
0.6
0.4
0.2
0
Pairwise Correlation Coefficient
y = -0.3129x + 1.0422
R2 = 0.9286
y = -0.1157x + 0.5818
R2 = 0.6699
0
0.5
1
1.5
2
2.5
3
3.5
Absolute Value TCDD Dose Difference (ng/g)

Figure 3.5 Percentage of ≥2-fold changes made up by genes with known functions (black, with selected genes listed inset), ESTs with no significant similarity to known genes (dark grey), clones whose identity was undetermined due to low quality sequence (light grey), and clones that were found to have no insert (white).

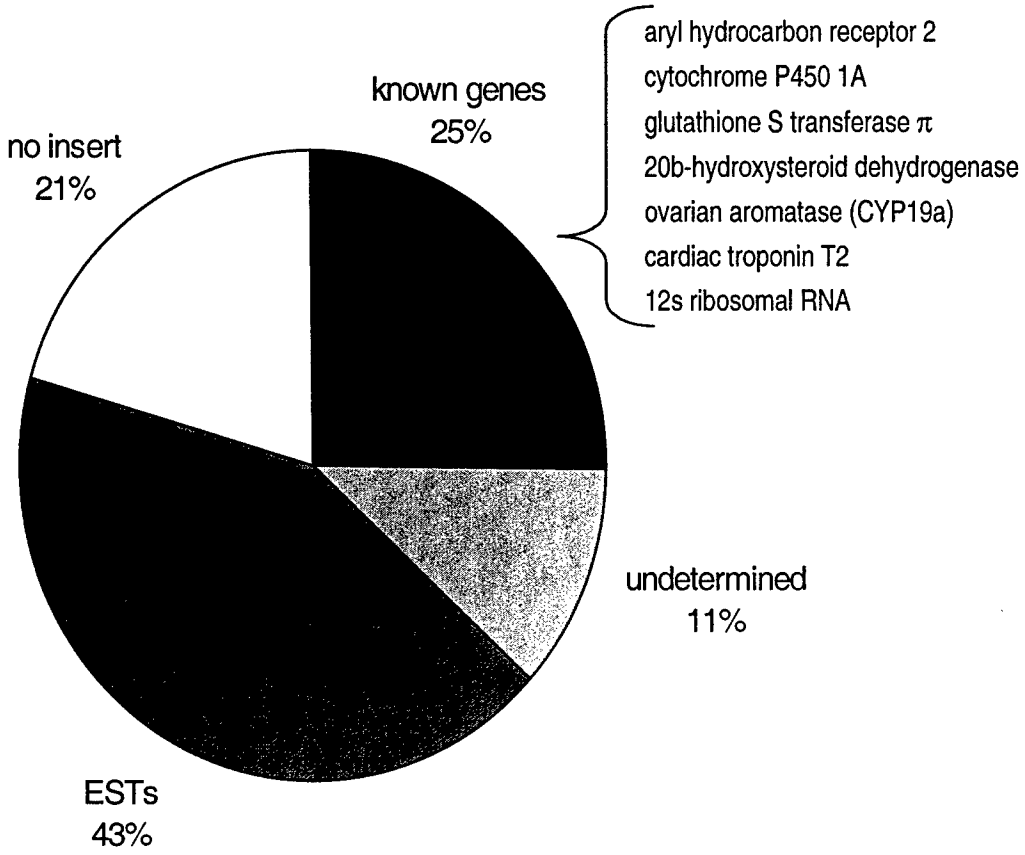

known genes
25%
no insert
21%
undetermined
11%
ESTs
43%
aryl hydrocarbon receptor 2
cytochrome P450 1A
glutathione S transferase π
20b-hydroxysteroid dehydrogenase
ovarian aromatase (CYP19a)
cardiac troponin T2
12s ribosomal RNA

Figure 3.6 Summary of gene expression results filtered according to a statistical confidence threshold of p-values $\leq 5 \times 10^{-4}$.

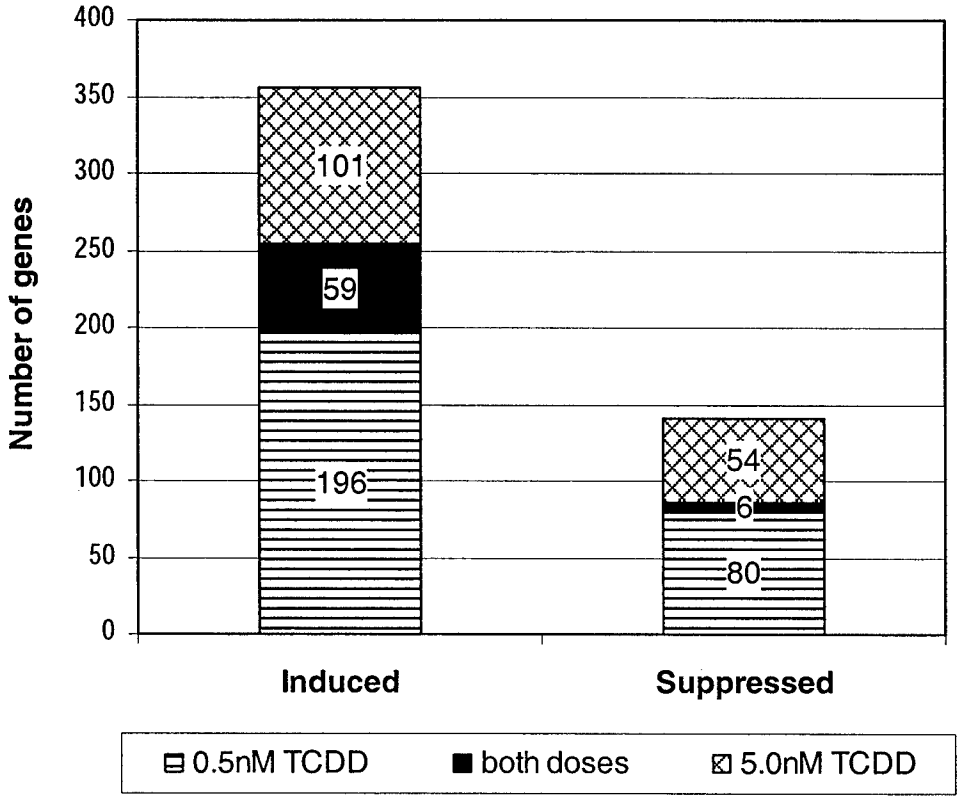

400
350
300
250
200
150
100
50
0
Number of genes
101
59
196
54
6
80
Induced
Suppressed
0.5nM TCDD
both doses
5.0nM TCDD

Figure 3.7 Frequency distribution of magnitude (absolute value fold change) of statistically significant (p-value $\leq 5 \times 10^{-4}$) mean expression ratios at 0.5nM (gray) and 5.0nM TCDD (black). Expression ratios >4.0 were limited to CYP1A and CYP1B1, and are not shown.

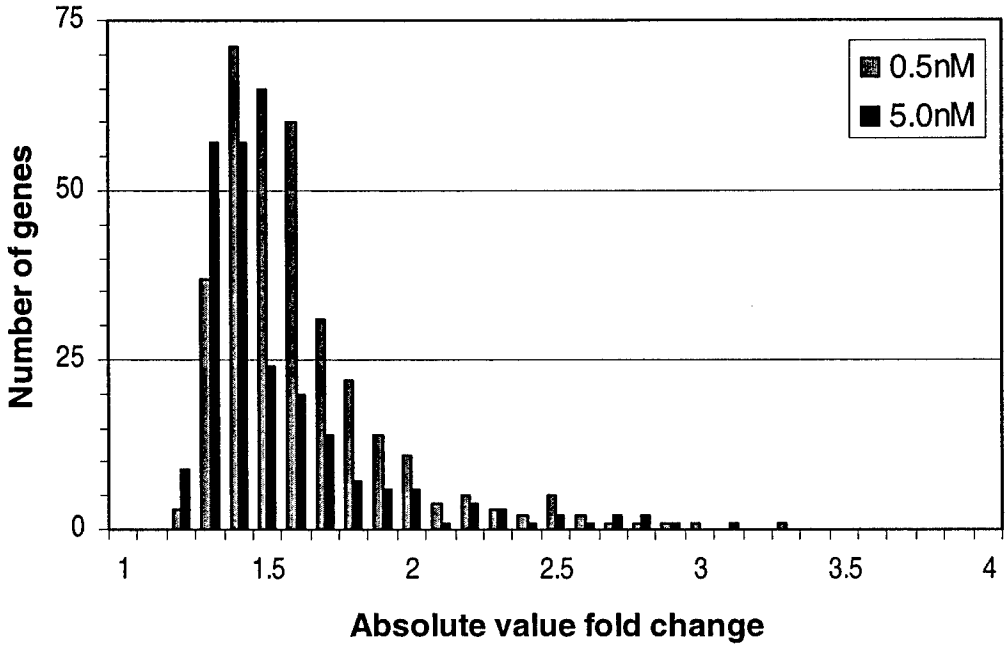

75
50
25
0
Number of genes
0.5nM
5.0nM
1
1.5
2
2.5
3
3.5
4
Absolute value fold change

Figure 3.8 Summary of TCDD-induced significantly differential expression (i.e. $\geq$1.8-fold change, p-values $\leq 5 \times 10^{-4}$), showing the excess of inductive responses and the relatively small number of genes whose expression was significantly altered at both TCDD dose levels.

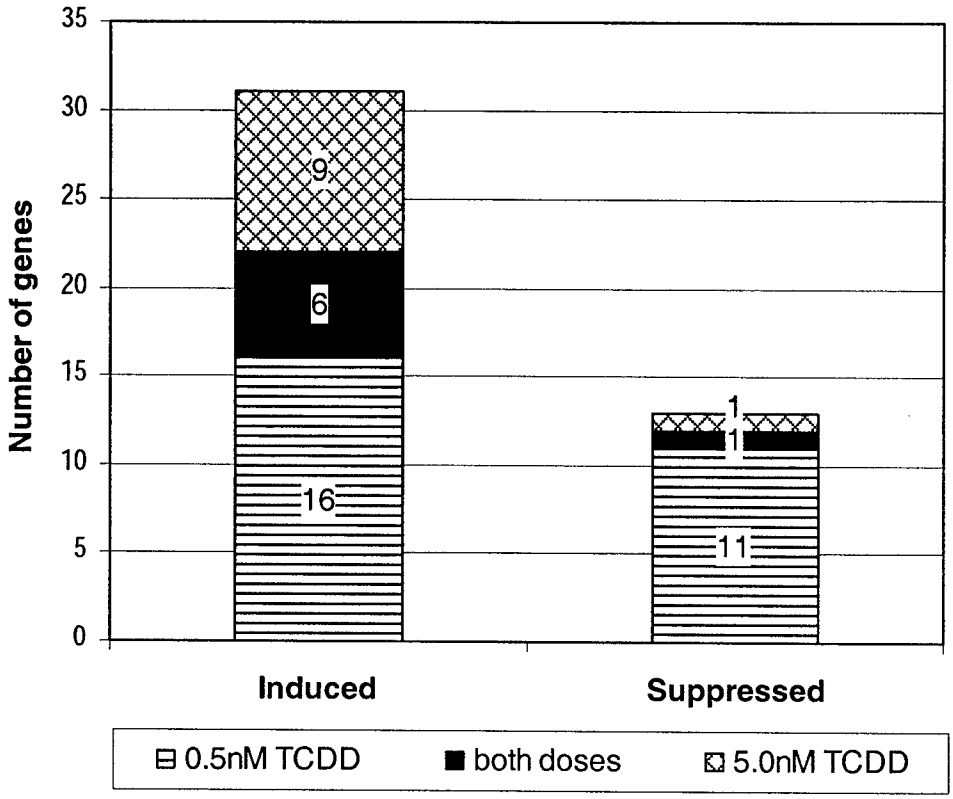
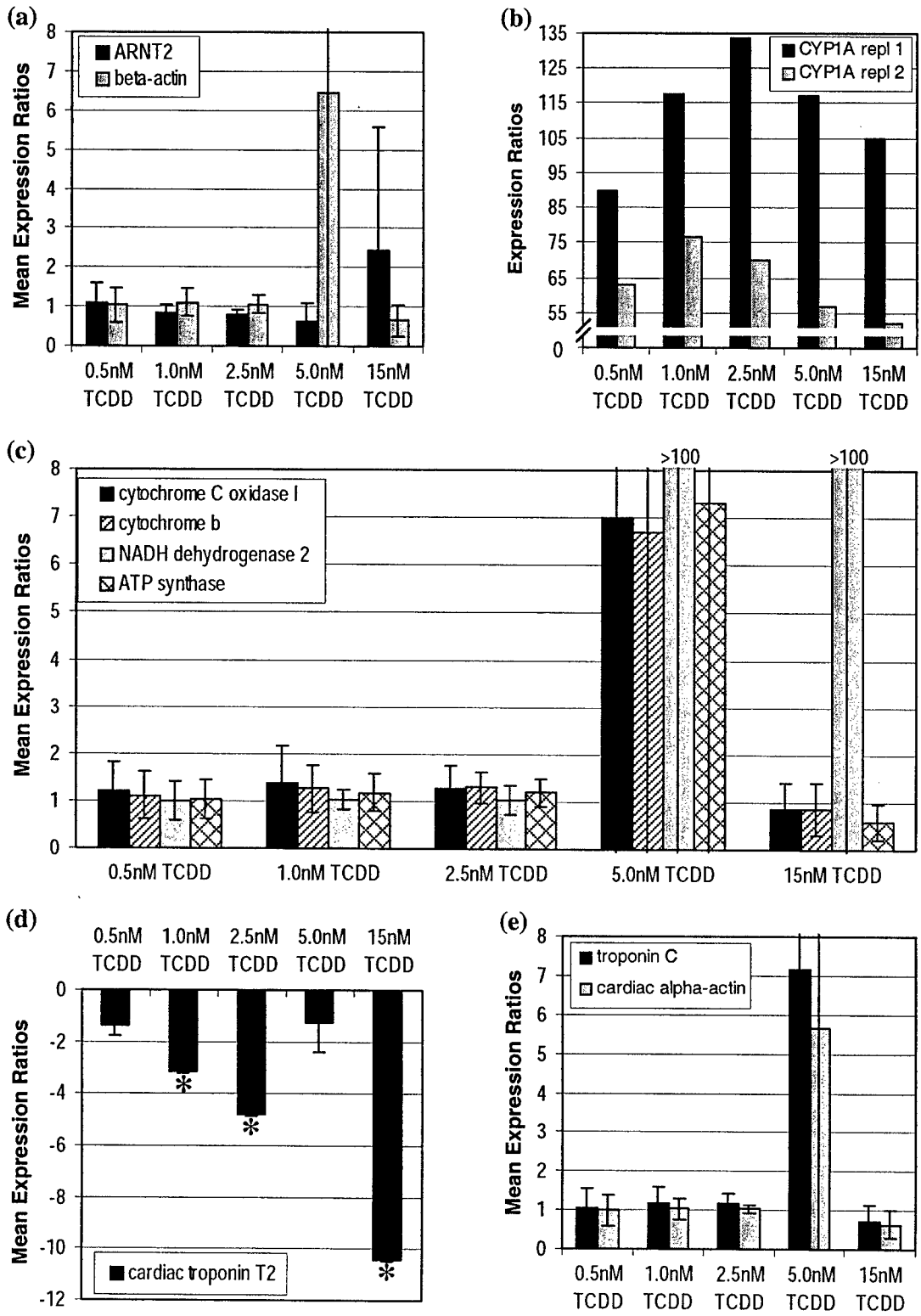

35
30
25
20
15
10
5
0
Number of genes
9
6
16
1
1
11
Induced
Suppressed
0.5nM TCDD
both doses
5.0nM TCDD

Figure 3.9 Gene expression ratios for selected genes, as determined by real-time RT-PCR. ARNT2 and β-actin provided negative controls (a), while CYP1A served as positive control (b). Subtle changes in expression of mitochondrial electron transfer proteins were not confirmed by RT-PCR (c, central panel). In contrast to microarray analyses, RT-PCR indicated strong suppression of cardiac troponin T2 (d). Relative expression levels for additional sarcomeric proteins, troponin C and cardiac α-actin, are also shown (e). Statistically significant results (single-factor ANOVA, p-values ≤ 0.01) are indicated by astrices.


(a)
Mean Expression Ratios
ARNT2
beta-actin
0.5nM TCDD
1.0nM TCDD
2.5nM TCDD
5.0nM TCDD
15nM TCDD
(b)
Expression Ratios
CYP1A repl 1
CYP1A repl 2
(c)
Mean Expression Ratios
cytochrome C oxidase I
cytochrome b
NADH dehydrogenase 2
ATP synthase
>100
>100
(d)
Mean Expression Ratios
cardiac troponin T2
(e)
Mean Expression Ratios
troponin C
cardiac alpha-actin

CHAPTER 4

EZR1: A novel unorthodox LTR-retroelement in zebrafish (*Danio rerio*)

Abstract

Retroelements make up nearly 40% of some vertebrate genomes and can influence gene expression and genome rearrangement. This chapter describes a group of novel, unorthodox, LTR-containing retroelements, EZR1 (Expressed Zebrafish Retroelement group 1), found in zebrafish. EZR1 elements consist of canonical LTRs flanking an integrase-like open reading frame and a non-coding region similar to retroviral envelope protein genes. As EZR1 sequences do not encode a reverse transcriptase, RNase H, or protease, these elements must be non-autonomous with respect to retrotransposition. Furthermore, they cannot be placed into any current LTR retroelement class.

The initial discovery of EZR1 resulted from our investigations of TCDD-altered gene expression in zebrafish embryos; EZR1 transcript levels approximately doubled following TCDD exposure (Chapter 3). AHR binding motifs are completely absent from the EZR1 LTR, indicating that observed EZR1 induction by TCDD cannot be attributed to direct AHR signaling. Alternatively, EZR1 induction may be a secondary effect of either CYP1A-mediated increases in AP-1 activity or cross-talk between AHR and GR. Given the abundance of EZR1 transcripts in the heart and reported involvement of certain endogenous retroviruses in cardiovascular disease, the relationship between EZR1 induction and cardiovascular toxicity caused by TCDD exposure warrants further investigation.

4.1 Introduction

Retroelements, such as short and long interspersed repeat elements (SINEs and LINEs), retrotransposons and endogenous retroviruses, are potentially mobile genetic elements that require an RNA intermediate for transposition. Such elements are integral components of eukaryotic genomes, composing as much as ~40% of total genomic material in mammals [172, 173]. Amplification of retroelements can have immediate impacts on gene expression (i.e. insertional mutagenesis) and local chromosomal structure, as well as lasting influence on recombination and genome rearrangement events [174]. Different classes of retroelements exhibit a variety of retrotransposition mechanisms, amplification rates, and fates in the host genome. Endogenous retroviruses and closely related retroelements comprise a distinct group defined by the presence of flanking long terminal repeats (LTRs) utilized in host genome integration.

While diverse at the level of primary sequence, LTRs possess several conserved features that can be used for *de novo* identification of LTR retroelements in genome sequences [175, 176]. Most LTRs are identical, direct repeats of approximately 300-500 bp in length, delineated by short (2-4 bp) inverted repeats, 5'-TG(TA) ... (TA)CA-3'. Additionally, integrated retroelements are flanked by direct repeats of 4-6 bp of host genomic sequence generated during the integration process. Canonical LTRs consist of three subunits – U3, R, and U5. The central R domain is delineated on the 5' end by transcription initiation sequences, and on the 3' end by a poly-adenylation signal. Since the R region is typically less than 80 bp in length, the proximal (5' LTR) poly-adenylation site is usually ignored. Many retroelements have an additional signal, a poly-adenylation downstream sequence (PADS), in the U3 region that further specifies use of the poly-adenylation site in the 3' LTR. Sequence elements responsible for regulation of gene expression are also generally found in the upstream U3 region.

The internal composition of LTR retroelements is similar to that of exogenous retroviruses, from which they are thought to have derived. As a general rule, the LTRs flank a small number of long open reading frames (ORFs), including homologs of the retroviral genes *gag*, *pol*, and *env*. *Gag* encodes structural proteins that form intracellular

nucleocapsid particles. The *pol* polyprotein typically yields an aspartic protease (Pro), a reverse transcriptase (RT), ribonuclease H (RNase H), and an integrase (Int) enzyme similar to the transposases found in DNA transposable elements. Finally, *env* encodes envelope glycoproteins necessary for host cell invasion by infectious retroviruses. Endogenous retroviruses tend to maintain all three genes, although often with inactivating mutations, while other retroelements generally lack an *env* ORF.

Traditionally, LTR retroelements have been divided into five major groups – BEL, Ty1/*copia*, Ty3/*gypsy*, DIRS1, and the vertebrate endogenous retroviruses. This classification system is based primarily on domain order within the *pol* ORF and phylogenetic analyses of reverse transcriptase genes. The DIRS1 group presents a challenge to this scheme; the defining features of DIRS1 retroelements are integrase genes distinct from either retroviral integrases or DNA transposon endonucleases, and unorthodox termini consisting of either split direct repeats or non-identical, inverted repeats [177, 178]. The recently discovered zebrafish retroelement, *bhikhari* (*bik*), also defies traditional classification [179]. *Bik* is flanked by extensive direct repeats manifesting all conserved features of LTRs, but it contains only a single ORF with no homology to any known protein.

Gene expression profiling of TCDD-exposed zebrafish embryos identified an EST cluster with weak homology to retroviral envelope proteins (Chapter 3). Further characterization of these ESTs, described herein, has revealed a novel, unorthodox LTR retroelement, EZR1 (Expressed Zebrafish Retroelement group 1). Examination of multiple EZR1 transcripts and genomic copies indicates that EZR1 elements consist of canonical LTRs flanking a single integrase-like ORF and a non-coding region similar to retroviral *env* genes. EZR1 elements lack a recognizable RT domain, and thus, cannot be placed within any existing LTR retroelement classes. EZR1 transcripts are abundant in normal embryonic and adult tissues, particularly the heart, and retroelement expression is enhanced by the environmental pollutant 2,3,7,8-tetrachlorodibenzo-*p*-dioxin. Predicted transcription factor binding sites provide hypotheses regarding the regulation of EZR1 expression. However, the biological function of EZR1, if any, remains uncertain.

4.2 Methods

Zebrafish Embryos

All fish used in these experiments were from an inbred line of wild-type TL zebrafish (*Danio rerio*) maintained in the Fishman laboratory facility at Massachusetts General Hospital. To obtain embryos, trios of one male and two female mature fish were held in divided mating tanks overnight. To ensure that all embryos were synchronous to within two cell cycles, embryos were collected within 30 min after removing the barrier the following morning. Embryos were maintained in Tübingen E3 egg water (5mM NaCl, 0.17mM KCl, 0.33mM $CaCl_2$, 0.33mM $MgSO_4$) at 28°C.

At 48 and 72 hours post fertilization (hpf), groups of 100 embryos were anaesthetized on ice and placed in 1ml 4% paraformaldehyde for 2 hrs, or overnight at 4°C. Fixed embryos were rinsed three times with 1ml methanol, and subsequently stored in 1ml methanol at -20°C. Additional embryos (72 hpf) were flash frozen in liquid nitrogen and stored at -80°C for subsequent RNA isolation.

in situ Hybridization

High resolution *in situ* hybridization was performed essentially as per previous description [180]. Briefly, digoxygenin-labeled anti-sense RNA probes were prepared by *in vitro* transcription from 1μg plasmid DNA using SP6 RNA polymerase. Fixed embryos were gradually rehydrated, then incubated with anti-sense probes, followed by primary and secondary antibodies, and finally, by chemiluminescent detection reagents. Embryos were then post-fixed in 4% PFA, photographed, and stored in phosphate buffered saline

RNA Preparation

Total RNA was prepared using TriZol Reagent (Invitrogen) according to manufacturer's protocol. Briefly, 50-100 embryos were homogenized in 1ml TriZol. Phase separation was accomplished by addition of chloroform (0.2 ml). RNA was precipitated from the aqueous phase using isopropanol (0.5 ml), and the resulting pellets

were washed with 1 ml 75% ethanol. Air-dried RNA was reconstituted in DEPC-treated water and stored at -80°C.

5' RACE and RT-PCR

cDNA was generated from total RNA using the SMART RACE kit (BD Biosciences Clontech) according to standard protocols. Briefly, total RNA (2 μg), 5'-RACE CDS primer (20 pmol), and SMART II A oligo (20 pmol) were incubated together 2 min at 70°C, then chilled on ice. Reverse transcription reactions (20 μl final volume) containing 1x first strand synthesis buffer, 2mM dithiothrietol, 1mM dNTPs, and Powerscript reverse transcriptase (2 μl) were then incubated 2 hrs at 42°C. cDNA was diluted 1:2.5 in dH_20 and stored at -20°C.

5' RACE PCRs were performed using Advantage 2 Polymerase (1 μl) in 50 μl reactions containing 1x reaction buffer, dNTPs (200μM each), and one gene-specific primer (20nM) plus either 1x Universal Primer Mix A or Nested Universal Primer A (1μM) (BD Biosciences Clontech). For all other PCR, Taq polymerase (Epicentre) was used with two equimolar gene-specific primers (20nM) under otherwise identical reaction conditions. The amplification program consisted of an initial 1 min denaturation step (95°C), followed by 30 cycles of 10 sec at 94°C, 10 sec at 58 °C, and 3 min at 72 °C. Completed PCRs were held at 4 °C prior to visualization by agarose gel electrophoresis and subsequent purification. Gene-specific PCR primer sequences (5' to 3') were as follows:

F1 – CCATGCAACCAGGATAAAACGAGC
R1 – GCCTGACAACACAGGATGGACAGG
R2 – CAGTCCCAATGTCCATAGCCACTTC
R3 – AGGTGCTCGTTTTATCCTGGTTGCATGG
R4 – GTTCTGGTTACAGCCACGACATCCGTCC

DNA Sequencing

Contaminating dNTPs and enzymes were removed from aliquots of PCR reactions (30μl) using QiaQuick PCR Purification spin columns according to standard protocols

(Qiagen). Purified PCR products (3μl) were ligated into pGEM-T Easy vector (1μl) in 10μl reactions containing 1x rapid ligation buffer and 1μl T4 DNA ligase (Promega). Ligations were incubated 2hrs at room temperature, then stored at 4°C prior to transformation into JM109 cells by standard heat-shock protocol (Promega). Transformed cells were grown overnight at 37°C on LB agar plates containing ampicillin (100μg/ml), IPTG and X-gal.

Selected colonies were transferred to liquid LB (100μg/ml ampicillin) and grown overnight at 37°C with agitation (220rpm). Bacterial cells from 4mls of liquid cultures were pelleted, and plasmid DNA was prepared using QiaPrep Spin Mini-Prep columns according to manufacturer's protocol (Qiagen). Universal SP6 and T7 primers (40pmol each) were used to amplify insert fragments from 1μl mini-prep'd DNAs in 100ul PCRs containing 1x PCR buffer, 2.0mM $MgCl_2$, 400μM dNTPs, and 5U Taq DNA polymerase (Qiagen). PCR products were purified using QiaQuick PCR Purification spin columns (Qiagen), and DNA concentrations were adjusted to 20ng/100 bp length. DNA sequencing reactions were performed by the Massachusetts General Hospital DNA sequencing core facility.

Radiation Hybrid Mapping

Mapping PCRs were performed using 5μl genomic DNA from the Goodfellow T51 radiation hybrid panel [181] in 10μl reactions containing 1x PCR buffer (Qiagen), 2.0mM $MgCl_2$, 200μM dNTPs, and 2μM each primer (F1 and R2, sequences above). An initial 30 sec denaturation step at 95°C was followed by 35 amplification cycles (30sec at 94°C, 30sec at 52°C, 1min at 72°C) and a final extension period of 7 min at 72°C. PCR products were stored temporarily at 4°C, then visualized by agarose gel electrophoresis and scored as present (1), absent (0), or ambiguous (2).

Sequence Analysis

Nucleotide sequence alignments were produced using ClustalX, with default settings modified to include gap opening penalty = 25 and gap extension penalty = 2. Paup 4.0

was used to perform a maximum likelihood analysis (HKY85 model assumed, stepwise addition of taxa). Bootstrap confidence values were calculated from 100 branch-swapping (tree-bisection-reconnection) replicates. This result was compared to the most probable tree morphology found by 100,000 generations (trees sampled every 100 generations) of an incrementally heated Metropolis-coupled Monte Carlo Markov Chain analysis (MrBayes v3.0, 4-by-4 nucleotide substitution model with rate variation according to determined gamma distribution).

Transcription factor binding motifs were identified by using MatInspector v2.2 to search the TransFac 4.0 database [182]. All other *in silico* sequence analysis was performed in GCG/SeqLab (Wisconsin Program Package). Zebrafish genomic sequence data used herein were produced by the Zebrafish Sequencing Group at the Sanger Institute and can be obtained freely from ftp://ftp.sanger.ac.uk/pub/zebrafish/.

4.3 Results

Characterization of TR004 transcripts

Teratogenic doses of TCDD had been found to cause robust induction of several closely related ESTs, collectively referred to as TR004 (Chapter 3). Specifically, the TR004 cluster consisted of 19 non-identical clones, ranging in length from <250 bp to nearly 2 kb. These transcript fragments were aligned to form a single assembly 1915 bp in length, terminating at a common poly-A tail. Despite obvious sequence similarity, this alignment revealed a large number of single nucleotide mismatches and ambiguities, as well as a region of ~500 bp near the 3' end that contained several sites of distinct sequence motifs and significant deletions. Thus, I undertook to better characterize the sequence and identity of the TR004 transcripts.

RT-PCR primers were designed against the initial 100 bp, represented by only a single clone, and the highly conserved final 185 bp of the assembled sequences. These primers, dubbed F1 and R1, amplified a unique band of ~1.7 kb (expected size of 1,685 bp) from RNA from 72hpf wildtype TL zebrafish embryos (data not shown). The F1-R1

fragment was cloned into the pGEM-T Easy vector, and four individual clones were sequenced using vector-targeted primers.

In addition, 5' RACE was used to obtain further upstream sequence. Nested primers, R3 and R4, were designed against the 5'-most 100 bp of the assembly; the upstream, or internal, primer (R4) was the reverse complement of the aforementioned primer F1. Used individually, each of these primers amplified a unique fragment of ~2.5 kb in length. Serial nested PCRs resulted in enhanced amplification of (presumably the same) ~2.5 kb fragment. As before, this fragment was cloned and two individual clones were sequenced to completion. Both clones contained inserts of 2515 bp in length. The 3' ends of these fragments overlapped the 5' end of the original TR004 assembly by 88 bp, and the F1-R1 PCR products by the 28 bp representing the common F1/R4 priming site.

All above sequences were assembled into a single contig of 4312 bp (Figure 4.1, top). No two clones in this assembly were identical, and 9.8% of the consensus sequence consisted of ambiguities. The primary sources of variation were two extended regions with >50% polymorphism rate and significant insertions/deletions in some subset of clones (Figure 4.2). Outside these variable regions, the mismatch rate was approximately 1/39 nucleotides, more than twice the polymorphism rate expected based on allelic variation. The 5' half of the assembly appeared to be more highly conserved than the 3' half, likely due to lesser sequence coverage (i.e. 2 clones versus 23).

The consensus sequence of the complete assembly was found to contain a single open reading frame spanning positions 1136 bp to 2518 bp, with a conserved poly-adenylation signal (AATAAA) at position 2521 bp (see schematic in Figure 4.1, bottom). The 462aa protein sequence putatively encoded by this open reading frame demonstrated strong homology to retroviral *pol* polyproteins. In particular, a complete integrase core catalytic domain was identified by probing NCBI's Conserved Domain Database. Directed searching revealed no further homology to reverse transcriptase or other typical *pol* constituents. To reflect this fact, this open reading frame was designated *int*.

As had been previously observed, the region downstream of the *int* ORF manifested moderate (e-values $\leq 4 \times 10^{-4}$) amino acid similarity to retroviral *env* gene products.

However, no open reading frame was detected in this region when either the TR004 consensus sequence or individual clones were interrogated. No similarity to known *gag* genes was detected. Nonetheless, based on the similarity, in both sequence and gene order, of the current *int* and *env* regions to retroviral genes, the described transcripts were collectively renamed EZR1, for Expressed Zebrafish Retroelement group 1.

Putative EZR1 LTR structure

The 5' and 3' termini of the EZR1 sequence assembly were found to constitute a 27 bp identical direct repeat, possibly indicative of the presence of long terminal repeats (Figure 4.1, bottom). To confirm the presence of LTRs, I identified zebrafish genomic sequences with regions of ≥98% identity to the final 185 bp of the EZR1 transcripts, presumed to be LTR sequence. Dot-plot comparison of the complete EZR1 consensus sequence to genomic sequence contig ctg9483 revealed two direct repeats of a sequence composed of ~500 bp similar to the 3' end of the EZR1 transcripts, followed by a region of similarity to the first ~50 bp of EZR1 transcripts (Figure 4.3).

To further characterize this putative LTR, additional genomic sequences were aligned with an artificial sequence construct consisting of the initial 100 bp of the TR004 consensus sequence appended (via the observed 27 bp overlap) to the final 700 bp. The 3' end of the LTR was defined as the position at which conservation between genomic sequences and either the putative LTR construct or other genomic contigs ended. This boundary fell 67 bp downstream of the 5' terminus EZR1 transcripts, and was marked by the canonical tetranucleotide sequence TACA.

A definitive 5' boundary could not be determined, as the only genomic copy containing a putative 5' LTR manifested distinctly different internal sequence and LTR structure (possibly as a result of sequence misassembly). Thus, the 5'-most tetranucleotide TGTA (inverted repeat of the final tetranucleotide) within the region conserved among all genomic and cDNA clones was used as a proxy for the 5' boundary. This designation was supported by the presence of an 11 bp poly-purine tract immediately upstream of this location. Assuming these boundaries produced a putative

LTR of ~630 bp in length (alignment length of 633 bp), only slightly longer than the usual 300-500 bp.

I used sequence motif searching to delineate the U3, R, and U5 domains within this putative LTR. The 5' boundary of the R domain was defined by a consensus transcription initiation sequence (TACG) at LTR consensus position 601 bp, directly abutting the observed 5' mRNA terminus at 605 bp (Figure 4.4). A canonical TATA box (overall 96% identity to 15 bp motif matrix) was found 28 bp upstream of the putative initiation site, lending further support to this prediction. A consensus poly-adenylation signal was detected at consensus positions 613-618, 15 bp upstream of observed poly-A tails (Figure 4.4). In addition, a strong GT-rich retroviral polyadenylation downstream sequence (PADS) element (matrix similarity score of 91.2%) was identified in the putative U3 region at positions 343-357. Thus, all data supported the existence of a typical, if short (12 bp), central R domain, flanked by U3 and U5 regions of 600 bp and 21 bp, respectively. This further confirmed that the observed transcripts conformed to the expected domain structure R-U5-internal-U3-R.

Genomic distribution of EZR1

The region of greatest sequence variability between individual transcripts was observed to fall largely within the putative LTR, strongly suggesting multiple EZR1 copies. LTR positions 316 bp to 606 bp were used for phylogenetic analyses, as sequence information for this region was available for all cDNA clones (excepting RACE products). Both the Bayesian most probable tree and the maximum likelihood 50% consensus tree placed the majority of EZR1 transcripts into two primary clades (Figure 4.5). However, resolution within clades was extremely poor and there were several discrepancies between results generated by the two phylogenetic methods. Together with the observed high polymorphism rate and the fact that no two transcripts were identical, these data suggested the presence of at least 7, and likely many more, distinct EZR1 loci in the zebrafish genome.

PCR primers located in the *int* ORF region amplified a band of expected size from nearly all of the 95 genomic DNA templates in the T51 radiation hybrid panel (data not

shown). This result suggested that most linkage groups carry at least one EZR1 copy. Not surprisingly, a nucleotide BLAST search of the zebrafish genome assembly (version 2.0, pre-released April 3, 2003) with the full-length (4.35 kb) EZR1 consensus sequence resulted in well over 100 significant matches. However, many matches were incomplete (the final ~200 bp of the EZR1 transcript was often missing) or included significant insertions/deletions or rearrangements.

Profile hidden Markov models (HMMs) are statistical models of multiple sequence alignments that provide greater sensitivity and discrimination in homology searching than does a traditional BLAST search. In an attempt to better estimate the EZR1 copy number in the zebrafish genome, a profile HMM was built from the aligned putative LTR sequences and used to search the zebrafish genome assembly. Nearly exact copies of the complete EZR1 LTR were found in 25 genomic sequence contigs (e-values $< 5x10^{-120}$) . An additional 20 sequence contigs contained related, often incomplete, sequences (e-value range $1.7x10^{-96} - 2.2x10^{-7}$).

Cardiac expression of EZR1

EZR1 transcripts were found to be extremely abundant in cDNA libraries from wildtype adult heart tissue, regardless of genetic strain or originating facility. EZR1 clones comprised approximately 0.4% (19 of 4,896) of adult heart cDNAs randomly selected for use in constructing the cDNA microarrays with which EZR1 was identified. This level of representation was comparable to that of structural genes, such as cardiac myosin, and mitochondrial energy production enzymes (Chapter 3). Similarly, a single UniGene cluster composed of ESTs >90% identical to the EZR1 *int* region accounted for 1.6% and 2.2% of sequences in two independent adult heart cDNA libraries.

EZR1-like ESTs were also detected, albeit in lesser quantities, in cDNA libraries from a variety of tissues and developmental stages. BLAST searching the zebrafish EST database revealed over 100 matches with e-value = 0.0. These ESTs were drawn from tissues including heart, brain, liver, kidney, ovary and testis, fin, and whole embryos ranging from shield stage (6hpf) to 5 days post fertilization. *In situ* hybridization using

three distinct EZR1 clones as probes further confirmed a broad expression pattern in 48 and 72hpf embryos (Figure 4.6).

Putative LTR regulatory elements

I identified potential transcription factor binding sites in the EZR1 LTR sequences by searching the TransFac 4.0 database of known binding motifs [182]. Matches were only accepted if all LTR sequences contained an absolutely conserved core sequence within a motif ≥85% identical to the complete corresponding weight matrix. As expected, most predicted transcription factor binding sites were located in the putative U3 region; 6 matches were discarded based on coincidence with either the TATA box or transcription initiation site, and no predicted sites fell within the U5 region. In all, 18 binding sites for 14 individual transcription factors were predicted (Table 4.1, Figure 4.4).

Binding sites for hematopoietic transcription factors accounted for a significant fraction of all matches. Of particular note, two strong (>94% matrix similarity) GATA-1 binding sites were predicted. Each of these sites was also identified as matching recognition sites for LMO-2/GATA-1 complexes. However, similarity was restricted to the GATA-1 half of the motif, with no corresponding similarity to LMO-2 specific sequences. The predicted BRN-2 binding site overlapped the GATA-1 site at position 483 bp over two thirds of its length; the significance (if any) of this finding is unknown. Similarly, the region beginning at 278 bp might be predicted to interact with either Ikaros-2 or MZF-1, as the predicted Ik-2 binding sequence is entirely encompassed by the MZF-1 motif. Finally, a well-conserved (92.5% matrix similarity) serum response factor recognition site was identified at position 534 bp.

Binding motifs for activator proteins were also conspicuous. Two conserved binding sites were predicted for the AP-1 (c-Fos/c-Jun) complex, and two more for AP-4. Several additional AP-1 binding motifs were found in some subset of EZR1 LTR sequences. The AP-1 site at 526 bp overlapped (in reverse orientation) an extremely well conserved TCF-11 binding sequence (98.1% matrix similarity).

The remaining motifs did not fall into any clear functional group. These included a glucocorticoid response element, and binding motifs for the cardiac homeobox factor Nkx-2.5 and the sex determining region Y protein.

No AHR/ARNT binding sites, or dioxin response elements (DREs), were found in the putative LTR sequences. One consensus DRE was found in the internal EZR1 sequence, 135 bp into the *int* ORF. Likewise, binding motifs for NFκB were absent from EZR1 sequences.

4.4 Discussion

A novel class of non-autonomous LTR retroelements

All available data indicate that EZR1 is an LTR-class retroelement. EZR1 transcribed sequences begin and end with direct repeat sequences. These ends can be recombined to generate a putative LTR sequence that is nearly identical to several zebrafish genomic DNA regions, and manifests all features of canonical LTRs (e.g. flanking inverted tetranucleotide repeats, a central R domain with strong initiation and poly-adenylation signals, and a U3 region rich in putative regulatory elements). EZR1 also contains other retroelement-specific sequence motifs, including a retroviral PADS element in the U3 region and a poly-purine tract immediately upstream of the 3' LTR.

The internal sequence and structure of EZR1 also bears relationship to retroviruses and derived retroelements. The single EZR1 open reading frame appears to encode a retroviral-type integrase protein; the presence of a conserved catalytic domain suggests the *possibility* of an active enzyme, but this has not been confirmed. The region downstream of the *int* ORF is devoid of significant open reading frames, but (if translated) demonstrates significant similarity to retroviral *env* proteins. This suggests the presence, at some point in the past, of an *env* gene that has been degraded by subsequent mutation. In vertebrate retroviruses, as well as Ty3/*gypsy* and BEL retrotransposons, integrase is the final domain of the *pol* gene. *Env* genes are generally absent from LTR retroelements, but are found downstream of the *pol* gene in

retroviruses. Thus, the *int-env* organization of EZR1 is strongly reminiscent of the 3' half of retroviral genomes (Figure 4.7).

What is truly unusual about EZR1 is the absence of elements from the 5' half of retroviral genomes. There is no evidence in any EZR1 sequence of a *gag* gene, past or present, or of the reverse transcriptase, RNase H or protease domains of *pol*. As the current scheme of LTR retroelement classification is based primarily on aspects of the *pol* gene, EZR1 cannot be fit into any existing LTR retroelement class. More importantly, without *gag* and *pol*, EZR1 lacks the means to generate reverse transcription machinery necessary for autonomous replication and retrotransposition. In this regard, EZR1 is similar to the recently discovered zebrafish retroelement, *bhikhari* [179]. *Bik* appears to be even more remotely related to other retroelements, as it contains a single ORF encoding a protein with no significant similarity to any known proteins (Figure 4.7).

EZR1 and *bik* are distinct from retroelement pseudogenes, the only other known non-autonomous LTR retroelements. Most pseudogenes differ from active relatives by single nucleotide mutations or modest deletions or rearrangements. Furthermore, retroelement pseudogenes are generally not replicated, and thus, are found at very low copy numbers [174]. In contrast, EZR1 and *bik* appear to represent independent, replicating lineages. Their internal gene content is vastly different from any autonomous retroelements, and both are represented by 25-100 copies per genome [179], suggesting amplification in the absence of autonomous retrotransposition capability. Thus, EZR1 and *bik* seem to constitute a novel class of non-autonomous LTR retroelements.

Whether these elements are currently active (i.e. transposing) is unknown. Both EZR1 and *bik* are abundantly expressed in normal zebrafish embryos and adult tissues. Indeed, the level of EZR1 expression is comparable to that of the yeast Ty1 retrotransposon family, which contains both autonomously and non-autonomously active elements [183]. However, expressed autonomous retroelements, upon which EZR1 and *bik* would depend for retrotransposition machinery, have yet to be found in zebrafish. Likewise, according to the timescale devised by Goncalves and colleagues [184], the range of sequence variation among EZR1 transcripts supports a history of sporadic EZR1

retrotransposition over a period of several tens of millions of years. Unfortunately, given the current state of misassembly of the zebrafish genome sequence, it is difficult to distinguish whether closely related EZR1 transcripts reflect recent transposition events or allelic variation on integrated loci. However, it seems likely that retrotransposition by EZR1 (or *bik*) is an infrequent event.

LTR regulation of EZR1 expression

Although no full-length cDNA has been examined, all evidence supports the conclusion that complete 4.35 kb retroelement transcripts are expressed under control of the EZR1 LTR. Experimental data and *in silico* predictions were in absolute agreement regarding the site of transcriptional initiation. Similarly, all EZR1 ESTs terminated at a single poly-adenylation site that was strongly supported by the presence of a retroviral PADS element upstream of the conserved poly-adenylation signal. Thus, all EZR1 elements appear to conform to the canonical domain structure R-U5-internal-U3-R. Furthermore, numerous non-identical transcripts, presumably originating from multiple loci, exhibited similar patterns of expression. These similarities argue that EZR1 expression is driven by common LTR sequences, rather than regulatory elements in a specific genomic context.

Putative regulatory elements might account for many aspects of the observed expression patterns. EZR1 is expressed at low or moderate levels throughout zebrafish embryos and adult fish. While the transcription initiation site and upstream TATA box both conform to expectations for a strong initiation site, no CCAAT box was found. Still, moderate basal expression might be accomplished, even without additional enhancers. TCF-11 is a ubiquitous transcriptional enhancer that might also contribute to general expression via the strong binding site predicted in the EZR1 U3 region.

The extraordinarily high levels of cardiac-specific expression inferred from EZR1-like EST abundance in adult heart cDNA libraries is more difficult to account for. One moderately conserved Nkx-2.5 binding site was identified. While Nkx-2.5 is a cardiac-specific homeobox transcription factor of known importance in heart development and

function in zebrafish [185, 186], it seems unlikely that Nkx-2.5 activity at a single site could drive EZR1 expression to levels of 0.5-2.5% of all cardiac transcripts.

Based on the number of hematopoietic regulatory elements detected, it seems likely that EZR1 is expressed in blood. Whether EZR1 expression would be restricted to certain blood cell-types is uncertain, as GATA-1, MZF-1 and Ikaros-2 are active primarily in erythroid, myeloid and lymphoid lineages, respectively. While it is possible that blood trapped in dissected hearts contributed to cardiac cDNA libraries, blood-specific expression is not likely to account for observed EZR1 EST quantities. Certainly, no other blood-specific genes are found at remotely comparable levels in these libraries. Thus, the origin of high-level cardiac expression of EZR1 remains elusive.

EZR1 induction by TCDD

The initial discovery of EZR1 was based on its transcriptional induction by 2,3,7,8 tetrachlorodibenzo-*p*-dioxin (TCDD), a widespread and persistent environmental contaminant with potent teratogenic properties. Finding a possible mechanism for this induction was a major goal of the current work. The predicted regulatory elements suggest unexpected mechanisms for induction of EZR1 by TCDD. Most toxic effects of TCDD are mediated by the aryl hydrocarbon receptor (AHR) [22, 53-59]. Furthermore, AHR and NFκB binding sites in the LTR of HIV are known to be necessary for potentiation of HIV infectivity observed in Hepa-1 cells following exposure to TCDD [187]. Thus, the absence of binding motifs for either of these factors was surprising.

Instead, the current results suggest that glucocorticoid receptor (GR) and/or the AP-1 complex might be responsible for induction of EZR1 expression by TCDD. GR has been implicated in TCDD-responsiveness of sequences placed under the control of the murine mammary tumor virus LTR [188]. Cross-talk between AHR and GR signaling has also been observed in other systems [189]. Thus, the glucocorticoid response element in the EZR1 LTR could be a target for indirect effects of TCDD.

The nature of the effect of TCDD on AP-1 has not been fully resolved, but TCDD has been shown to enhance both expression of constituent proteins and AP-1 DNA binding activity under some conditions [152, 190-194]. Thus, the multiple potential AP-1

binding sites in the EZR1 LTR might account for the observed increase in EZR1 transcript levels. It is interesting to note that both AP-1 induction and HIV potentiation by TCDD require enzymatic activity, and presumably reactive oxygen production, by CYP1A1 [187, 190]. Thus, reactive oxygen may constitute a common stimulus for activation of both endogenous and exogenous retroelements by TCDD.

Biological implications of EZR1 expression

The implications of specific retroelement expression independent of transposition are matter for speculation. EZR1 is abundantly expressed in cardiac tissue, possibly in response to specific LTR elements, while *bhikhari* expression in developing mesoderm is driven by activin signaling [179]. Similarly, a murine endogenous retrovirus-like gene is expressed in early mouse embryos and *may* be necessary for progression from 2 to 4 cells [195]. Certain *Drosophila* retrotransposons are also expressed in tissue-specific patterns during embryogenesis. Specificity of expression is one possible indicator of recruitment of novel genomic elements to cellular functions. However, the nature of the functions that might be performed by either EZR1 or *bik* are completely unknown.

The repercussions of EZR1 induction by TCDD are also unclear. It has been hypothesized, by Barbara McClintock and others, that mobile genetic elements may be activated in response to environmental stress in order to facilitate potentially beneficial genome rearrangements (the genome shock theory). Certainly, there is evidence to indicate that diverse mobile elements are activated by a variety of stimuli, including chemicals similar to TCDD [196, 197]. However, given the inability of EZR1 to retrotranspose autonomously, this does not provide a satisfactory explanation in this case.

Perhaps a more relevant analogy would be the induction of endogenous retroviruses in certain disease states. Particularly interesting is the correlation between elevated levels of certain endogenous retroviral transcripts in myocardium and cardiovascular disease in rats [198, 199]. Likewise, EZR1 transcript abundances were observed to be dose-dependently increased by doses of TCDD that caused cardiovascular toxicity and disrupted cardiomyocyte gene expression (Chapter 3). Thus, possible links between EZR1 expression and cardiac malfunction in zebrafish warrant further investigation.

Table 4.1 Predicted transcription factor binding sites in putative EZR1 LTR sequences.

Transcription Factor Name	Matrix Score	Binding Motif	Position (bp)	Strand
AP-1 (Activator Protein 1, c-Fos/c-Jun)	0.902	yaTGACttcwg	47 - 57	plus
	0.882	taTGACtagcc	526 - 536	plus
AP-4 (Activator Protein 4)	0.883	cgCAGCttca	449 - 458	plus
	0.858	atCAGCccct	560 - 569	plus
BRN-2 (Brain-specific POU factor 2)	0.900	cagatatgAAATa(t/g)g	485 - 500	plus
GATA-1	0.948	CtcaGATAtgaaa	483 - 495	plus
	0.944	ggctGATAgcaga	554 - 566	minus
GFI-1 (Growth Factor Independent 1)	0.871	KggracatAATCwgaag	45 - 68	minus
GRE (Glucocorticoid Response Element)	0.918	aggacaaaaTGTTctc	290 - 305	minus
HNF-3b (Hepatocyte Nuclear Factor 3b)	0.888	tgtatTATTttcctt	141 - 155	plus
	0.879	tacaaTGTTtgatga	184 - 198	plus
Ik-2 (Ikaros-like 2)	0.882	agagGGGActga	278 - 289	plus
MZF-1 (Myeloid Zinc Finger 1)	0.986	agaGGGGa	278 - 285	plus
Nkx-2.5 (tinman homolog)	0.884	agAAGTg	334 - 340	plus
SRF (serum response factor)	0.925	gcCCATatttggag	534 - 547	plus
SRY (sex-determining region Y)	0.874	taagACAAaatg	511 - 522	plus
TCF11 homodimer	0.981	GTCAtacagcatt	519 - 531	minus
TH1E47 (Hand1/E47 heterodimer)	0.895	accatggtCTGGtttc	429 - 444	plus

Figure 4.1 Schematic illustration of EZR1 sequence assembly (top), including TR004 ESTs (black), PCR (blue) and 5' RACE (red) products, and the resultant consensus sequence (bottom). Gaps in sequence coverage are indicated by dotted lines. Regions of concentrated sequence variability (pale green), the integrase ORF (*Int*), and a non-coding region of homology to retroviral *env* genes (pale grey) are depicted on the consensus sequence schematic. Flanking direct repeats are indicated by black triangles, with sequences inset.

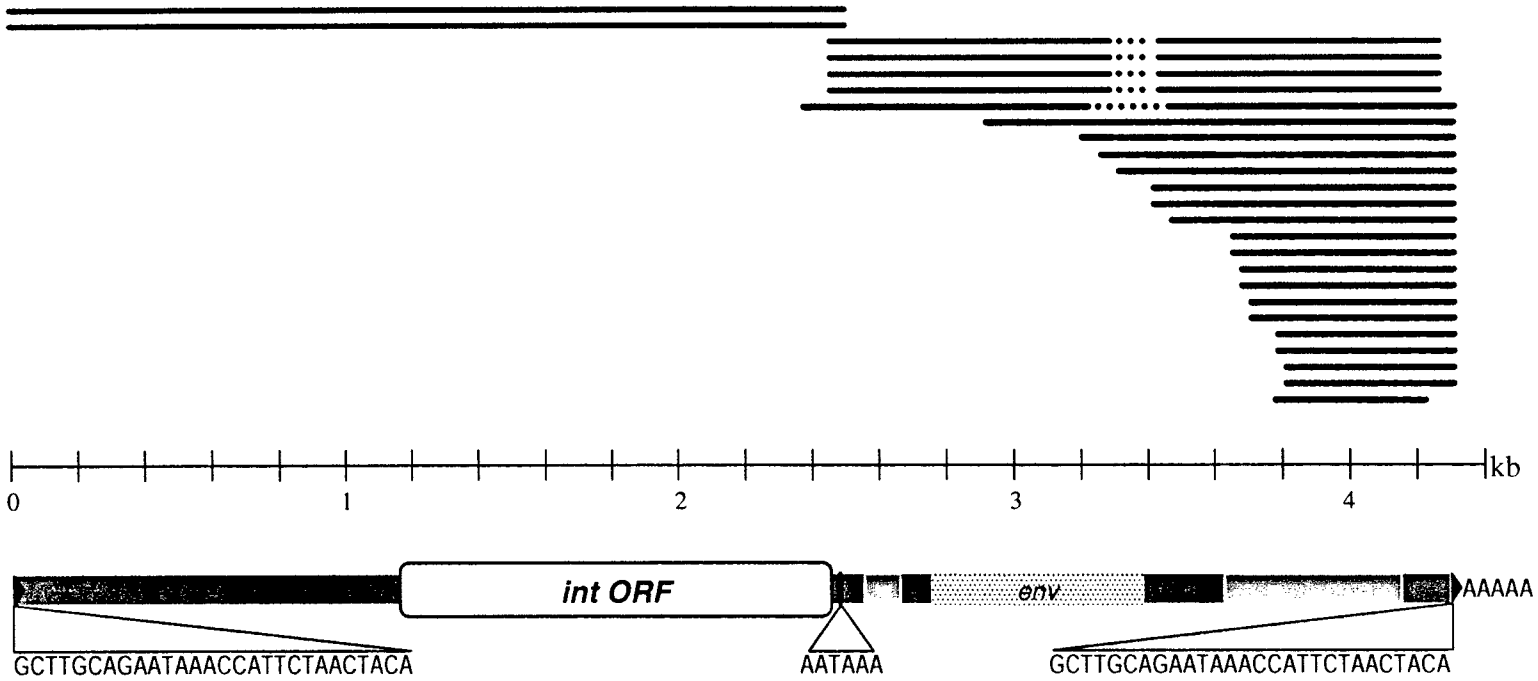
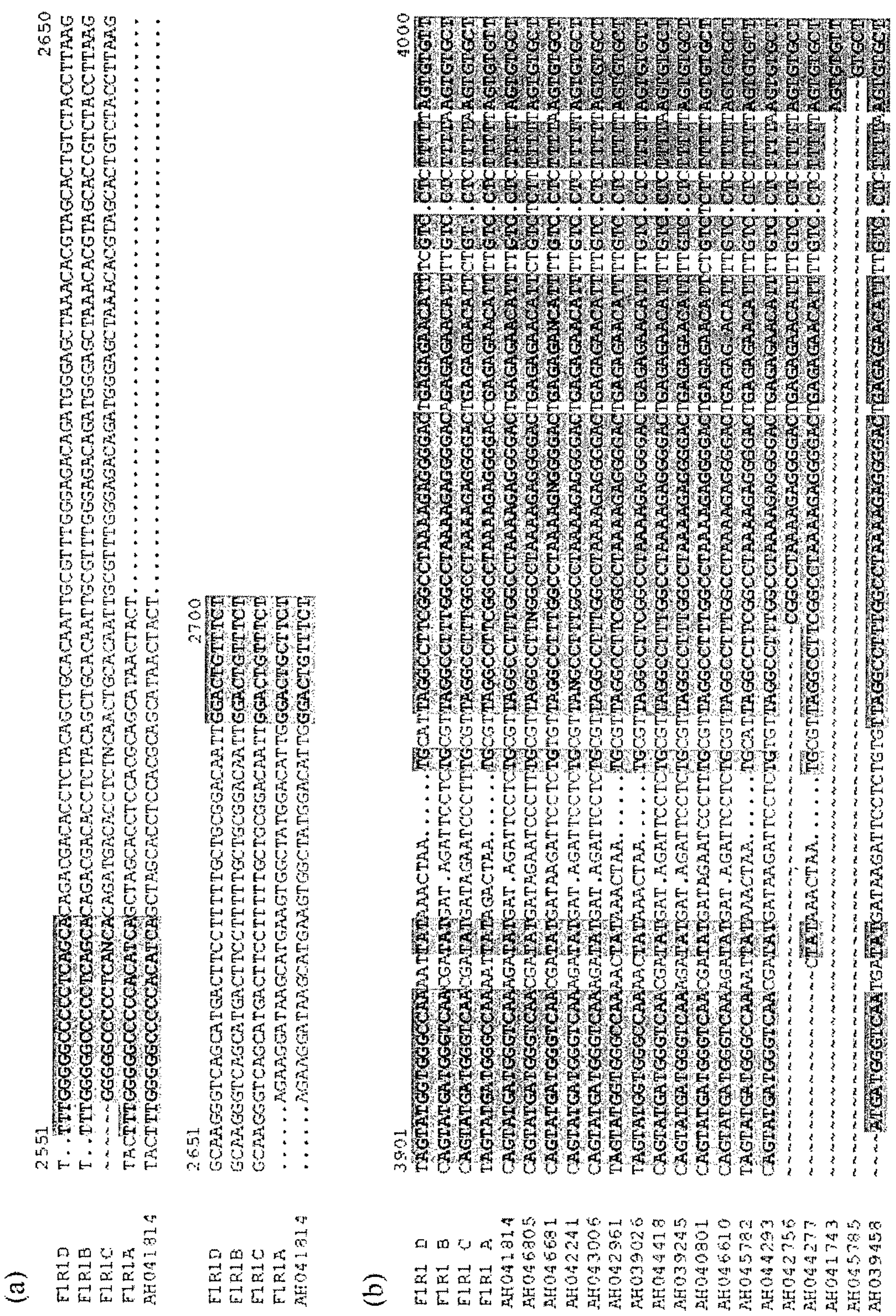

0
1
2
3
4
kb
int ORF
env
AAAAA
GCTTGCAGAATAAACCATTCTAACTACA
AATAAA
GCTTGCAGAATAAACCATTCTAACTACA

Figure 4.2 Multiple sequence alignments taken from major variable regions spanning positions 2567 to 2690 (a) and 3689 to 4114 (b) in the EZR1 sequence assembly. Gaps inserted during sequence alignment are indicated by periods (“.”), lack of sequence data by tildes (“~”). Positions that are ≥95% conserved are highlighted in grey.


(a)
F1R1D
F1R1B
F1R1C
F1R1A
AH041814
2551
2650
2651
2700
(b)
F1R1 D
F1R1 B
F1R1 C
F1R1 A
AH041814
AH046805
AH046681
AH042241
AH043006
AH042961
AH039026
AH044418
AH039245
AH040801
AH046610
AH045782
AH044293
AH042756
AH044277
AH041743
AH045785
AH039458
3901
4000

```
          4001                                                                                              4100
F1R1 D    TGCAGCTACGTGCAGAAGTGCTAGTGGTTTTGTGTGTATATGTATTTGTGTGTCAGAAATATGAGGCCATGGTCTGGTTTCAGTCTTGCAGGTGTTAGAA
F1R1 B    TGCA.TGAGAAGCAGAAGTGTCTGTGGTTT..TGTGTGTGTGTGTGTATCGGAG...GTCTGAAGGTCAACGTCTGGTTTCGGTCTTGCAGGTGTTAAAA
F1R1 C    TGCA.TGAGAAGCAGAAGTGTCTGTGGTTTTTTGTGTGTGTGTGTGTATCGGAG...GTCTGAAGGTCAACGTCTGGTTTCGGTCTTGCAGGTGTTAAAA
F1R1 A    TGCAGCTACGTGCAGAAGTGCTTGTGGTTTTGTGTGTATATGTACTTGTGTGTCAGAAATATGAGGCCATGGTCTGGTTTCAGTTTTGCAGGTGTTAGAA
AH041814  TGCA.TGAGAAGCAGAAGTGTCTGTGGTTT..TGTGTGTGTGTGTGTATCGGAG...GTCTGAAGGTCAACGTCTGGTTTCAGTCTTGCAGGTGTTAAAA
AH046805  TGCAGCTACGTGCAGAAGTGCTTGTGGTTTTGTGTATATGGGNGNGTATATGTCAGAAATATGAGTCCATGGTCTGGTTTCAGTCTTGCAGGTGTTAGAA
AH046691  TGCA.TGAGAAGCAGAAGTGTCTGTGGTTTTTTGNGNGNGTGTGTGTATCGG................................................
AH042241  TGCA.TGAGAAGCAAAAGTGTCTGTGGGTT..TGTGTGGGTGGGGGTATNGGNG...GTTTNANGGTCANCGTNNGGTTTCANTNTTNCNGGNGTTAAAA
AH043006  TGCA.TGAGAAGCAGAAGTGTCTGTGGTTT..TGTGTGTGTGTGTGTATCGGAG...GTCTGAAGGTCAACGTCTGGTTTCAGTCTTGCAGGTGTTAAAA
AH042961  TGCAGCTACGTGCAGAAGTGCTTGTGGTTT..TGTGTATATGTGTGTATATGTCAGAAATATGAGGCCATGGTCTGGTTTCAGTCTTGCAGGTGTTAGAA
AH039026  TGCAGCTACGTGCAGAAGTGCTTGTGGTTTTGTGTGTATATGTATTTGTGTGTCAGAAATATGAGGCCATGGTCTGGTTTCAGTCTTGCAGGTGTTAGAA
AH044418  TGCA.TGAGAAGCAGAAGTGTCTGTGGTTT..TGTGTGTGTGTGTGTATCGGAG...GTCTGAAGGTCAACGTCTGGTTTCGGTCTTGCAGGTGTTAAAA
AH039245  TGCA.TGAGAAGCAGAAGTGTCTGTGGTTT..TGTGTGTGTGTGTGTATCGGAG...GTCTGAAGGTCAACGTCTGGTTTCAGTCTTGCAGGTGTTAAAA
AH040801  TGCAGCTACGTGCAGAAGTGCTTGTGGTTTTGTGTATATGTGTGTGTATATGTCAGAAATATGAGGCCATGGTCTGGTTTCAGTCTTGCAGGTGTTAGAA
AH046610  TGCA.TGAGAAGCAGAAGTGTCTGTGGTTT..TGTGTGTGTGTGTGTATCGGAG...GTCTGAAGGTCAACGTCTGGTTTCAGTCTTGCAGGTGTTAAAA
AH045782  TGCAGCTACGTGCAGAAGTGCTTGTGGTTTTGTGTGTATATGTATTTGTGTGTCAGAAATATGAGGCCATGGTCTGGTTTCAGTCTTGCAGGTGTTAGAA
AH044293  TGCA.TGAGAAGCAGAAGTGTCTGTGGTTTTTTGTGTGTGTGTGTGTATCGGAG...GTCTGAAGGTCAAC.............................
AH042756  TGCAGCTACGTGCAGAAGTGCTTGTGGTTT..TGTGTATATGTGTGTATATGTCAGAAATATGAGGCCATGGTCTGGTTTCAGTCTTGCAGGTGTTAGAA
AH044277  TGCAGCTACGTGCAGAAGTGCTTGTGGTTT..TGTGTATATGTGTGTATATGTCAGAAATATGAGGCCATGGTCTGGTTTCAGTCTTGCAGGTGTTAGAA
AH041743  TGCAGCTACGTGCAGAAGTGCTTGTGGTTTTGTGTGTATATGTATTTGTGTGTCAGAAATATGAGGCCATGGTCTGGTTTCAGTCTTGCAGGTGTTAAAA
AH045785  TGCA.TGAGAAGCAGAAGTGTCTGTGGTTTTTTGTGTGTGTGTGTGCATCGGAG...GTCTGAAGGTCAAC.............................
AH039458  TGCA.TGAGAAGCAGAAGTGTCTGTGGTTT..TGTGTGTGTGTGTGCATCGGAG...GTCTGAAGGTCAATGTCTGGTTTCAGTCTTGCAG~~~~~~~~~

          4101                                             4150
F1R1D     ATATGAGACCATGGTCTGGTTTCATTTCGCAGCTTCACCCACGCTGGAAA
F1R1B     ATATGAGACCATGGTCTGGTTTCATTTCGCAGCTTCACCCACGCTGCAAA
F1R1C     ATATGAGACCATGGTCTGGTTTCATTTCGCAGCTTCACCCACGCTGGAAA
F1R1A     GTCTGAGGCCATGGTCTGGTTTCATTTCGCAGCTTCACCCACGCTGGAAA
AH041814  ATATGAGACCATGGTCTGGTTTCATTTCGCAGCTTCACCCACGCTGGAAA
AH046805  ATATGAGACCATGGTCTGGTTTCATTTCGCAGCTTCACCCACGCTGGAAA
AH046691  ..............TCTGGTTTCATTTCGCAGCTTCACCCCCGCTGGAAA
AH042241  ATANGNGCCCANGGTTNGGTTTCNTTTNNCAGNTTCCCCCCCNNNGGAAA
AH043006  A~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
AH042961  ATATGAGACCATGGTCTGGTTTCATTTCGCAGCTTCACCCACGCTGGAAA
AH039026  ATATGAGACCATGGTCTGGTTTCATTTCGCAGCTTCACCCACGCTGGAAA
AH044418  ATATGAGACCATGGTCTGGTTTCATTTCGCAGCTTCACCCACGCTGCAAA
AH039245  ATATGAGACCATGGTCTGGTTTCATTTCGCAGCTTCACCCACGCTGGAAA
AH040801  ATATGAGACCATGGTCTGGTTTCATTTCGCAGCTTCACCCACGCTGGAAA
AH046610  ATATGAGACCATGGTCTGGTTTCATTTCGCAGCTTCACCCACGCTGGAAA
AH045782  ATATGAGACCATGGTCTGGTTTCATTTCGCAGCTTCACCCACGGTGGAAA
AH044293  .............GTCTGGTTTCATTTCGCAGCTTCACCCACGCTGGAAA
AH042756  ATATGAGACCATGGTCTGGTTTCATTTCGCAGCTTCACCCACGCTGGAAA
AH044277  ATATGAGACCATGGTCTGGTTTCATTTCGCAGCTTCACCCACGCTGGAAA
AH041743  ATATGAGACCATGGTCTGGTTTCATTTCGCAGCTTCACCCACGCTGGAAA
AH045785  .............GTCTGGTTTCATTTCGCAGCTTCACCCACGCTGGAAA
AH039458  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
```

Figure 4.3 Dot-plot comparison of the complete EZR1 consensus sequence to zebrafish genomic sequence contig ctg9483, positions 90,000-95,000.

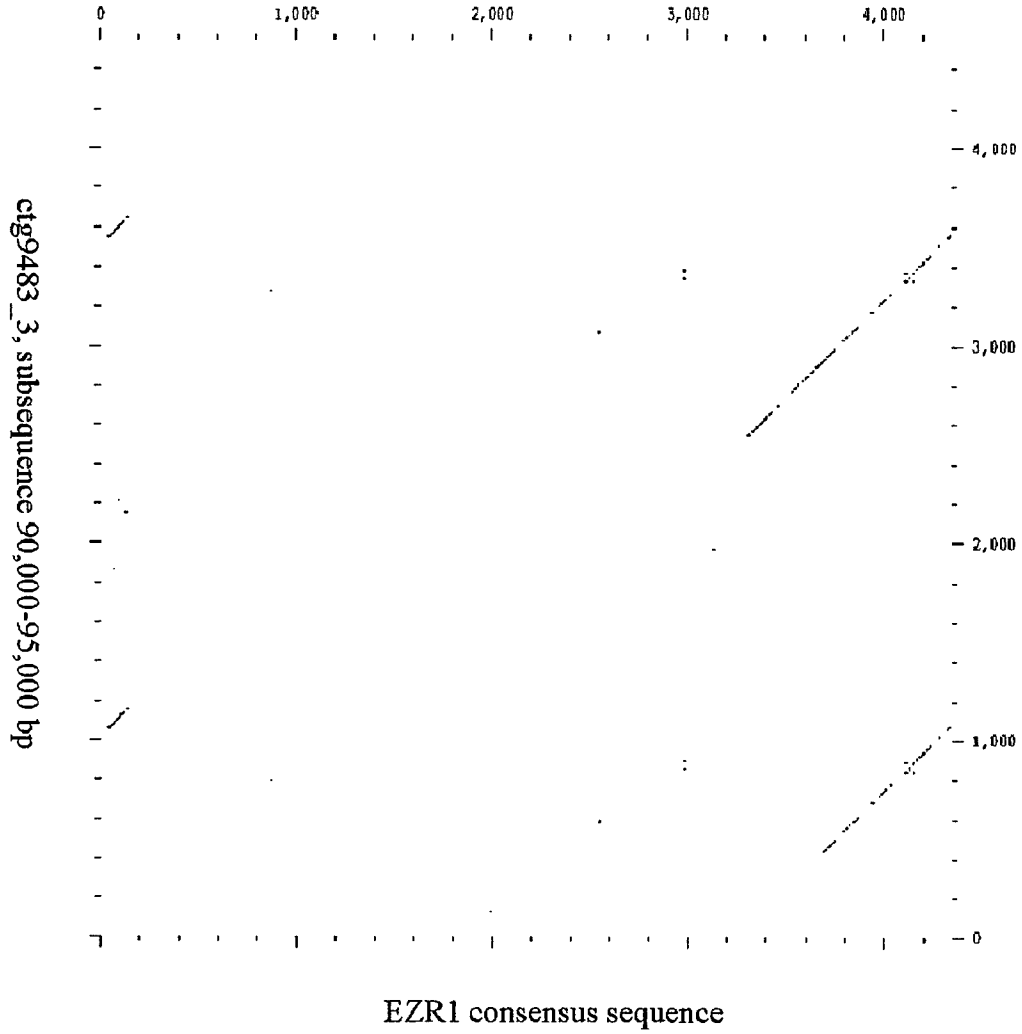

0
1,000
2,000
3,000
4,000
4,000
3,000
2,000
1,000
0
ctg9483_3, subsequence 90,000-95,000 bp
EZR1 consensus sequence

Figure 4.4 Schematic illustration of the putative EZR1 LTR structure, showing U3, R, and U5 domain boundaries and the complete R domain sequence (inset). Triangles depict locations of predicted binding sites for GATA-1 (red), MAF-1/Ik-2 (orange), AP-1 (blue), and TCF-11 (green). Triangles above the grey bar indicate motifs on the plus strand, those below represent motifs on the minus strand.

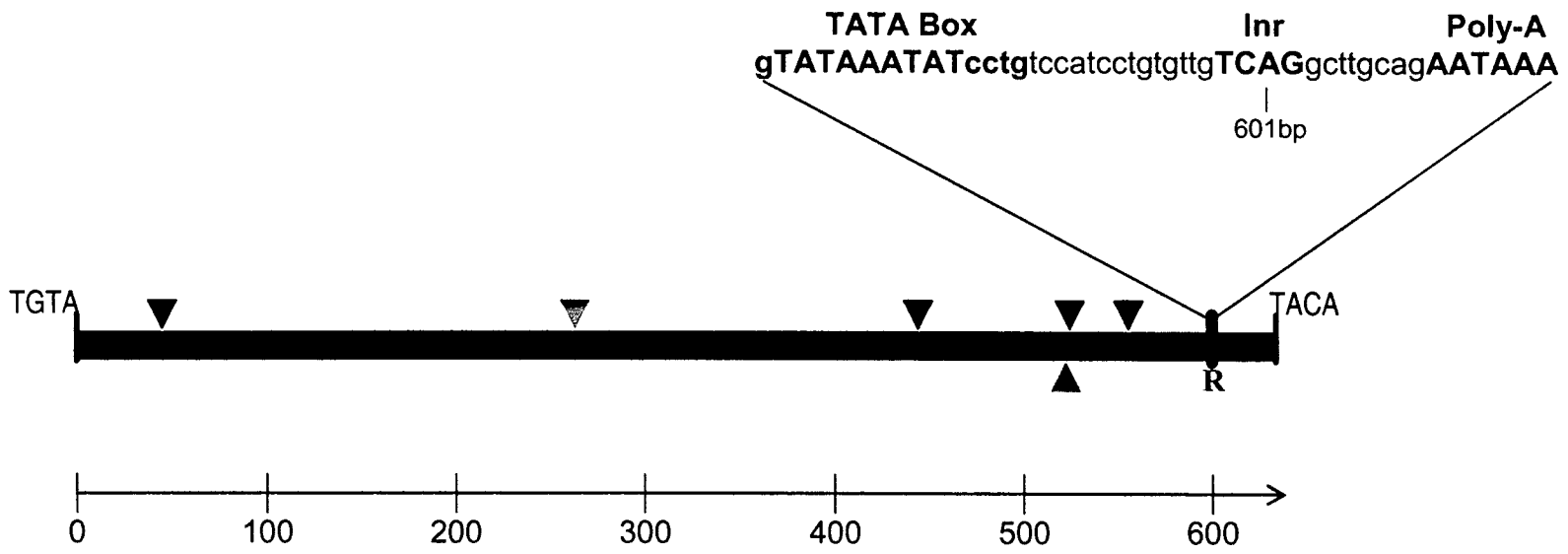

TATA Box
Inr
Poly-A
gTATAAATATcctgtccatcctgtgttgTCAGgcttgcagAATAAA
601bp
TGTA
TACA
R
0
100
200
300
400
500
600

Figure 4.5 Most probable tree morphology, as determined by Bayesian inference of phylogeny, for EZR1 LTR sequences. Nodes with >50% support are labeled with both Bayesian posterior probabilities and maximum likelihood bootstrap support values.

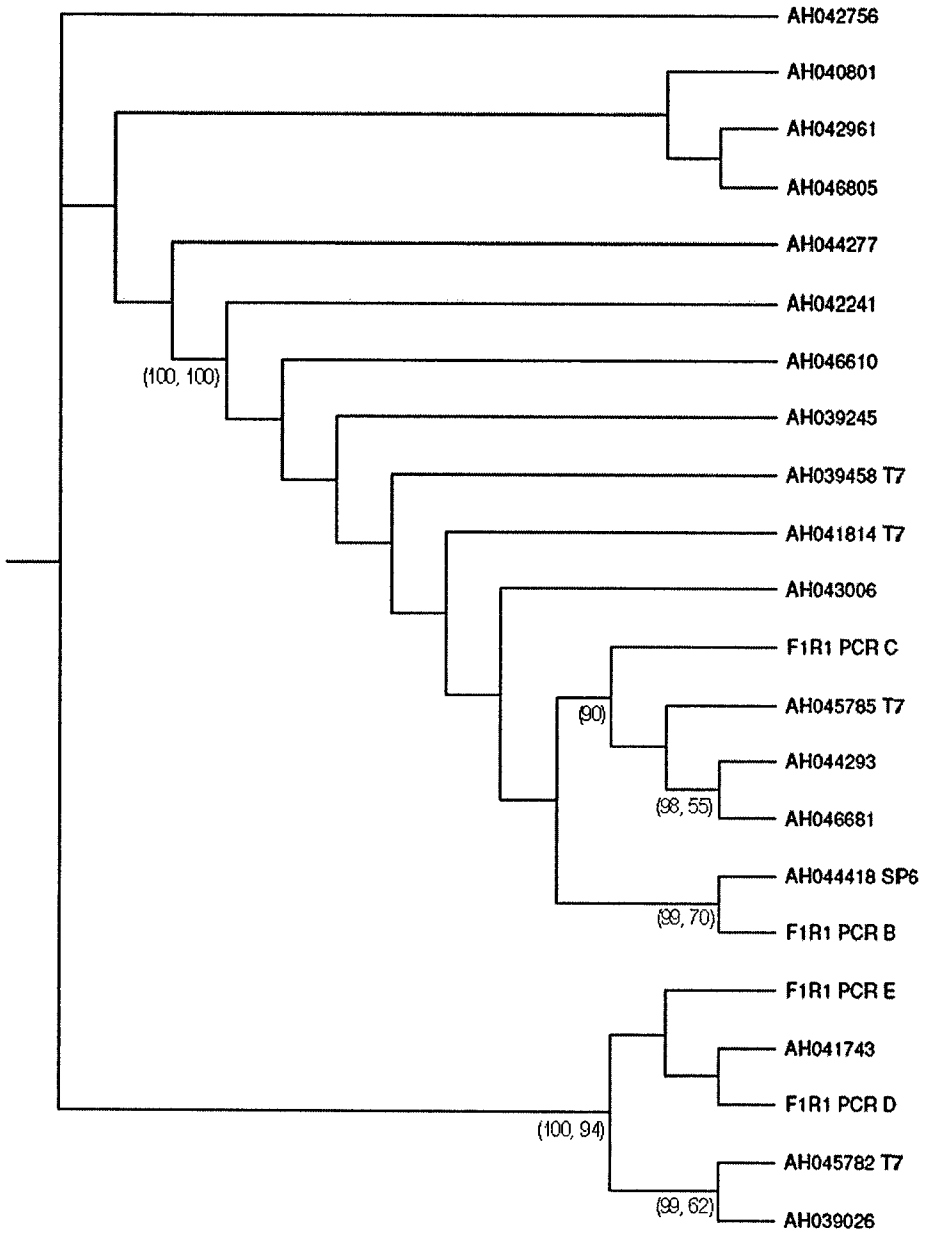

AH042756
AH040801
AH042961
AH046805
AH044277
AH042241
(100, 100)
AH046610
AH039245
AH039458 T7
AH041814 T7
AH043006
F1R1 PCR C
(90)
AH045785 T7
AH044293
(98, 55)
AH046681
AH044418 SP6
(99, 70)
F1R1 PCR B
F1R1 PCR E
AH041743
F1R1 PCR D
(100, 94)
AH045782 T7
(99, 62)
AH039026

Figure 4.6 Representative photographs showing spatial distribution of EZR1 expression in 72hpf zebrafish embryos, as visualized by *in situ* hybridization with anti-sense probes generated from TR004 ESTs AH041814 (a) and AH042756 (b) (blue/purple staining). General, high-level expression of cytochrome c oxidase subunit I is shown for comparison (c).

(a)

(b)

(c)

Figure 4.7 Schematic depiction of canonical domain structures for vertebrate endogenous retroviruses (a), LTR retrotransposons (b), zebrafish bhikhari (c), and EZR1 (d). Green arrows indicate protein coding regions. LTR = long terminal repeat, PBS = primer binding site, PPT = poly-purine tract, ORF = open reading frame, gag = group antigen gene, pol = poly-protein gene, env = envelope protein gene, int = integrase gene.

(a) endogenous retroviruses

(b) LTR retrotransposons

(c) *bhikhari*

(d) EZR1

CHAPTER 5

Conclusions and future work

5.1 Summary

The primary goal of this thesis was to identify TCDD-responsive genes likely to be causally involved in processes of cardiovascular embryotoxicity. Toward this end, we constructed microarrays using cDNA libraries derived from zebrafish embryonic and adult heart tissue. Three sets of embryonic heart arrays were used in methodological testing that led to the development of an effective workflow for high-quality array synthesis and use. These protocols were then used in the production of adult heart microarrays, AH001/A and AH002A/B. AH001/A arrays were used for gene expression profiling of TCDD-treated zebrafish embryos. The results and implications of this work are discussed in greater detail below.

AH002A/B arrays continue to be applied to a variety of collaborative projects. In conjunction with Dr. Hiroki Teraoka's laboratory, I have used AH002 arrays to analyze gene expression in zebrafish embryos exposed to thiuram, an increasingly widespread environmental contaminant. I have been working with Dr. Afonso Bainy to characterize the effects of phenobarbitol and related chemicals on hepatic gene expression in adult zebrafish. Expression profiling has been an integral part Dr. Elaine Joseph's efforts to understand the mechanism by which mutation of a general transcriptional elongation factor translates into a specific cardiac phenotype in zebrafish embryos. Other projects, including examination of zebrafish embryos in which AHR or CYP1A have been knocked down using morpholinos, are in their infancy. In light of the expanding scope of ongoing work, a careful assessment of the current microarray strategy is in order.

5.2 Zebrafish adult heart cDNA microarrays

The approach taken in generating adult heart cDNA microarrays, specifically the use of uncharacterized, redundant probe sets, was (and is) unique. As the full ramifications of such a strategy could not be predicted, going forward with this work required a leap of faith. From a technical perspective, this faith has been born out; there is no evidence of any negative impact on the quality of microarrays or hybridization data. However,

effects on more abstract aspects of this work, such as rate of progress and meaningful interpretation of gene expression data, have been mixed.

As hoped, this approach significantly accelerated the synthesis of microarrays by eliminating months of sequence analysis and cDNA library manipulation. However, much of the time saved at the beginning of the project has simply been delayed until the end stages of each project. At this point, nearly 500 adult heart cDNA library clones have been sequenced as a result of selection by microarray analyses. This number will continue to grow, as AH002 arrays are being used in on-going investigations of several rather disparate conditions (described later in this chapter). High-throughput DNA sequencing and sequence analysis can be largely automated, and does not require the same degree of care and scrutiny that has been applied to targeted sequencing. Thus, characterizing 5000 or more clones would likely have required little additional time, and arraying sequenced clones would have both hastened and improved data analysis.

The most pressing problem in interpreting data based on targeted *a posteriori* sequencing stems from the obvious fact that the identities of unchanged clones remain unknown. Thus, it is unknown whether there are additional clones representing a gene deemed to be differentially expressed that would indicate no change. In this way, the current strategy might lead to overestimations of responsiveness to a given stimulus. In the case of mitochondrial genes, the degree of induction remains uncertain due to a wide range of observed expression ratios. Gene expression results for other genes, such as AHR2, cardiac troponin T2, and myosin isoforms, were significantly less variable. Nonetheless, accurate estimates of variability are crucial in assessing "real" changes, and are not necessarily obtained by targeted *a posteriori* sequencing.

Short of sequencing all arrayed clones, one way of resolving such ambiguity would be to perform microarray hybridizations with labeled DNA from individual differentially expressed clones. This would identify all clones with a given sequence, thereby allowing a complete interpretation of relative expression data for that specific gene. However, even with only 44 observed TCDD-responsive genes, this process would be extremely labor- and resource-intensive. Given current DNA sequencing costs, single-pass

sequences for at least 2500 clones could be obtained at approximately the same price as 50 hybridizations. As AH001 print lots have been nearly exhausted, a major sequencing effort involving that clone set would be unwarranted. A limited number of hybridizations with individual high-priority clones, such as cardiac troponin T2 and some mitochondrial genes, would serve to boost confidence in conclusions drawn from work with AH001 microarrays. In the case of AH002 arrays, though, sequencing of arrayed clones may be the most time- and cost-effective route to high quality gene expression data.

Despite certain drawbacks, the current method presented unique advantages. The discovery of a novel and unusual retroelement, EZR1, highlighted the opportunity for gene discovery via microarrays. Although microarray analysis is often referred to as "blind" or "hypothesis free," the probe design/selection phase of microarray synthesis incorporates significant presuppositions and biases. Certainly, under the priority ranking system applied to the zebrafish embryonic heart library, the combination of excessive redundancy, high genomic copy number, and uncertain gene identity would have eliminated EZR1 ESTs. The *completely* blind nature of the current approach prevented the imposition of such biases, thereby enabling the completely unexpected discovery of EZR1. It is interesting to note that the only other similarly unconventional LTR retroelement, *bhikhari*, was discovered using differential display, another completely blind screening method.

5.3 Expressed Zebrafish Retroelement 1

The discovery of EZR1 is possibly the most intriguing single result of this thesis. EZR1 is a moderate copy LTR retroelement that bears significant resemblance to endogenous retroviruses, but lacks both a *gag* gene and a reverse transcriptase domain. Without these components, EZR1 cannot autonomously function as a retrovirus, or even a transposon. While there are many inactive retroelements in vertebrate genomes, EZR1 is striking in that it is highly expressed in normal cardiac tissue and is robustly induced by TCDD. There are three alternative ways to account for the EZR1 expression pattern – (1) expression of EZR1 elements may be specifically regulated by elements in the LTR, (2) a subset of EZR1 elements may be integrated into the 3' untranslated regions of

cardiac-expressed, TCDD-responsive genes, or (3) EZR1 elements may be induced as part of a more general stress response including activation of mobile elements.

Specific regulation of EZR1 by LTR sequences seems most likely, largely on the basis of observed coordinated expression of 19 distinct EZR1 copies. The alternative explanation, that EZR1 elements have been integrated into 3' UTRs of 19 separate cardiac-expressed genes that are all similarly induced by TCDD, seems contrived. Specific, LTR-driven transcriptional regulation by cellular factors has been clearly demonstrated for developmental expression of *bhikhari* [179], and for induction of HIV activity in TCDD-treated cells [187]. Specific regulation would not necessarily indicate any functionality, but suggests at least the possibility of adoption into some (unknown) cellular process. The same idea has been raised with regard to normal developmental expression of both *bhikhari* [179] and murine ERV-L elements [195].

If, on the other hand, mobile element activation by TCDD is a general phenomenon, there may be broad implications with regard to TCDD toxicity. Elevated levels of endogenous retroviral transcripts are associated with mammalian heart disease [198, 199], providing a potential mechanism for TCDD cardiovascular toxicity. Endogenous retroviruses have been implicated in any number of autoimmune diseases [200], an intriguing observation in light of known immunosuppressive effects of TCDD. Finally, activation of transposable elements might have repercussions for carcinogenesis or even next-generation congenital disease. While compelling in its universality, this hypothesis does not account for high levels of EZR1 expression in normal tissue. Of course, a combination of regulatory mechanisms is possible.

These speculations suggest several avenues for further investigation in this area. First and foremost, full-length genomic and cDNA copies of EZR1 elements should be amplified by PCR and sequenced to absolutely confirm the structure of EZR1; this should be a straightforward process given the sequence data contained in this thesis. Two further experiments would help distinguish between the above-mentioned hypotheses. Firstly, inverse genomic PCR could be used to determine the genomic context of integrated EZR1 elements. Unfortunately, due to the high rate of misassembly, the

zebrafish genome project in its current state is of little help in this endeavor. Secondly, full-length EZR1 clones could be subjected to standard promoter analysis to determine regulatory LTR sequences. While it would be interesting to know whether TCDD causes a general activation of mobile DNA elements, current knowledge of zebrafish transposons is too limited to allow a broad sampling. This question might be better addressed using human or murine cell culture.

5.4 TCDD-induced dilated cardiomyopathy

It will be important to further investigate the nature, origin, and impact of TCDD-induced changes in expression of cardiac troponin T2 and cardiac myosins. The microarray and RT-PCR data in this thesis are in direct conflict regarding the direction of change in expression of cardiac troponin T2, but agree that TCDD exposure results in significant differential expression of cardiac troponin T2. Resolving this difference should be a top priority. However, it is difficult to say how this should be accomplished. Additional microarray hybridizations using dye-swapping might be tried, although this seems unlikely to resolve the issue since there are no indications that amino-allyl post-labeling is subject to dye bias. As troponins and myosins are highly expressed, Northern blot analysis should be possible and would provide a third data source. It would also be interesting to examine upstream genomic DNA sequence for known regulatory elements. Such information could not substitute for experimental validation, but might add weight to one set of observations by providing a potential mechanism for either induction or suppression.

Ultimately, any change in cardiac troponin T2 is consistent with cardiomyopathy, as is induction of myosins. Furthermore, TCDD-induced dilated cardiomyopathy accompanied by elevated myosin levels has been clearly demonstrated chick embryos [32, 33]. Thus, the relevant question is no longer "What is the nature of cardiac impacts of TCDD?" but rather "How does TCDD cause cardiomyopathy?" Given the current data, it is impossible to know whether observed changes in gene expression are causally related to toxic impacts or whether they are secondary manifestations of toxicity.

Obviously, clarifying this relationship will be crucial in advancing our understanding of toxic mechanisms.

One way this might be accomplished would be to examine gene expression at multiple times during the progression of TCDD toxicity. As the window of susceptibility to TCDD cardiovascular toxicity is limited to a 24-hour period (48-72 hpf) in zebrafish [24], it would be possible to obtain a high resolution temporal map of gene expression with perhaps a dozen sampling times. In this way, likely causative events (i.e., changes in gene expression that precede toxic impacts) could be teased apart from later secondary responses. Time-courses at multiple TCDD concentrations could also provide information about the interplay between dose level and exposure time. Dose-dependent differences in gene expression profiles observed at 72 hpf may be reflective of differential rates of progression of toxicity rather than completely distinct processes; so-called "low dose specific" responses may simply occur at an earlier time at higher doses.

Thus, establishing a three-dimensional dose-time-response surface would be extremely informative. It was originally hoped that this could be accomplished as part of this thesis work, but the labor and cost required for such an experiment was prohibitive. However, RT-PCR could be used to describe dose-time-response relationships for a limited number of genes identified by microarray analyses; real-time RT-PCR data presented in Chapter 3 are a step in this direction.

5.5 TCDD and reactive oxygen species (ROS)

Stimulation of reactive oxygen production is an increasingly common theme in the study of TCDD, one touched on by several aspects of this thesis. Cytochrome P450 1A(1) [64] and mitochondria have been identified as AHR-dependent sources of ROS [163, 164]. Whereas current knowledge can be compiled into a detailed hypothesis covering steps from AHR activation to decoupling of specific steps in the CYP1A catalytic cycle (see Chapter 1), no such context exists for mitochondrial ROS production. TCDD-induced differential expression of mitochondrial and downstream energy transfer genes has been observed in this and other gene expression profiling work [123, 124], and suggests a mechanism for decoupling of mitochondrial electron transfer associated with

ROS production [165]. In turn, mitochondrial ROS production might be one pathway to cardiomyopathy and cardiovascular failure; evidence of a causal link between excess reactive oxygen and cardiac pathologies is increasing [201-203].

Accounting for differential expression of mitochondrial genes is more difficult, and consideration of this question leads to the conclusion that mitochondrial dysfunction may, itself, be a secondary effect of other reactive oxygen production (i.e., CYP1A). The events triggering mitochondrial ROS production related to heart disease are unknown. TCDD-induced mitochondrial ROS production requires AHR [164], but AHR modulation of gene expression must be via indirect means as there are no AHR binding elements in regulatory regions of the zebrafish mitochondrial genome. Data from CYP1A1-null knock-out mice suggest that CYP1A1 may be the link between AHR and mitochondria [70]. As CYP1A(1) has no inherent transcriptional regulatory capacity, such effects would presumably be mediated by ROS.

Indeed, reactive oxygen may be a common regulatory force governing many of the changes in gene expression documented in this thesis. As previously noted, there is growing support for a role of reactive oxygen in generation of various heart diseases; the exact nature of that role is uncertain, but would likely involve modulation of gene expression. Thus, it would be extremely interesting to see which aspects of TCDD gene expression profiles can be mimicked by direct exposure of zebrafish embryos to reactive oxygen species, such as hydrogen peroxide. Conversely, it would be interesting to compare TCDD gene expression profiles in the presence and absence of anti-oxidants.

In particular, reactive oxygen signaling via the redox-sensitive transcription factors NF-κB and AP-1 may be important in TCDD toxicity. Both factors are activated by TCDD [190] and have been implicated, by this and other work, in various TCDD responses. NF-κB is necessary for activation of HIV by TCDD [187, 204], while AP-1 activity could contribute to hypothesized LTR-driven regulation of EZR1. AP-1 binding sites in upstream regions of mammalian glutathione S transferase genes are thought to be responsible for ROS activation [153, 204]. As the zebrafish genome project progresses, it will be important to search upstream regions of TCDD-responsive genes for AP-1 and

NF-κB binding sites. It would also be useful to directly probe the roles of NF-κB and AP-1 in cardiovascular embryotoxicity by either (a) over-expressing these proteins in 48-72 hpf embryos, or (b) using morpholino technology to knock down functional protein levels in TCDD-treated zebrafish embryos.

5.6 Alternative mechanisms of gene expression regulation

While this work has focused on transcriptional modulation by TCDD, there are alternative pathways that might lead to altered expression profiles. For example, elevated levels of mRNAs for mitochondrially-encoded genes might reflect an increase in the number of mitochondria per cell. Increased mitochondrial density is seen in cases of elevated oxygen and/or energy demand [205], and could be part of an adaptive response to either general physiological stress caused by toxicant exposure, or to specific cardiac impairment by TCDD. This explanation would account for the lack of relevant transcription factor binding sites in the mitochondrial genome, as well as the universality of mitochondrial gene induction. This alternative could also have implications for toxicity; proportionate increases in mitochondrion abundance and mitochondrial gene expression might be less likely to result in aberrant ROS production than would overexpression of specific proteins within mitochondria. Thus, it would be interesting to determine mitochondrial density, both generally and within cardiomyocytes, in TCDD-treated zebrafish embryos.

Alterations in RNA stability might contribute to observed differential expression of some genes. Specific regulation of RNA degradation is a feature of developmental processes, steroid hormone signaling, and stress responses caused by hypoxia [206]. Modified RNA stability and secondary structure is an important aspect of retroviral replication and gene expression. Mitochondrial genes are also subject to post-transcriptional regulation [207-209]. Furthermore, observed differential expression of pyrimidine 5' nucleotidase and ribosomal RNAs and proteins might have implications for RNA degradation and post-transcriptional processing. Further investigation of RNA stability following TCDD treatment would be intriguing, as it represents a potentially AHR-independent mechanism of regulation of gene expression.

5.7 Conclusions

The ability to target specific cellular processes and formulate detailed hypotheses regarding TCDD embryotoxicity is an indicator of the progress that has been made over the course of this thesis. In large part, the impetus to generate zebrafish cardiovascular microarrays stemmed from frustration with a fundamental lack of information about cardiac impacts of TCDD. The strength of microarrays is not in their ability to address specific hypotheses, but rather, to generate a large body of observations that can be collated and sorted to produce workable theories and testable hypotheses. This work has provided a significant body of observations and hypotheses on which to build future investigation of the mechanisms of TCDD cardiovascular embryotoxicity. Furthermore, the discovery of EZR1 has provided an intriguing introduction into the poorly explored area of chemical regulation of endogenous retroelements, and remaining ESTs offer opportunities for exploration of novel aspects of TCDD activity.

APPENDIX A.

Expression of vascular endothelial growth factor in early zebrafish embryos is unaffected by 2,3,7,8-tetrachlorodibenzo-*p*-dioxin exposure

A.1 Introduction

TCDD (2,3,7,8-tetrachlorodibenzo-p-dioxin) is a potent and environmentally widespread teratogen that severely disrupts cardiovascular development. The hallmarks of embryonic TCDD exposure are edema, hemorrhage, craniofacial malformations, and early life stage mortality. This suite of symptoms, similar to blue sac syndrome in salmonid fish, has been observed in over a dozen fish species exposed to TCDD and related chemicals [9-15]. Detailed study of cardiovascular embryotoxicity in zebrafish (*Danio rerio*) has revealed additional impacts, including circulatory failure [22-24], loss of erythrocytes [24], and reductions in heart size and cardiac contractile strength [31]. Similar phenotypes have been observed in embryos of birds [16-20] and rodents [21] exposed to TCDD and related compounds.

The embryotoxic impacts of TCDD are primarily mediated by the aryl hydrocarbon receptor (AHR) [22, 53-59], a basic-helix-loop-helix Per-ARNT-Sim (bHLH-PAS) protein that functions as a ligand-activated transcription factor with a broad affinity for aromatic hydrocarbons [36]. Toxicity-eliciting events downstream of AHR activation are poorly understood, but several lines of evidence have suggested a possible role for vascular endothelial growth factor (VEGF).

Vascular endothelial growth factor (VEGF) is an endothelial cell-specific mitogen that is responsible for dictating formation and organization of new blood vessels, as well as regulating permeability of vessels. Vascular endothelial cells are known to be sensitive targets for enzyme induction, apoptosis, and morphological alteration caused by TCDD [26-28]. The phenotype of overexpression of VEGF in avian embryos shares some features with TCDD toxicity, including increased vascular permeability and widespread edema [210].

There are multiple avenues by which AHR might influence VEGF expression; the most direct route would be cross-talk between AHR and hypoxia inducible factor 1α (HIF-1α). AHR and HIF-1α share a common dimerization partner, aryl hydrocarbon receptor nuclear translocator (ARNT). ARNT is absolutely necessary for both TCDD-activated AHR signaling and HIF-1α dependent induction of VEGF [89, 211, Park, 1999

#1706]. HIF-1α-like factor (HLF) also interacts with ARNT to form a transcription factor that regulates VEGF expression [212]. It has been hypothesized that elevated levels of active AHR might lead to competition for ARNT binding and disruption of signaling pathways governed by factors with lesser ARNT binding affinities. Competition for ARNT binding has been demonstrated between HIF-1α and AHR *in vitro* [213], but there is no compelling evidence of an effect on either signaling pathway [90-92].

Two indirect pathways, involving cytokines or reactive oxygen species as intermediates, provide alternative means for AHR modulation of VEGF expression. VEGF is subject to induction by various cytokines and growth factors, including interleukin-1β, tumor necrosis factor-α and transforming growth factor-β1 (Neufeld et al. 1999), all of which are up-regulated by TCDD exposure [214]. VEGF expression and transcript stability are also increased by reactive oxygen species, including superoxide and hydrogen peroxide [215, 216]. Accordingly, TCDD exposure causes AHR-dependent reactive oxygen production and oxidative damage that has been associated with toxic impacts [26, 64, 86, 87, 163, 164].

Based on an abundance of potential mechanisms for regulation of VEGF by TCDD, we undertook to determine whether TCDD exposure alters VEGF expression in zebrafish embryos. Two RT-PCR methods were used to assess expression levels of β-actin (negative control), CYP1A (positive control), and VEGF in 12 hpf and 24 hpf embryos following exposure to a toxic (~ED65 for cardiovascular impacts) dose of TCDD.

A.2 Methods

Embryos and RNA

Prior to 6 hours post fertilization (hpf), synchronous zebrafish (*Danio rerio*) embryos were injected with triolein (vehicle) or 3pg TCDD, or left uninjected. Groups of at least 25 embryos were flash-frozen at either 12 hpf or 24 hpf, then held at –80°C (Table A.1). Total RNA was isolated from whole embryos using 10μl RNA STAT-60 per mg of tissue, according to the manufacturer's suggested protocol. RNA was dissolved in

DEPC-treated water and quantified by means of UV spectrophotometry (Shimadzu UV-2401PC with UV Photometric software). The integrity of RNA was confirmed by agarose gel electrophoresis.

Competitive RT-PCR

Competitor template construction

Heterologous competitor DNA templates were constructed using the PanVera Competitive DNA Construction Kit (PanVera Corporation, Madison WI), which provides a λDNA template that can be used to construct competitors of any size up to 600bp. PCR primers were designed to amplify unique fragments of the λDNA template. The corresponding gene-specific primer sequence (i.e. sense or anti-sense for a given target) was then appended to the 5' end of each primer. These composite primers were used to synthesize heterologous DNA competitors ~10% shorter than gene-specific PCR products. Composite primers for a VEGF competitor amplified a 328 bp fragment of the λDNA template (bold indicates gene-specific primer sequences; see below):

5' – **ctcgcggctctcctccatctg**tgtgaagacgacgcgaaattcagc – 3'

5' – **cttctgcctttggcctgcattc**ggaaaccagtttcttgttgttcg – 3'

Primer sequences used to obtain the β-actin competitor were:

5' – **cgacccagacatcagggagtg**tgtgaagacgacgcgaaattcagc – 3'

5' – **gtccagggccacatagcacag**acgccgcgaccaggagaacg – 3'

To 25μl 2X Premix Solution (PanVera Corporation, Madison WI) were added primers (10pmol each) and dH_2O to a final reaction volume of 50μl. An initial 5-minute denaturation step at 94°C preceeded 30 cycles of 30 seconds at 94°C, 30 seconds at 60°C, then 45 seconds at 72°C. This was followed by a final 7-minute extension step at 72°C.

PCR products were purified according to manufacturer's specifications using SUPREC™-02 (PanVera Corporation, Madison WI). Purified PCR products were analyzed in agarose gel (2% in 1X TAE buffer). Competitor DNA templates of the desired sizes were cut from the gel and extracted from agarose using GENECLEAN® III

Kit (BIO 101, Vista CA). JM109 cells were transfected with recombinant pGEM-T Easy vectors containing the appropriate competitor DNA fragment (Promega Corporation, Madison WI). Mini-preps of plasmid DNA (Qiagen Spin Prep Mini-Preps) were quantified by UV spectrophotometry and gel electrophoresis and used in competitive PCR reactions.

Reverse Transcription

Reverse transcription reactions (50 μl) consisted of 50 ng total RNA, MuLV Reverse Transcriptase (2.5U/μl), random hexameric primers (2.5μM), dNTPs (1mM), 5mM $MgCl_2$, 1X PCR Buffer II, and Rnase Inhibitor (1U/μl) (all reagents by Perkin Elmer Applied Biosystems). To allow complete priming, reactions were incubated at 25°C for 10 minutes. Reverse transcription was carried out for 15 minutes at 42°C. Reactions were then heated to 99°C for 5 minutes, and finally, cooled to 5°C for 5 minutes. cDNA was stored overnight at 4°C, then at -20°C until use.

Competitive PCR

PCR was carried out using reagents from Perkin Elmer Applied Biosystems (Foster City CA). 50μl reaction volumes contained AmpliTaq Gold DNA Polymerase (1U), 2.5mM $MgCl_2$, 1X PCR Gold Buffer, cDNA from 10ng (β-Actin) or 100ng (all others) total RNA, 0.2μM primers (sequences and product sizes can be found in Table 1), and appropriate competitor template. PCR conditions were as described for the construction of competitor DNA templates.

PCR quantification

Aliquots of PCR reactions were subjected to polyacrylamide gel electrophoresis in 6% TBE gels (NOVEX). PCR products were detected by staining with ethidium bromide (1μg/ml in 1x TBE). Gels were digitally photographed and negative images were subjected to spot densitometry analysis (ChemImager). Ethidium bromide fluorescence intensity for each band was plotted against the known competitor concentration, and best-fit trend lines were determined for competitor and target template individually. The

absolute quantity of target template was calculated as the intersection of the two lines (e.g., Figure A.1).

A.3 Results

Standard RT-PCR

Initially, expression of β-actin, VEGF, and CYP1A was examined using standard RT-PCR methods; reaction conditions were identical to those described for competitive PCR. β-actin was used as a control for technical variation; fluorescence intensities for VEGF and CYP1A PCR products visualized by gel electrophoresis were normalized to β-actin intensities. These data indicated no change in either VEGF or CYP1A expression at 12 hpf, but suggested TCDD-specific induction of both genes at 24 hpf (Figure A.2). This trend was not statistically significant (two-factor ANOVA, p-values >0.05), and was only apparent after normalization to β-actin. These observations suggested that variation in β-actin measurements was a significant confounding factor in these analyses.

Competitive RT-PCR

Two competitive RT-PCR experiments were conducted. In the first, we evaluated message levels for β-actin and VEGF in RNA from 12 hpf and 24 hpf control, triolein-treated, and TCDD-injected embryos. There was insufficient RNA from 12 hpf triolein-injected embryos for another replicate. Thus, in the second experiment, β-actin, VEGF and CYP1A levels were measured in all RNA samples except that from 12 hpf triolein-treated embryos

Mean β-actin expression levels did not differ significantly between any of the treatment groups at either 12 hpf or 24 hpf (two-factor ANOVA, p-values >0.05). However, there was a trend toward reduced β-actin levels in the 24 hpf triolein-treated embryos (Figure A.3). As this trend was observed in both PCR replicates, it seemed unlikely that this was an artifact of human error, such as pipetting in accuracy. This observation brought into question the validity of normalizing VEGF and CYP1A data to

β-actin measurements. Both raw data and normalized values are presented, and lead to the same conclusions.

At 12 hpf, CYP1A levels were insufficient to allow precise quantification using current PCR conditions. However, PCRs with $1x10^5$ copies of CYP1A competitor yielded target and competitor product bands of approximately equal strength (data not shown), indicating approximately 1000 copies CYP1A/ng RNA at 12 hpf, regardless of TCDD treatment. Basal CYP1A expression at 24 hpf was much greater than at 12 hpf, but still $<1x10^5$ copies/ng RNA in both control samples. Both raw and β-actin normalized data unequivocally indicate that CYP1A expression at 24 hpf was significantly induced by TCDD (single-factor ANOVA, p-value <0.05). Raw data indicated ~110-fold increase in CYP1A mRNA copy number (Figure A.4), while normalized data suggested more moderate induction of ~45-fold (Figure A.5).

Basal VEGF expression levels more than doubled from an average of $8.6x10^4$ copies/ng RNA at 12hpf to $2.4x10^5$copies/ng RNA at 24hpf (two-factor ANOVA, p-value <0.05); this increase in constitutive expression was also reflected in β-actin normalized data (Figure A.5). Neither raw data (Figure A.4) nor normalized values (Figure A.5) indicated any significant effect of TCDD treatment on VEGF expression levels (two-factor ANOVA, p-values >0.05).

A.4 Discussion

Ultimately, normalization to β-actin seemed to be the most appropriate way of handling all RT-PCR data. Each gene showed a trend toward lower copy numbers in the 24 hpf triolein sample, suggesting some fundamental difference in this RNA preparation. Perhaps this sample was contaminated by genomic DNA, resulting in an underestimation of actual total RNA used in each reaction.

Furthermore, normalized values for CYP1A induction at 24 hpf closely accord with other reports of CYP1A induction in zebrafish embryos. Microarray analysis of 72 hpf embryos exposed to ~2pg TCDD indicated 29-fold induction, whereas real-time RT-PCR analyses at the same dose level indicated 60-90-fold induction (Chapter 3). In another

case, >20-fold induction of CYP1A was observed at several times following exposure of embryos to 1.5nM TCDD in egg water, a concentration that should produce embryo burdens approximately twice current levels [42].

All RT-PCR data lead to the same conclusion regarding vascular endothelial growth factor expression, namely that it is not significantly impacted by exposure to toxic doses of TCDD. In light of more recent advances in understanding of TCDD toxicity, this is not surprising. At the time this work was conceived, it was generally thought that cardiovascular impacts late in development were the result subtle disruptions in early patterning events. However, it is now clear that the window of susceptibility for cardiovascular toxicity actually falls between 48 and 72 hpf [24]. Thus, an early change in expression would not necessarily have implicated VEGF in TCDD toxicity. Likewise, the observed lack of change does not rule out the possibility of involvement of VEGF in TCDD toxicity.

The relationship between TCDD exposure and VEGF expression appears to be complex, depending heavily on cell type and dose level. VEGF expression was *induced* in human hepatoma cells [93], but not in murine liver, spleen or thymus [95]. In yet another study, TCDD *suppressed* VEGF expression in one lung epithelial cell line, but did not affect expression in another [94]. As embryonic development, itself, comprises a complex and unique cellular environment it would be interesting to know if VEGF expression is altered later in development. Unfortunately, VEGF was not well represented on zebrafish cardiovascular cDNA arrays used to examine TCDD-influenced gene expression at 72 hpf (Chapters 2 & 3). However, other work to clarify the response of VEGF to embryonic TCDD exposure may already by underway [S. Billiard, pers. comm.].

Table A.1 Number of embryos and total tissue weight in flash-frozen samples from which total RNA was isolated.

TREATMENT	12hpf # Embryos	12hpf Weight (mg)	24hpf # Embryos	24hpf Weight (mg)
Uninjected	67	70.5	55	50.7
Triolein	28	38.8	29	25.8
3pg TCDD	44	52.6	41	38.1

Figure A.1 Example of competitive RT-PCR data, showing linear best-fit trendlines and equations used to calculate absolute copy number for target sequences. Densitometry results for VEGF target sequence are indicated by squares, results for competitor sequence by diamonds.

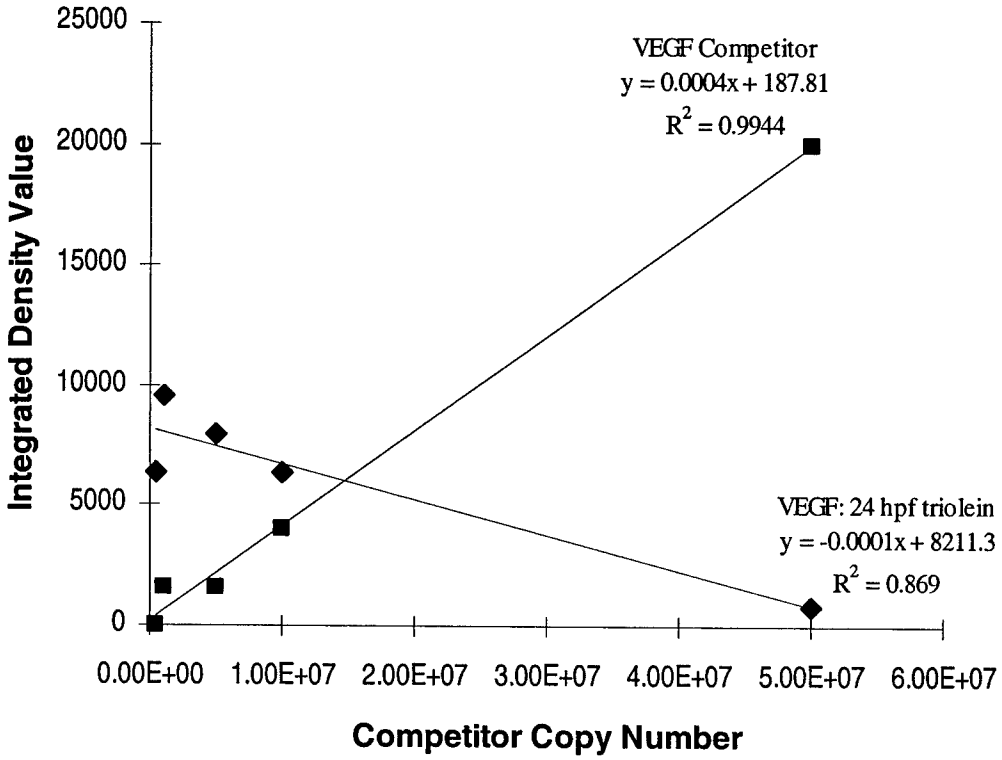

25000
20000
15000
10000
5000
0
Integrated Density Value
VEGF Competitor
y = 0.0004x + 187.81
R2 = 0.9944
VEGF: 24 hpf triolein
y = -0.0001x + 8211.3
R2 = 0.869
0.00E+00
1.00E+07
2.00E+07
3.00E+07
4.00E+07
5.00E+07
6.00E+07
Competitor Copy Number

Figure A.2 Relative expression levels of vascular endothelial growth factor (striped) and cytochrome P450 1A (solid) in control and TCDD-injected zebrafish embryos. Data from duplicate RT-PCR experiments were normalized to β-actin. Mean values are shown with error bars representing one standard deviation.

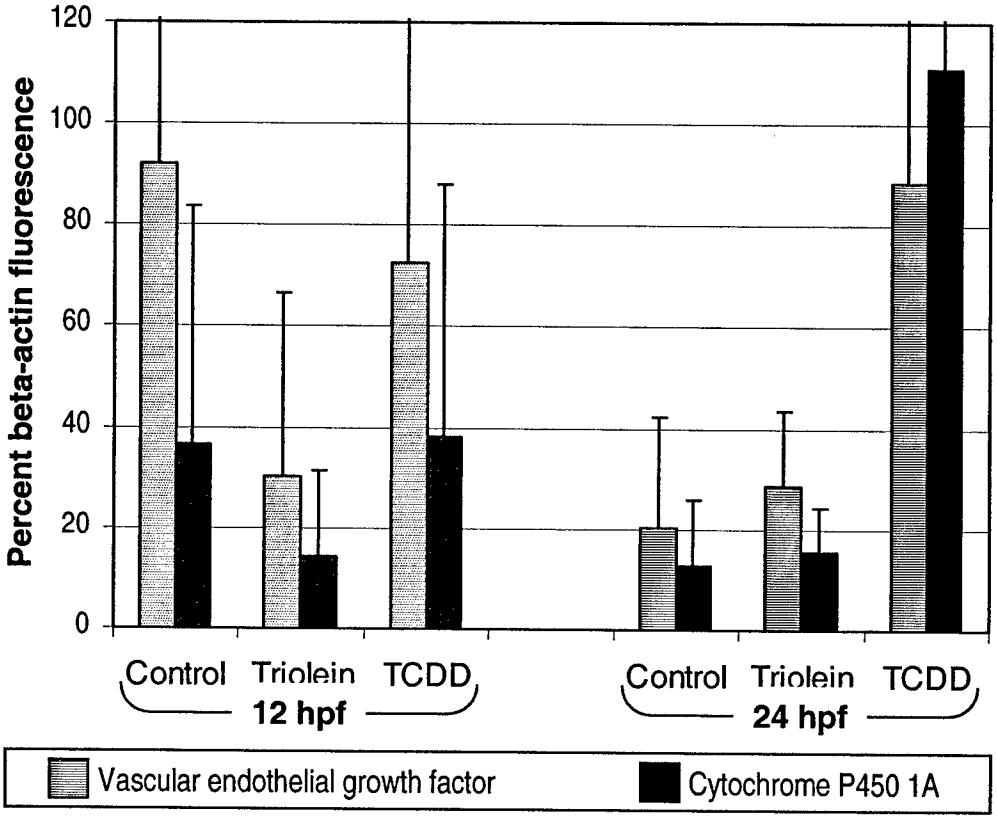

Percent beta-actin fluorescence
120
100
80
60
40
20
0
Control
Triolein
TCDD
12 hpf
Control
Triolein
TCDD
24 hpf
Vascular endothelial growth factor
Cytochrome P450 1A

Figure A.3 Absolute β-actin mRNA expression levels, as measured in two replicates of competitive RT-PCR.

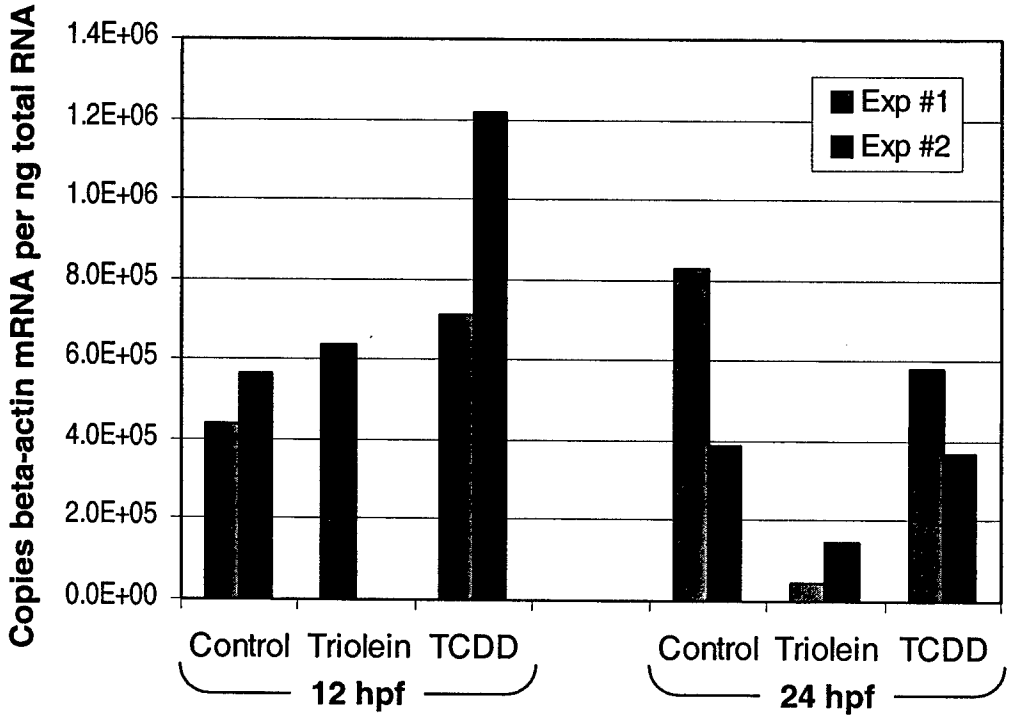

Copies beta-actin mRNA per ng total RNA
1.4E+06
1.2E+06
1.0E+06
8.0E+05
6.0E+05
4.0E+05
2.0E+05
0.0E+00
Exp #1
Exp #2
Control
Triolein
TCDD
12 hpf
Control
Triolein
TCDD
24 hpf

Figure A.4 Absolute quantitation by of VEGF and CYP1A mRNA levels in control and TCDD-injected zebrafish embryos. VEGF data are mean values of two competitive RT-PCR replicates; error bars represent one standard deviation. CYP1A values were derived from a single experiment.

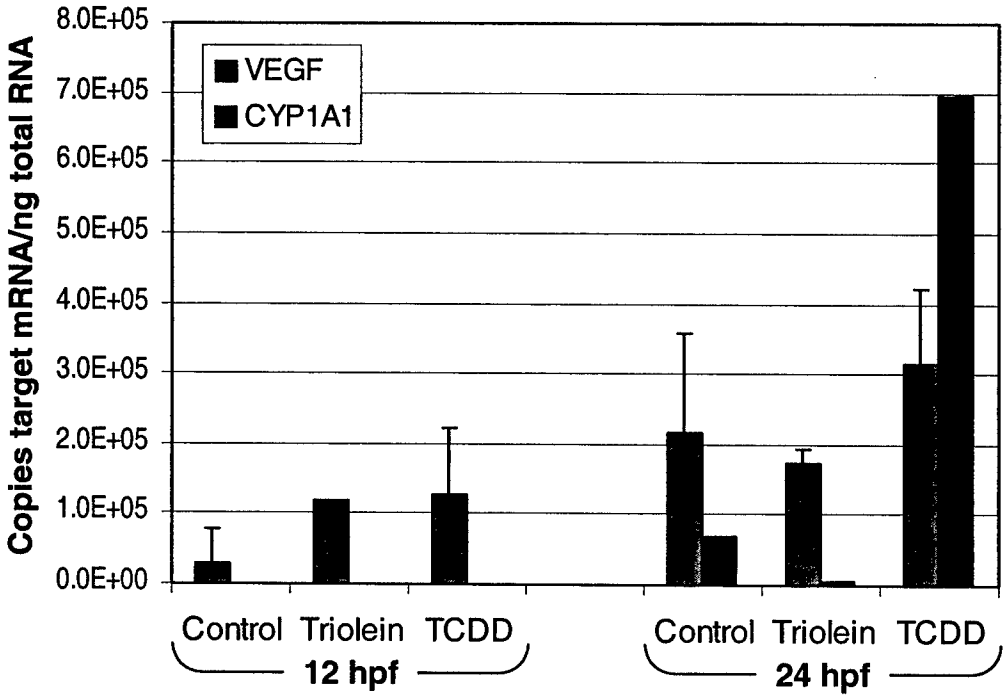
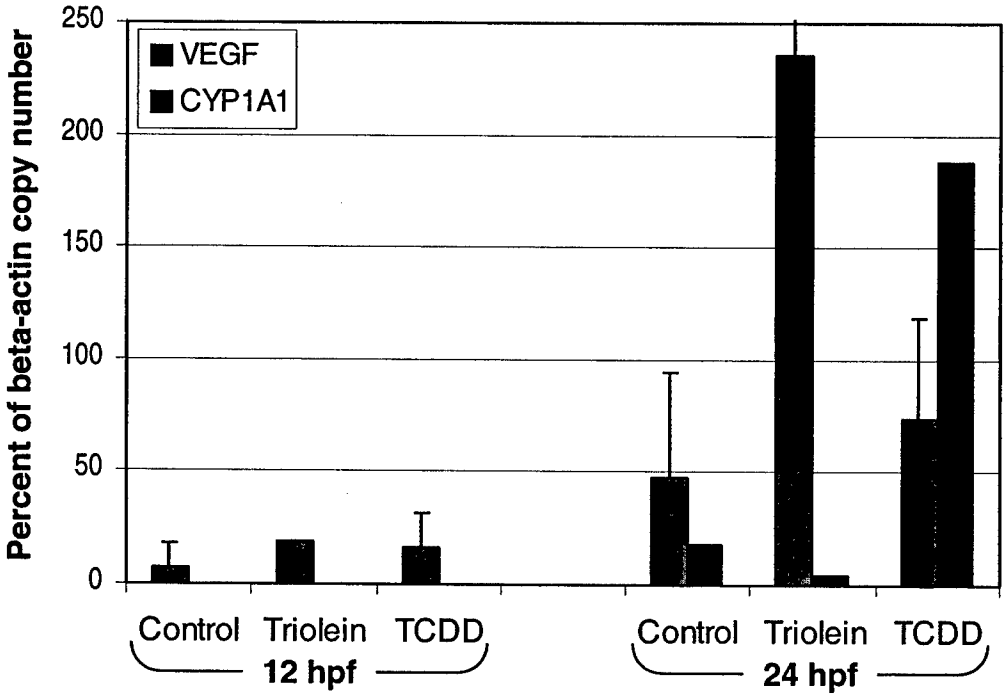

Copies target mRNA/ng total RNA
8.0E+05
7.0E+05
6.0E+05
5.0E+05
4.0E+05
3.0E+05
2.0E+05
1.0E+05
0.0E+00
VEGF
CYP1A1
Control
Triolein
TCDD
12 hpf
Control
Triolein
TCDD
24 hpf

Figure A.5 Mean VEGF and CYP1A mRNA levels expressed as a proportion of β-actin values.


Percent of beta-actin copy number
250
200
150
100
50
0
VEGF
CYP1A1
Control
Triolein
TCDD
12 hpf
Control
Triolein
TCDD
24 hpf

APPENDIX B.

Preliminary evaluation of cross-species hybridization efficiency

B.1 Introduction

Although still very young, the field of toxicogenomics is already making great strides in the areas of elucidating molecular mechanisms of toxicity and defining chemical-specific expression profiles [99, 100]. Ultimately, the goal of much of this work is the development of diagnostic and predictive biomarkers for pre-clinical, clinical, and environmental applications [101-103]. DNA microarrays are the primary tool being used in such work.

However, the large body of DNA sequence data needed to support microarray design severely limits the number of species for which microarrays are available. *Fundulus heteroclitus*, the mummichog or saltmarsh killifish, is a small marine fish that has been used extensively in both developmental biology and ecotoxicology. The *F. heteroclitus* genome is poorly characterized; less than 50 genes have been cloned and a genome project has only recently been undertaken. The situation is similar (or worse) for many environmentally and economically important fish species (e.g. *Salmo salmieri*, *Oncorhynchus mykiss*). Small-scale custom macroarrays have been used to investigate gene expression in certain environmental settings [113]. However, high-density arrays for *Fundulus* and other environmentally important species are several years away, at best.

There is evidence that microarrays constructed with material specific to one species can be used to assay gene expression in closely related species [217, 218]. Thus, it was of interest to determine whether zebrafish microarrays (Chapter 2) might be used to study gene expression in other fish species. To this end, we prepared labeled cDNA from both zebrafish and *F. heteroclitus* heart RNA, and compared the strength and patterns of hybridization to zebrafish cDNA microarrays. Preliminary analyses indicate that, with further optimization, cross-species hyridization may be an extremely informative tool.

B.2 Methods

mRNA from *Fundulus heteroclitus* adult heart tissue (Sibel Karchner) and total RNA from zebrafish adult heart tissue was used to generate amino-allyl post-labeled cDNA, according to previously described protocols (Chapter 3). Single-color hybridizations to

AH001 arrays were performed at 55°C, with all other conditions as previously described (Chapter 3).

B.3 Results

Side-by-side visual inspection of same-species and cross-species hybridizations revealed obvious similarities in patterns of relative signal strength among features (Figure B.1). To quantify this relationship, feature intensities from 3787 features on Cy3 hybridizations were compared directly (Figure B.2). In the vast majority of cases, same-species hybridization produced higher fluorescence intensity. Several hundred features with cross-species fluorescence intensities ≥2-fold higher than same-species intensities were separated from the main body of data. Each group showed moderate levels of correlation between same-species and cross-species fluorescence intensities ($R^2 = 0.59$ and 0.75).

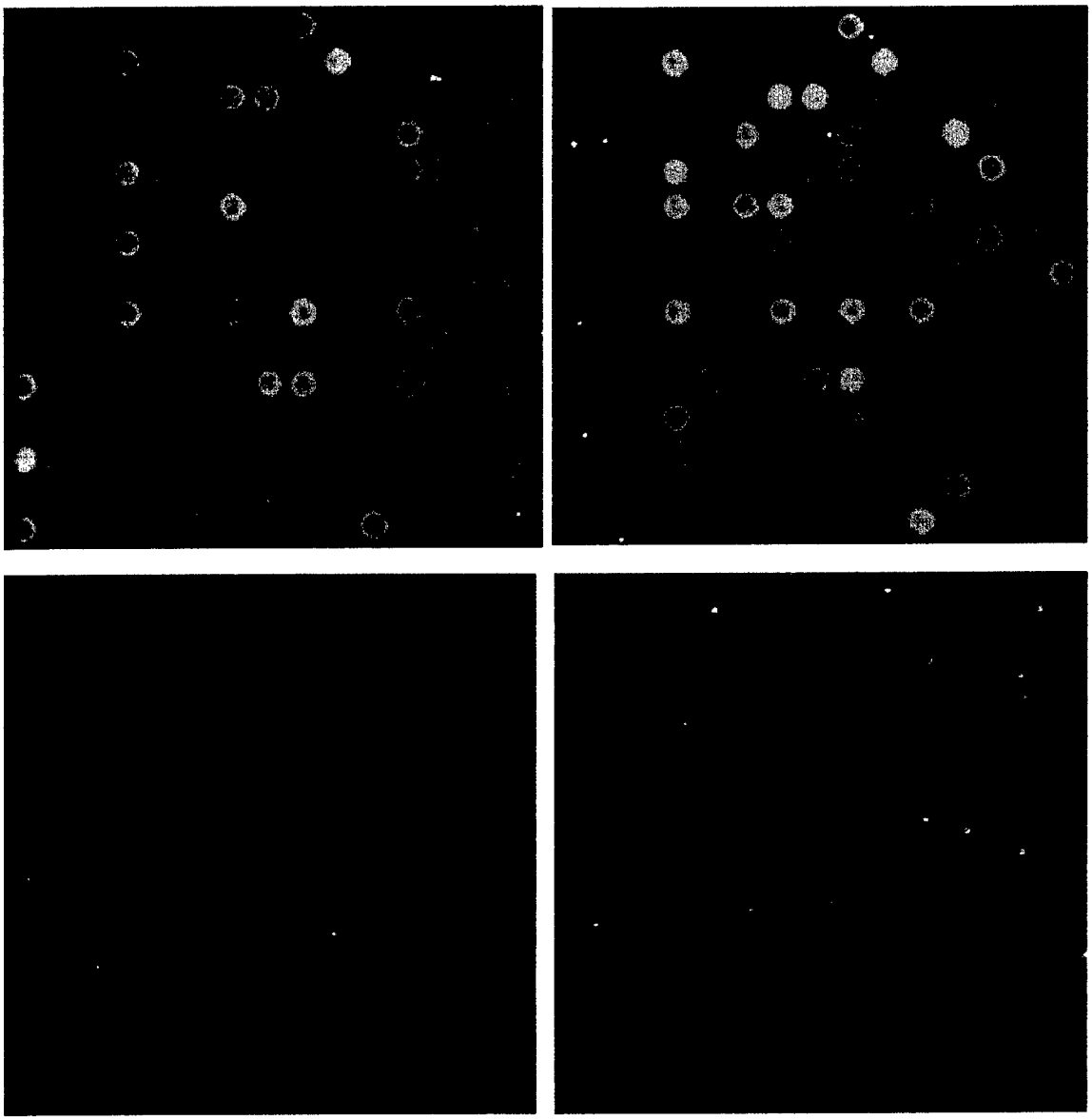
B.4 Discussion

These results suggest that, while less efficient than same-species hybridization, cross-species hybridization to zebrafish microarrays may be used to detect gene expression in fish species for which DNA arrays are not available. A general correlation between same-species and inter-species hybridization results was readily apparent upon inspection of either hybridization images or resulting numerical data. Similarly, results of hybridization of pig RNA to human microarray tracked closely with results from human-human hybridizations [218].

Neither this preliminary work nor published investigations of inter-species hybridization has adequately addressed the potential for non-specific hybridization. In the current case, significant outliers and only moderate support for a regression trendline fitted to the main body of data both suggest an unexplained source of variation affecting some subset of *Fundulus* genes. A high level of variation in cross-species results for 6% of arrayed human genes also suggested gene-specific artifacts [218]. Such variance might be reduced by increasing the stringency of cross-species hybridizations; further work is needed to determine an ideal hybridization temperature for use of *Fundulus*

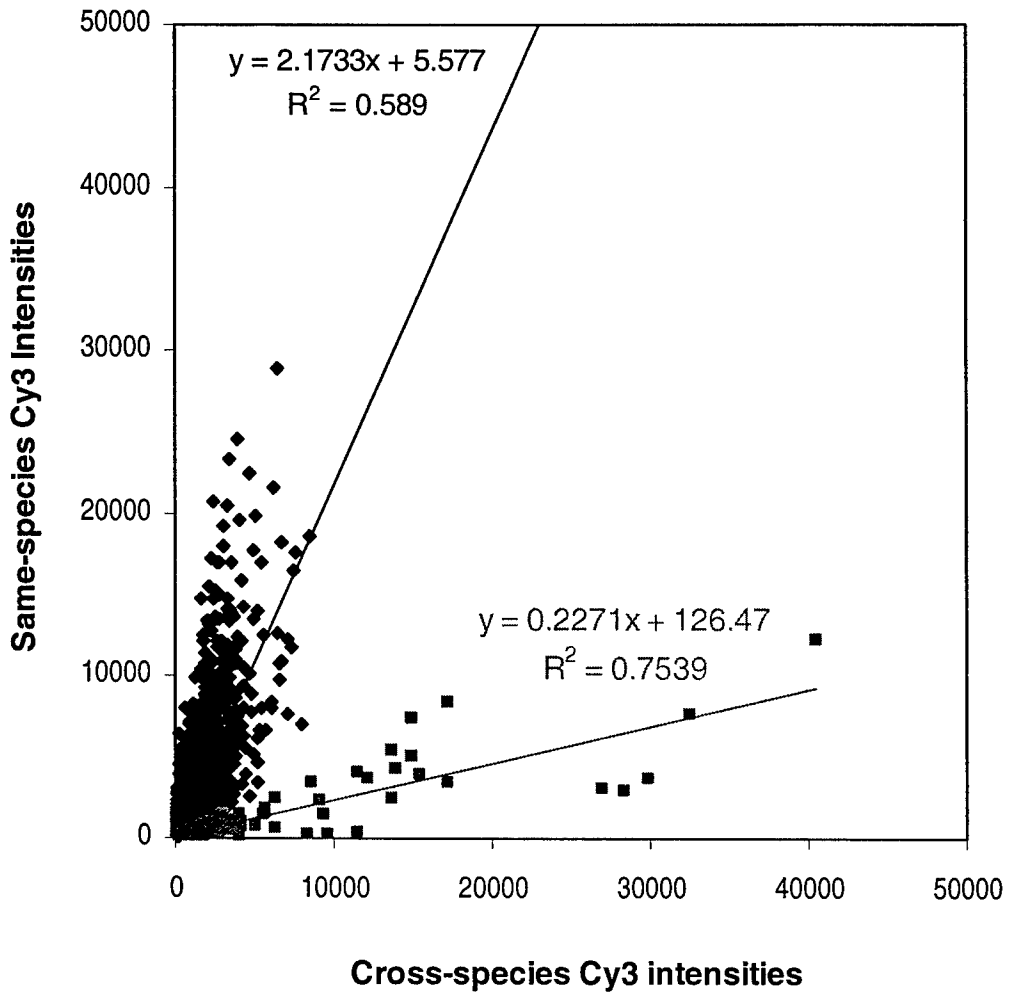
heteroclitus samples with zebrafish arrays. However, it is also possible that a large portion of variability is due to actual biological differences (i.e., differences in basal levels of expression of certain genes). Thus, it would be interesting to assay non-specific hybridization using individual *Fundulus* gene transcripts. Such work has yielded important information about specificity of arrayed cDNA probes [138], and might contribute to the development of general guidelines for conditions of cross-species hybridization.

Figure B.1 Representative quadrants from zebrafish (which arrays) hybridized with cDNA from zebrafish heart tissue (left panel) or from *Fundulus heteroclitus* heart tissue (right panel).



Danio rerio *Fundulus heteroclitus*

Figure B.2 Correlation between Cy3 feature intensities generated by hybridization of AH001 arrays with either zebrafish or *Fundulus heteroclitus* adult heart cDNA. 3787 features with intensities of at least 100 rfu on both hybridizations were compared. Features with cross-species fluorescence intensities at least 2-fold higher than same-species values were analyzed separately (grey squares).


y = 2.1733x + 5.577
R2 = 0.589
y = 0.2271x + 126.47
R2 = 0.7539
Same-species Cy3 Intensities
Cross-species Cy3 intensities
0
10000
20000
30000
40000
50000

APPENDIX C.

The role of cytochrome P450 1A in TCDD embryotoxicity: preliminary results

C.1 Introduction

Several lines of evidence have implicated cytochrome P450 1A (CYP1A) in the mechanism of TCDD toxicity. Induction of CYP1A enzymes by aromatic hydrocarbons was first reported more than thirty years ago [60, 61], and has since been shown to be strictly AHR-dependent [62, 63]. CYP1A induction co-localizes with target regions for TCDD toxicity, such as vascular endothelium, and follows similar dose-response curves as toxic end-points [26, 27, 68, 69]. Finally, blocking CYP1A enzymatic activity protects zebrafish embryos against circulatory dysfunction [22]. However, direct and conclusive proof of a role for CYP1A in processes of TCDD toxicity have been elusive.

Morpholino technology provides a rapid method for functional knock-down of specific protein expression in zebrafish. We have attempted to use morpholinos to knock down CYP1A expression and induction by TCDD. However, this effort has been confounded by the sporadic appearance of what is most likely an artifactual phenotype. As a result, focus has shifted to the analysis of CYP1A morphant embryos being generated by Dr. Hiroki Teraoka's laboratory.

C.2 Methods

Gene-specific morpholinos and a fluorescein-tagged standard control morpholino were obtained from GeneTools, LLC. One morpholino, referred to simply as ATG, was designed to span the translational start site of CYP1A:

5' – GGAAGAATAGTCAGAGCCATTGCTG – 3'

Another morpholino (I2) targeted the splice acceptor site at the boundary of intron 2 and exon 2:

5' – TAACCCACCCACCTTATCGAACGTA – 3'

A four-base mismatch, called $I2^{neg}$, served as a specific negative control:

5' - TAtCCCtCCCACCTTATgGAAgGTA

Two commercial preparations each of I2 and $I2^{neg}$, one incorporating a fluoroscein tag and one without, were used interchangeably in experimental work.

Stock solutions of 1 mM morpholino in 1x Danieau's were stored at 4°C and used to generate 50-500 μM working solutions. Morpholinos or buffer alone were injected into the cells of 1- to 4-cell zebrafish embryos. When using fluoroscein-tagged morpholinos, embryos were selected for evenly distributed high-level fluorescence at 18-24 hpf.

C.3 Results

The I2 morpholino produced a range of deletions in the 3' portion of intron 2; truncated transcripts were detected by RT-PCR, cloned and sequenced (data not shown). The primary lesion was a 30 bp deletion (811-841 bp), presumably resulting in an in-frame deletion of amino acids 271-280. No such lesions were detected in CYP1A transcripts from embryos injected with $I2^{neg}$ morpholino.

Experiments involving the ATG-targeted morpholino revealed an abnormal phenotype consisting of reduced cranial size, skeletal malformations (i.e., twisted tail), and disruption of peripheral circulation (Figure C.1). Both I2 and, to a lesser degree, $I2^{neg}$ morpholinos produced an indistinguishable phenotype (Figure C.1). Over the course of five additional experiments, occurrence of this phenotype was sporadic and did not appear to bear any relationship to the source (i.e., different preparations) or concentration of morpholino used. However, this phenotype was never observed in embryos injected with either 1x Danieau's buffer or GeneTools' standard control (data not shown).

C.4 Discussion

The fact that two morpholinos targeting disparate regions of the CYP1A gene produced the same abnormal phenotype, while buffer and standard control morpholino did not, would tend to suggest that this is a specific effect. However, observations of the same phenotype in embryos injected with the specific negative control morpholino, $I2^{neg}$, suggest otherwise. It is clear that $I2^{neg}$ has no effect on CYP1A transcript processing, and thus, probably does not interfere with functional protein expression. Thus, the observed phenotype cannot be the result of specific knock-down of CYP1A.

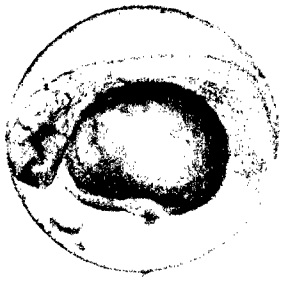
One possible alternative is that these morpholinos are interacting with another undefined cytochrome P450 gene. At this time, relatively little is known about CYP family genes in zebrafish. It would be extremely interesting to identify genes that, based on sequence similarity, might be interacting with the morpholinos used here.

Members of Dr. Hiroki Teraoka's laboratory have recently published their findings that a different CYP1A morpholino protects against TCDD circulatory impacts without causing any confounding effects [54]. In order to address the initial question of interest, we have begun to work with Dr. Teraoka to characterize gene expression in untreated and TCDD-exposed CYP1A morphant embryos using AH002A/B arrays.

Figure C.1 CYP1A morphant phenotype, as observed at 50 hpf. This phenotype was produced sporadically by morpholinos targeted against either the boundary of intron 2 and exon 2 (a), or the translational start site (b).

(a) 500μM I2

(b) 500μM ATG



LITERATURE CITED

1. Safe, S. (1993) Toxicology, structure-function relationship, and human and environmental health impacts of polychlorinated biphenyls: progress and problems. *Environ Health Perspect* **100**, 259-68.

2. Weaver, G. (1984) PCB contamination in and around New Bedford, Massachusetts. *Environ Sci Technol* **18**(1), 22A-27A.

3. (1999) Polychlorinated dibenzo-*p*-dioxins and related compounds update: Impact on fish advisories, pp. 6. U.S. Environmental Protection Agency.

4. Tanabe, S. (1985) DISTRIBUTION BEHAVIOR AND FATE OF POLYCHLORINATED BIPHENYLS IN THE MARINE ENVIRONMENT: LECTURE BY THE MEMBER AWARDED THE OKADA PRIZE OF THE OCEANOGRAPHICAL SOCIETY OF JAPAN FOR 1985. *J Oceanogr Soc Jpn* **41**(5), 358-370.

5. Safe, S., Safe, L. and Mullin, M. (1985) Polychlorinated biphenyls (PCBs) -- congener-specific analysis of a commercial mixture and a human milk extract. *J AGric Food Chem* **33**, 24-29.

6. Wu, W. Z., Zhang, Q. H., Schramm, K. W., Xu, Y. and Kettrup, A. (2000) Distribution, transformation, and long-term accumulation of polychlorinated dibenzo-p-dioxins and dibenzofurans in different tissues of fish and piscivorous birds. *Ecotoxicol Environ Saf* **46**(3), 252-7.

7. NRC. (2000) *Hormonally Active Agents in the Environment.* National Academy Press.

8. NRC. (2000) *Scientific Frontiers in Developmental Toxicology and Risk Assessment.* National Academies Press.

9. Elonen, G. E., Spehar, R. L., Holcombe, G. W., Johnson, R. D., Fernandez, J. D., Erickson, R. J., Tietge, J. E. and Cook, P. M. (1998) Comparative toxicity of 2,3,7,8-tetrachlorodibenzo-*p*-dioxin to seven freshwater fish species during early life-stage development. *Environ Toxicol Chem* **17**(3), 472-483.

10. Henry, T. R., Spitsbergen, J. M., Hornung, M. W., Abnet, C. C. and Peterson, R. E. (1997) Early life stage toxicity of 2,3,7,8-tetrachlorodibenzo-p-dioxin in zebrafish (Danio rerio). *Toxicol Appl Pharmacol* **142**(1), 56-68.

11. Helder, T. (1981) Effects of 2,3,7,8-tetrachlorodibenzo-dioxin (TCDD) on early life stages of rainbow trout (Salmo gairdneri, Richardson). *Toxicology* **19**(2), 101-12.

12. Spitsbergen, J. M., Walker, M. K., Olson, J. R. and Peterson, R. E. (1991) Pathologic alterations in early life stages of lake trout, *Salvelinus namaycush*, exposed to 2,3,7,8-tetrachlorodibenzo-*p*-dioxin as fertilized eggs. *Aquat Toxicol* **19**, 41-72.

13. Walker, M. K. and Peterson, R. E. (1994) Toxicity of 2,3,7,8-tetrachlorodibenzo-*p*-dioxin to brook trout (*Salvelinus fontinalis*) during early development. *Environ Toxicol Chem* **13**(5), 817-820.

14. Wisk, J. D. and Cooper, K. R. (1990) The stage specific toxicity of 2,3,7,8-tetrachlorodibenzo-*p*-dioxin in embryos of the Japanese medaka (*Oryzias latipes*). *Environ Toxicol Chem* **9**, 1159-1169.

15. Toomey, B. H., Bello, S., Hahn, M. E., Cantrell, S., Wright, P., Tillitt, D. E. and Di Giulio, R. T. (2001) 2,3,7,8-Tetrachlorodibenzo-p-dioxin induces apoptotic cell death and cytochrome P4501A expression in developing Fundulus heteroclitus embryos. *Aquat Toxicol* **53**(2), 127-38.

16. Brunstrom, B., Broman, D. and Naf, C. (1991) Toxicity and EROD-inducing potency of 24 polycyclic aromatic hydrocarbons (PAHs) in chick embryos. *Arch Toxicol* **65**(6), 485-9.

17. Brunstrom, B. (1988) Sensitivity of embryos from duck, goose, herring gull, and various chicken breeds to 3,3',4,4'-tetrachlorobiphenyl. *Poult Sci* **67**(1), 52-7.

18. Brunstrom, B. and Lund, J. (1988) Differences between chick and turkey embryos in sensitivity to 3,3',4,4'-tetrachloro-biphenyl and in concentration/affinity of the hepatic receptor for 2,3,7,8-tetrachlorodibenzo-p-dioxin. *Comp Biochem Physiol C* **91**(2), 507-12.

19. Brunstrom, B. and Darnerud, P. O. (1983) Toxicity and distribution in chick embryos of 3,3',4,4'- tetrachlorobiphenyl injected into the eggs. *Toxicology* **27**(2), 103-10.

20. Powell, D. C., Aulerich, R. J., Meadows, J. C., Tillitt, D. E., Giesy, J. P., Stromberg, K. L. and Bursian, S. J. (1996) Effects of 3,3',4,4',5-pentachlorobiphenyl (PCB 126) and 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD) injected into the yolks of chicken (Gallus domesticus) eggs prior to incubation. *Arch Environ Contam Toxicol* **31**(3), 404-9.

21. Poland, A. and Knutson, J. C. (1982) 2,3,7,8-Tetrachlorodibenzo-*p*-dioxin and related halogenated aromatic hydrocarbons: examination of the mechanism of toxicity. *Annu Rev Pharmacol Toxicol* **22**, 517-554.

22. Dong, W., Teraoka, H., Yamazaki, K., Tsukiyama, S., Imani, S., Imagawa, T., Stegeman, J. J., Peterson, R. E. and Hiraga, T. (2002) 2,3,7,8-tetrachlorodibenzo-p-dioxin toxicity in the zebrafish embryo: local circulation failure in the dorsal midbrain is associated with increased apoptosis. *Toxicol Sci* **69**(1), 191-201.

23. Teraoka, H., Dong, W., Ogawa, S., Tsukiyama, S., Okuhara, Y., Niiyama, M., Ueno, N., Peterson, R. E. and Hiraga, T. (2002) 2,3,7,8-Tetrachlorodibenzo-p-dioxin toxicity in the zebrafish embryo: altered regional blood flow and impaired lower jaw development. *Toxicol Sci* **65**(2), 192-9.

24. Belair, C. D., Peterson, R. E. and Heideman, W. (2001) Disruption of erythropoiesis by dioxin in the zebrafish. *Dev Dyn* **222**(4), 581-94.

25. Hornung, M. W., Spitsbergen, J. M. and Peterson, R. E. (1999) 2,3,7,8-Tetrachlorodibenzo-p-dioxin alters cardiovascular and craniofacial development and function in sac fry of rainbow trout (Oncorhynchus mykiss). *Toxicol Sci* **47**(1), 40-51.

26. Cantrell, S. M., Lutz, L. H., Tillitt, D. E. and Hannink, M. (1996) Embryotoxicity of 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD): the embryonic vasculature is a physiological target for TCDD-induced DNA damage and apoptotic cell death in Medaka (Orizias latipes). *Toxicol Appl Pharmacol* **141**(1), 23-34.

27. Guiney, P. D., Smolowitz, R. M., Peterson, R. E. and Stegeman, J. J. (1997) Correlation of 2,3,7,8-tetrachlorodibenzo-p-dioxin induction of cytochrome P4501A in vascular endothelium with toxicity in early life stages of lake trout. *Toxicol Appl Pharmacol* **143**(2), 256-73.

28. Guiney, P. D., Walker, M. K., Spitsbergen, J. M. and Peterson, R. E. (2000) Hemodynamic dysfunction and cytochrome P4501A mRNA expression induced by 2,3,7,8-tetrachlorodibenzo-p-dioxin during embryonic stages of lake trout development. *Toxicol Appl Pharmacol* **168**(1), 1-14.

29. Chen, J. N., Haffter, P., Odenthal, J., Vogelsang, E., Brand, M., van Eeden, F. J., Furutani-Seiki, M., Granato, M., Hammerschmidt, M., Heisenberg, C. P., Jiang, Y. J., Kane, D. A., Kelsh, R. N., Mullins, M. C. and Nusslein-Volhard, C. (1996) Mutations affecting the cardiovascular system and other internal organs in zebrafish. *Development* **123**, 293-302.

30. Stainier, D. Y., Fouquet, B., Chen, J. N., Warren, K. S., Weinstein, B. M., Meiler, S. E., Mohideen, M. A., Neuhauss, S. C., Solnica-Krezel, L., Schier, A. F., Zwartkruis, F., Stemple, D. L., Malicki, J., Driever, W. and Fishman, M. C. (1996) Mutations affecting the formation and function of the cardiovascular system in the zebrafish embryo. *Development* **123**, 285-92.

31. Bello, S., Heideman, W. and Peterson, R. (2002) Dioxin impairs osmoregulation in the zebrafish larva. *The Toxicologist* **66**(1-S), Abstract #826.

32. Walker, M. K. and Catron, T. F. (2000) Characterization of cardiotoxicity induced by 2,3,7, 8-tetrachlorodibenzo-p-dioxin and related chemicals during early chick embryo development. *Toxicol Appl Pharmacol* **167**(3), 210-21.

33. Ivnitski, I., Elmaoued, R. and Walker, M. K. (2001) 2,3,7,8-Tetrachlorodibenzo-*p*-dioxin (TCDD) Inhibition of Coronary Development is Preceded by a Decrease in Myocyte Proliferation and an Increase in Cardiac Apoptosis. *Teratol* **64**, 201-212.

34. Fan, L., Ovadia, M., Friedman, D. M. and Rifkind, A. B. (2000) Ventricular preexcitation sensitive to flecainide in late stage chick embryo ECGs: 2,3,7,8-tetrachlorodibenzo-p-dioxin impairs inotropic but not chronotropic or dromotropic responses to isoproterenol and confers resistance to flecainide. *Toxicol Appl Pharmacol* **166**(1), 43-50.

35. Sedmera, D., Pexieder, T., Vuillemin, M., Thompson, R. P. and Anderson, R. H. (2000) Developmental patterning of the myocardium. *Anat Rec* **258**, 319-337.

36. Hankinson, O. (1995) The aryl hydrocarbon receptor complex. *Annu Rev Pharmacol Toxicol* **35**, 307-40.

37. Hahn, M. E. (2002) Aryl hydrocarbon receptors: diversity and evolution. *Chem Biol Interact* **141**(1-2), 131-60.

38. Hahn, M. E. (1998) The aryl hydrocarbon receptor: a comparative perspective. *Comp Biochem Physiol C Pharmacol Toxicol Endocrinol* **121**(1- 3), 23-53.

39. Walker, M. K., Pollenz, R. S. and Smith, S. M. (1997) Expression of the aryl hydrocarbon receptor (AhR) and AhR nuclear translocator during chick cardiogenesis is consistent with 2,3,7,8-tetrachlorodibenzo-p-dioxin-induced heart defects. *Toxicol Appl Pharmacol* **143**(2), 407-19.

40. Walker, M. K., Heid, S. E., Smith, S. M. and Swanson, H. I. (2000) Molecular characterization and developmental expression of the aryl hydrocarbon receptor

from the chick embryo. *Comp Biochem Physiol C Toxicol Pharmacol* **126**(3), 305-19.

41. Karchner, S. I., Powell, W. H. and Hahn, M. E. (1999) Identification and functional characterization of two highly divergent aryl hydrocarbon receptors (AHR1 and AHR2) in the teleost Fundulus heteroclitus. Evidence for a novel subfamily of ligand-binding basic helix loop helix-Per-ARNT-Sim (bHLH-PAS) factors. *J Biol Chem* **274**(47), 33814-24.

42. Andreasen, E. A., Spitsbergen, J. M., Tanguay, R. L., Stegeman, J. J., Heideman, W. and Peterson, R. E. (2002) Tissue-specific expression of AHR2, ARNT2, and CYP1A in zebrafish embryos and larvae: effects of developmental stage and 2,3,7,8-tetrachlorodibenzo-p-dioxin exposure. *Toxicol Sci* **68**(2), 403-19.

43. Lahvis, G. P., Lindell, S. L., Thomas, R. S., McCuskey, R. S., Murphy, C., Glover, E., Bentz, M., Southard, J. and Bradfield, C. A. (2000) Portosystemic shunting and persistent fetal vascular structures in aryl hydrocarbon receptor-deficient mice. *Proc Natl Acad Sci U S A* **97**(19), 10442-7.

44. Lund, A., Kanagy, N. and Walker, M. (2002) Aryl hydrocarbon receptor (AHR) null mice exhibit hypertension and increased plasma endothelin levels. *The Toxicologist* **66**(1-S), Abstract #37.

45. Thackaberry, E. A., Gabaldon, D. M., Walker, M. K. and Smith, S. M. (2002) Aryl hydrocarbon receptor null mice develop cardiac hypertrophy and increased hpoxia-inducible factor-1α in the absence of cardiac hypoxia. *Cardiovasc Toxicol* **2**(4), 263-274.

46. Mehrabi, M. R., Steiner, G. E., Dellinger, C., Kofler, A., Schaufler, K., Tamaddon, F., Plesch, K., Ekmekcioglu, C., Maurer, G., Glogar, H. D. and Thalhammer, T. (2002) The arylhydrocarbon receptor (AhR), but not the AhR-nuclear translocator (ARNT), is increased in hearts of patients with cardiomyopathy. *Virchows Arch* **441**(5), 481-489.

47. Heid, S. E., Walker, M. K. and Swanson, H. I. (2001) Correlation of cardiotoxicity mediated by halogenated aromatic hydrocarbons to aryl hydrocarbon receptor activation. *Toxicol Sci* **61**(1), 187-96.

48. Pohjanvirta, R., Viluksela, M., Tuomisto, J. T., Unkila, M., Karasinska, J., Franc, M. A., Holowenko, M., Giannone, J. V., Harper, P. A., Tuomisto, J. and Okey, A. B. (1999) Physicochemical differences in the AH receptors of the most TCDD-susceptible and the most TCDD-resistant rat strains. *Toxicol Appl Pharmacol* **155**(1), 82-95.

49. Hestermann, E. V., Stegeman, J. J. and Hahn, M. E. (2000) Relative contributions of affinity and intrinsic efficacy to aryl hydrocarbon receptor ligand potency. *Toxicol Appl Pharmacol* **168**(2), 160-172.

50. Moriguchi, T., Motohashi, H., Hosoya, T., Nakajima, O., Takahashi, S., Ohsako, S., Aoki, Y., Nishimura, N., Tohyama, C., Fujii-Kuriyama, Y. and Yamamoto, M. (2003) Distinct response to dioxin in an arylhydrocarbon receptor (AHR)-humanized mouse. *Proc Natl Acad Sci U S A* **100**(10), 5652-7.

51. Powell, W. H., Bright, R., Bello, S. M. and Hahn, M. E. (2000) Developmental and tissue-specific expression of AHR1, AHR2, and ARNT2 in dioxin-sensitive and -resistant populations of the marine fish Fundulus heteroclitus. *Toxicol Sci* **57**(2), 229-39.

52. Bello, S. M., Franks, D. G., Stegeman, J. J. and Hahn, M. E. (2001) Acquired resistance to Ah receptor agonists in a population of Atlantic killifish (Fundulus heteroclitus) inhabiting a marine superfund site: in vivo and in vitro studies on the inducibility of xenobiotic metabolizing enzymes. *Toxicol Sci* **60**(1), 77-91.

53. Dong, W., Teraoka, H., Kondo, S. and Hiraga, T. (2001) 2, 3, 7, 8-tetrachlorodibenzo-p-dioxin induces apoptosis in the dorsal midbrain of zebrafish embryos by activation of arylhydrocarbon receptor. *Neurosci Lett* **303**(3), 169-72.

54. Teraoka, H., Dong, W., Tsujimoto, Y., Iwasa, H., Endoh, D., Ueno, N., Stegeman, J. J., Peterson, R. E. and Hiraga, T. (2003) Induction of cytochrome P450 1A is required for circulation failure and edema by 2,3,7,8-tetrachlorodibenzo-p-dioxin in zebrafish. *Biochem Biophys Res Commun* **304**(2), 223-8.

55. Fernandez-Salguero, P. M., Hilbert, D. M., Rudikoff, S., Ward, J. M. and Gonzalez, F. J. (1996) Aryl-hydrocarbon receptor-deficient mice are resistant to 2,3,7,8-tetrachlorodibenzo-p-dioxin-induced toxicity. *Toxicol Appl Pharmacol* **140**(1), 173-9.

56. Mimura, J., Yamashita, K., Nakamura, K., Morita, M., Takagi, T. N., Nakao, K., Ema, M., Sogawa, K., Yasuda, M., Katsuki, M. and Fujii-Kuriyama, Y. (1997) Loss of teratogenic response to 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD) in mice lacking the Ah (dioxin) receptor. *Genes Cells* **2**(10), 645-54.

57. Staples, J. E., Murante, F. G., Fiore, N. C., Gasiewicz, T. A. and Silverstone, A. E. (1998) Thymic alterations induced by 2,3,7,8-tetrachlorodibenzo-p-dioxin are strictly dependent on aryl hydrocarbon receptor activation in hemopoietic cells. *J Immunol* **160**(8), 3844-54.

58. Peters, J. M., Narotsky, M. G., Elizondo, G., Fernandez-Salguero, P. M., Gonzalez, F. J. and Abbott, B. D. (1999) Amelioration of TCDD-induced teratogenesis in aryl hydrocarbon receptor (AhR)-null mice. *Toxicol Sci* **47**(1), 86-92.

59. Lin, T. M., Ko, K., Moore, R. W., Simanainen, U., Oberley, T. D. and Peterson, R. E. (2002) Effects of aryl hydrocarbon receptor null mutation and in utero and lactational 2,3,7,8-tetrachlorodibenzo-p-dioxin exposure on prostate and seminal vesicle development in C57BL/6 mice. *Toxicol Sci* **68**(2), 479-87.

60. Nebert, D. W. and Gelboin, H. V. (1969) The in vivo and in vitro induction of aryl hydrocarbon hydroxylase in mammalian cells of different species, tissues, strains, and developmental and hormonal states. *Arch Biochem Biophys* **134**(1), 76-89.

61. Nebert, D. W. (1969) Changes in aryl hydrocarbon hydroxylase activity and microsomal P450 during polycyclic hydrocarbon treatment of mammalian cells in culture. *Biochem Biophys Res Commun* **36**(6), 885-90.

62. Zaher, H., Yang, T. J., Gelboin, H. V., Fernandez-Salguero, P. and Gonzalez, F. J. (1998) Effect of phenobarbital on hepatic CYP1A1 and CYP1A2 in the Ahr-null mouse. *Biochem Pharmacol* **55**(2), 235-8.

63. Shimada, T., Inoue, K., Suzuki, Y., Kawai, T., Azuma, E., Nakajima, T., Shindo, M., Kurose, K., Sugie, A., Yamagishi, Y., Fujii-Kuriyama, Y. and Hashimoto, M. (2002) Arylhydrocarbon receptor-dependent induction of liver and lung cytochromes P450 1A1, 1A2, and 1B1 by polycyclic aromatic hydrocarbons and polychlorinated biphenyls in genetically engineered C57BL/6J mice. *Carcinogenesis* **23**(7), 1199-207.

64. Schlezinger, J. J., White, R. D. and Stegeman, J. J. (1999) Oxidative inactivation of cytochrome P-450 1A (CYP1A) stimulated by 3,3',4,4'-tetrachlorobiphenyl: production of reactive oxygen by vertebrate CYP1As. *Mol Pharmacol* **56**(3), 588-97.

65. Rifkind, A. B., Kanetoshi, A., Orlinick, J., Capdevila, J. H. and Lee, C. (1994) Purification and biochemical characterization of two major cytochrome P-450 isoforms induced by 2,3,7,8-tetrachlorodibenzo-p-dioxin in chick embryo liver. *J Biol Chem* **269**(5), 3387-96.

66. Gilday, D., Gannon, M., Yutzey, K., Bader, D. and Rifkind, A. B. (1996) Molecular cloning and expression of two novel avian cytochrome P450 1A enzymes induced by 2,3,7,8-tetrachlorodibenzo-p-dioxin. *J Biol Chem* **271**(51), 33054-9.

67. Gilday, D., Bellward, G. D., Sanderson, J. T., Janz, D. M. and Rifkind, A. B. (1998) 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD) induces hepatic cytochrome P450-dependent arachidonic acid epoxygenation in diverse avian orders: regioisomer selectivity and immunochemical comparison of the TCDD-induced P450s to CYP1A4 and 1A5. *Toxicol Appl Pharmacol* **150**(1), 106-16.

68. Cantrell, S. M., Joy-Schlezinger, J., Stegeman, J. J., Tillitt, D. E. and Hannink, M. (1998) Correlation of 2,3,7,8-tetrachlorodibenzo-p-dioxin-induced apoptotic cell death in the embryonic vasculature with embryotoxicity. *Toxicol Appl Pharmacol* **148**(1), 24-34.

69. Gannon, M., Gilday, D. and Rifkind, A. B. (2000) TCDD induces CYP1A4 and CYP1A5 in chick liver and kidney and only CYP1A4, an enzyme lacking arachidonic acid epoxygenase activity, in myocardium and vascular endothelium. *Toxicol Appl Pharmacol* **164**(1), 24-37.

70. Senft, A. P., Dalton, T. P., Nebert, D. W., Genter, M. B., Puga, A. and Shertzer, H. G. (2001) Dioxin-stimulated production of reactive oxygen in liver mitochondria from CYP1A1+/+ and CYP1A1-/- mice. *The Toxicologist* **60**(1-S), Abstract #1282.

71. Stohs, S. J. (1990) Oxidative stress induced by 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD). *Free Radic Biol Med* **9**(1), 79-90.

72. Stohs, S. J., Shara, M. A., Alsharif, N. Z., Wahba, Z. Z. and al-Bayati, Z. A. (1990) 2,3,7,8-Tetrachlorodibenzo-p-dioxin-induced oxidative stress in female rats. *Toxicol Appl Pharmacol* **106**(1), 126-35.

73. Stohs, S. J., Alsharif, N. Z., Shara, M. A., al-Bayati, Z. A. and Wahba, Z. Z. (1991) Evidence for the induction of an oxidative stress in rat hepatic mitochondria by 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD). *Adv Exp Med Biol* **283**, 827-31.

74. Alsharif, N. Z., Lawson, T. and Stohs, S. J. (1994) Oxidative stress induced by 2,3,7,8-tetrachlorodibenzo-p-dioxin is mediated by the aryl hydrocarbon (Ah) receptor complex. *Toxicology* **92**(1-3), 39-51.

75. Alsharif, N. Z., Schlueter, W. J. and Stohs, S. J. (1994) Stimulation of NADPH-dependent reactive oxygen species formation and DNA damage by 2,3,7,8-tetrachlorodibenzo-p-dioxin in rat peritoneal lavage cells. *Arch Environ Contam Toxicol* **26**(3), 392-7.

76. Tritscher, A. M., Seacat, A. M., Yager, J. D., Groopman, J. D., Miller, B. D., Bell, D., Sutter, T. R. and Lucier, G. W. (1996) Increased oxidative DNA damage in livers of 2,3,7,8-tetrachlorodibenzo- p-dioxin treated intact but not ovariectomized rats. *Cancer Lett* **98**(2), 219-25.

77. Hassoun, E. A., Wilt, S. C., Devito, M. J., Van Birgelen, A., Alsharif, N. Z., Birnbaum, L. S. and Stohs, S. J. (1998) Induction of oxidative stress in brain tissues of mice after subchronic exposure to 2,3,7,8-tetrachlorodibenzo-p-dioxin. *Toxicol Sci* **42**(1), 23-7.

78. Shertzer, H. G., Nebert, D. W., Puga, A., Ary, M., Sonntag, D., Dixon, K., Robinson, L. J., Cianciolo, E. and Dalton, T. P. (1998) Dioxin causes a sustained oxidative stress response in the mouse. *Biochem Biophys Res Commun* **253**(1), 44-8.

79. Slezak, B. P., Hatch, G. E., DeVito, M. J., Diliberto, J. J., Slade, R., Crissman, K., Hassoun, E. and Birnbaum, L. S. (2000) Oxidative stress in female B6C3F1 mice following acute and subchronic exposure to 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD). *Toxicol Sci* **54**(2), 390-8.

80. Hassoun, E. A., Li, F., Abushaban, A. and Stohs, S. J. (2001) Production of superoxide anion, lipid peroxidation and DNA damage in the hepatic and brain tissues of rats after subchronic exposure to mixtures of TCDD and its congeners. *J Appl Toxicol* **21**(3), 211-9.

81. Wyde, M. E., Wong, V. A., Kim, A. H., Lucier, G. W. and Walker, N. J. (2001) Induction of hepatic 8-oxo-deoxyguanosine adducts by 2,3,7,8-tetrachlorodibenzo-p-dioxin in Sprague-Dawley rats is female-specific and estrogen-dependent. *Chem Res Toxicol* **14**(7), 849-55.

82. Latchoumycandane, C., Chitra, C. and Mathur, P. (2002) Induction of oxidative stress in rat epididymal sperm after exposure to 2,3,7,8-tetrachlorodibenzo-p-dioxin. *Arch Toxicol* **76**(2), 113-8.

83. Slezak, B. P., Hamm, J. T., Reyna, J., Hurst, C. H. and Birnbaum, L. S. (2002) TCDD-mediated oxidative stress in male rat pups following perinatal exposure. *J Biochem Mol Toxicol* **16**(2), 49-52.

84. Hassoun, E. A., Wang, H., Abushaban, A. and Stohs, S. J. (2002) Induction of oxidative stress in the tissues of rats after chronic exposure to TCDD, 2,3,4,7,8-pentachlorodibenzofuran, and 3,3',4,4',5- pentachlorobiphenyl. *J Toxicol Environ Health A* **65**(12), 825-42.

85. Latchoumycandane, C., Chitra, K. C. and Mathur, P. P. (2003) 2,3,7,8-Tetrachlorodibenzo- p-dioxin (TCDD) induces oxidative stress in the epididymis and epididymal sperm of adult rats. *Arch Toxicol* **77**(5), 280-4.

86. Hassoun, E. A., Walter, A. C., Alsharif, N. Z. and Stohs, S. J. (1997) Modulation of TCDD-induced fetotoxicity and oxidative stress in embryonic and placental tissues of C57BL/6J mice by vitamin E succinate and ellagic acid. *Toxicology* **124**(1), 27-37.

87. Latchoumycandane, C. and Mathur, P. P. (2002) Effects of vitamin E on reactive oxygen species-mediated 2,3,7,8- tetrachlorodi-benzo-p-dioxin toxicity in rat testis. *J Appl Toxicol* **22**(5), 345-51.

88. Schlezinger, J. J., Keller, J., Verbrugge, L. A. and Stegeman, J. J. (2000) 3,3',4,4'-Tetrachlorobiphenyl oxidation in fish, bird and reptile species: relationship to cytochrome P450 1A inactivation and reactive oxygen production. *Comp Biochem Physiol C Toxicol Pharmacol* **125**(3), 273-86.

89. Park, H. (1999) Aromatic hydrocarbon nuclear translocator as a common component for the hypoxia- and dioxin-induced gene expression. *Mol Cells* **9**(2), 172-8.

90. Chan, W. K., Yao, G., Gu, Y. Z. and Bradfield, C. A. (1999) Cross-talk between the aryl hydrocarbon receptor and hypoxia inducible factor signaling pathways. Demonstration of competition and compensation. *J Biol Chem* **274**(17), 12115-23.

91. Gassmann, M., Kvietikova, I., Rolfs, A. and Wenger, R. H. (1997) Oxygen- and dioxin-regulated gene expression in mouse hepatoma cells. *Kidney Int* **51**(2), 567-74.

92. Pollenz, R. S., Davarinos, N. A. and Shearer, T. P. (1999) Analysis of aryl hydrocarbon receptor-mediated signaling during physiological hypoxia reveals lack of competition for the aryl hydrocarbon nuclear translocator transcription factor. *Mol Pharmacol* **56**(6), 1127-37.

93. Puga, A., Maier, A. and Medvedovic, M. (2000) The transcriptional signature of dioxin in human hepatoma HepG2 cells. *Biochem Pharmacol* **60**(8), 1129-42.

94. Martinez, J. M., Afshari, C. A., Bushel, P. R., Masuda, A., Takahashi, T. and Walker, N. J. (2002) Differential toxicogenomic responses to 2,3,7,8-tetrachlorodibenzo-p- dioxin in malignant and nonmalignant human airway epithelial cells. *Toxicol Sci* **69**(2), 409-23.

95. Zeytun, A., McKallip, R. J., Fisher, M., Camacho, I., Nagarkatti, M. and Nagarkatti, P. S. (2002) Analysis of 2,3,7,8-tetrachlorodibenzo-p-dioxin-induced gene expression profile in vivo using pathway-specific cDNA arrays. *Toxicology* **178**(3), 241-60.

96. Nuwaysir, E. F., Bittner, M., Trent, J., Barrett, J. C. and Afshari, C. A. (1999) Microarrays and toxicology: the advent of toxicogenomics. *Mol Carcinog* **24**(3), 153-9.

97. Selkirk, J. and Tennant, R. W. (2002) *Gordon Conference on Toxicogenomics* **Available at *http://www.grc.uri.edu/programs/2003/toxico.htm***.

98. Ballatori, N., Boyer, J. L. and Rockett, J. C. (2003) Exploiting genome data to understand the function, regulation, and evolutionary origins of toxicologically relevant genes sequences. *Environ Health Perspect* **111**(6), 871-5.

99. Waring, J. F., Jolly, R. A., Ciurlionis, R., Lum, P. Y., Praestgaard, J. T., Morfitt, D. C., Buratto, B., Roberts, C., Schadt, E. and Ulrich, R. G. (2001) Clustering of hepatotoxins based on mechanism of toxicity using gene expression profiles. *Toxicol Appl Pharmacol* **175**, 28-42.

100. Hamadeh, H. K., Bushel, P. R., Jayadev, S., Martin, K., DiSorbo, O., Sieber, S., Bennett, L., Tennant, R., Stoll, R., Barrett, J. C., Blanchard, K., Paules, R. S. and Afshari, C. A. (2002) Gene expression analysis reveals chemical-specific profiles. *Toxicol Sci* **67**(2), 219-31.

101. Thomas, R. S., Rank, D. R., Penn, S. G., Zastrow, G. M., Hayes, K. R., Pande, K., Glover, E., Silander, T., Craven, M. W., Reddy, J. K., Jovanovich, S. B. and Bradfield, C. A. (2001) Identification of toxicologically predictive gene sets using cDNA microarrays. *Mol Pharmacol* **60**(6), 1189-1194.

102. Hamadeh, H. K., Bushel, P. R., Jayadev, S., DiSorbo, O., Bennett, L., Li, L., Tennant, R., Stoll, R., Barrett, J. C., Paules, R. S., Blanchard, K. and Afshari, C. A. (2002) Prediction of compound signature using high density gene expression profiling. *Toxicol Sci* **67**(2), 232-40.

103. Waring, J. F., Gum, R., Morfitt, D., Jolly, R. A., Ciurlionis, R., Heindel, M., Gallenberg, L., Buratto, B. and Ulrich, R. G. (2002) Identifying toxic mechanisms using DNA microarrays: evidence that an experimental inhibitor of cell adhesion molecule expression signals through the aryl hydrocarbon nuclear receptor. *Toxicology* **181-182**, 537-50.

104. Waters, M., Boorman, G., Bushel, P., Cunningham, M., Irwin, R., Merrick, A., Olden, K., Paules, R., Selkirk, J., Stasiewicz, S., Weis, B., Houten, B. V., Walker,

N. and Tennant, R. (2003) Systems Toxicology and the Chemical Effects in Biological Systems (CEBS) Knowledge Base. *Environ Health Perspect* **111**(6), 811-24.

105. Velculescu, V. E., Zhang, L., Vogelstein, B. and Kinzler, K. W. (1995) Serial analysis of gene expression. *Science* **270**(5235), 484-7.

106. Velculescu, V. E., Zhang, L., Zhou, W., Vogelstein, J., Basrai, M. A., Bassett, D. E., Jr., Hieter, P., Vogelstein, B. and Kinzler, K. W. (1997) Characterization of the yeast transcriptome. *Cell* **88**(2), 243-51.

107. Velculescu, V. E., Vogelstein, B. and Kinzler, K. W. (2000) Analysing uncharted transcriptomes with SAGE. *Trends Genet* **16**(10), 423-5.

108. Liang, P., Averboukh, L., Keyomarsi, K., Sager, R. and Pardee, A. B. (1992) Differential display and cloning of messenger RNAs from human breast cancer versus mammary epithelial cells. *Cancer Res* **52**(24), 6966-8.

109. Liang, P. and Pardee, A. B. (1992) Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. *Science* **257**(5072), 967-71.

110. Jiang, H., Lin, J. J., Su, Z. Z., Goldstein, N. I. and Fisher, P. B. (1995) Subtraction hybridization identifies a novel melanoma differentiation associated gene, mda-7, modulated during human melanoma differentiation, growth and progression. *Oncogene* **11**(12), 2477-86.

111. Jiang, H., Kang, D. C., Alexandre, D. and Fisher, P. B. (2000) RaSH, a rapid subtraction hybridization approach for identifying and cloning differentially expressed genes. *Proc Natl Acad Sci U S A* **97**(23), 12684-9.

112. Larkin, P., Folmar, L. C., Hemmer, M. J., Poston, A. J., Lee, H. S. and Denslow, N. D. (2002) Array technology as a tool to monitor exposure of fish to xenoestrogens. *Mar Environ Res* **54**(3-5), 395-9.

113. Larkin, P., Sabo-Attwood, T., Kelso, J. and Denslow, N. D. (2002) Gene expression analysis of largemouth bass exposed to estradiol, nonylphenol, and p,p'-DDE. *Comp Biochem Physiol B Biochem Mol Biol* **133**(4), 543-57.

114. Selmin, O., Lucier, G. W., Clark, G. C., Tritscher, A. M., Vanden Heuvel, J. P., Gastel, J. A., Walker, N. J., Sutter, T. R. and Bell, D. A. (1996) Isolation and characterization of a novel gene induced by 2,3,7,8- tetrachlorodibenzo-p-dioxin in rat liver. *Carcinogenesis* **17**(12), 2609-15.

115. Dong, L., Ma, Q. and Whitlock, J. P., Jr. (1997) Down-regulation of major histocompatibility complex Q1b gene expression by 2,3,7,8-tetrachlorodibenzo-p-dioxin. *J Biol Chem* **272**(47), 29614-9.

116. Gao, L., Dong, L. and Whitlock, J. P., Jr. (1998) A novel response to dioxin. Induction of ecto-ATPase gene expression. *J Biol Chem* **273**(25), 15358-65.

117. Roman, B. L. and Peterson, R. E. (1998) In utero and lactational exposure of the male rat to 2,3,7,8- tetrachlorodibenzo-p-dioxin impairs prostate development. 1. Effects on gene expression. *Toxicol Appl Pharmacol* **150**(2), 240-53.

118. Donat, S. and Abel, J. (1998) Analysis of gene expression in lung and thymus of TCDD treated C57BL/6 mice using differential display RT-PCR. *Chemosphere* **37**(9-12), 1867-72.

119. Ma, Q., Baldwin, K. T., Renzelli, A. J., McDaniel, A. and Dong, L. (2001) TCDD-inducible poly(ADP-ribose) polymerase: a novel response to 2,3,7,8-tetrachlorodibenzo-p-dioxin. *Biochem Biophys Res Commun* **289**(2), 499-506.

120. Svensson, C. and Lundberg, K. (2001) Immune-specific up-regulation of adseverin gene expression by 2,3,7,8- tetrachlorodibenzo-p-dioxin. *Mol Pharmacol* **60**(1), 135-42.

121. Kolluri, S. K., Balduf, C., Hofmann, M. and Gottlicher, M. (2001) Novel target genes of the Ah (dioxin) receptor: transcriptional induction of N-myristoyltransferase 2. *Cancer Res* **61**(23), 8534-9.

122. Cao, Z., Tanguay, R. L., McKenzie, D., Peterson, R. E. and Aiken, J. M. (2003) Identification of a putative calcium-binding protein as a dioxin- responsive gene in zebrafish and rainbow trout. *Aquat Toxicol* **63**(3), 271-82.

123. Frueh, F. W., Hayashibara, K. C., Brown, P. O. and Whitlock, J. P., Jr. (2001) Use of cDNA microarrays to analyze dioxin-induced changes in human liver gene expression. *Toxicol Lett* **122**(3), 189-203.

124. Kurachi, M., Hashimoto, S., Obata, A., Nagai, S., Nagahata, T., Inadera, H., Sone, H., Tohyama, C., Kaneko, S., Kobayashi, K. and Matsushima, K. (2002) Identification of 2,3,7,8-tetrachlorodibenzo-p-dioxin-responsive genes in mouse liver by serial analysis of gene expression. *Biochem Biophys Res Commun* **292**(2), 368-77.

125. Son, D. S. and Rozman, K. K. (2002) 2,3,7,8-Tetrachlorodibenzo-p-dioxin (TCDD) induces plasminogen activator inhibitor-1 through an aryl hydrocarbon

receptor-mediated pathway in mouse hepatoma cell lines. *Arch Toxicol* **76**(7), 404-13.

126. Funseth, E., Pahlman, M., Eloranta, M. L., Friman, G. and Ilback, N. G. (2002) Effects of coxsackievirus B3 infection on the acute-phase protein metallothionein and on cytochrome P-4501A1 involved in the detoxification processes of TCDD in the mouse. *Sci Total Environ* **284**(1-3), 37-47.

127. Nishimura, N., Miyabara, Y., Suzuki, J. S., Sato, M., Aoki, Y., Satoh, M., Yonemoto, J. and Tohyama, C. (2001) Induction of metallothionein in the livers of female Sprague-Dawley rats treated with 2,3,7 ,8-tetrachlorodibenzo-p-dioxin. *Life Sci* **69**(11), 1291-303.

128. Spitsbergen, J. M. and Kent, M. L. (2003) The state of the art of the zebrafish model for toxicology and toxicologic pathology research--advantages and current limitations. *Toxicol Pathol* **31 Suppl**, 62-87.

129. Driever, W., Solnica-Krezel, L., Schier, A. F., Neuhauss, S. C., Malicki, J., Stemple, D. L., Stainier, D. Y., Zwartkruis, F., Abdelilah, S., Rangini, Z., Belak, J. and Boggs, C. (1996) A genetic screen for mutations affecting embryogenesis in zebrafish. *Development* **123**, 37-46.

130. Allred, P. M. and Strange, J. R. (1977) The effects of 2,4,5-trichlorophenoxyacetic acid and 2,3,7,8-tetrachlorodibenzo-p-dioxin on developing chicken embryos. *Arch Environ Contam Toxicol* **6**(4), 483-9.

131. Kennedy, S. W., Lorenzen, A., Jones, S. P., Hahn, M. E. and Stegeman, J. J. (1996) Cytochrome P4501A induction in avian hepatocyte cultures: a promising approach for predicting the sensitivity of avian species to toxic effects of halogenated aromatic hydrocarbons. *Toxicol Appl Pharmacol* **141**(1), 214-30.

132. Schena, M., Shalon, D., Davis, R. W. and Brown, P. O. (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270**(5235), 467-470.

133. Behr, M. A., Wilson, M. A., Gill, W. P., Salamon, H., Schoolnik, G. K., Rane, S. and Small, P. M. (1999) Comparative genomics of BCG vaccines by whole-genome DNA microarray. *Science* **284**(5419), 1520-1523.

134. Pollack, J. R., Perou, C. M., Alizadeh, A. A., Eisen, M. B., Pergamenschikov, A., Williams, C. F., Jeffrey, S. S., Botstein, D. and Brown, P. O. (1999) Genome-wide analysis of DNA copy-number changes using cDNA microarrays. *Nat Genet* **23**(1), 41-46.

135. Lieb, J. D. (2003) Genome-wide mapping of protein-DNA interactions by chromatin immunoprecipitation and DNA microarray hybridization. In *Functional Genomics: Methods and Protocols* (ed. M. J. Brownstein and A. Khodursky), pp. 99-110. Humana Press.

136. Holloway, A. J., van Laar, R. K., Tothill, R. W. and Bowtell, D. D. (2002) Options available--from start to finish--for obtaining data from DNA microarrays II. *Nat Genet* **32 Suppl**, 481-9.

137. Kane, M. D., Jatkoe, T. A., Stumpf, C. R., Lu, J., Thomas, J. D. and Madore, S. J. (2000) Assessment of the sensitivity and specificity of oligonucleotide (50mer) microarrays. *Nucleic Acids Res* **28**(22), 4552-7.

138. Xu, W., Bak, S., Decker, A., Paquette, S. M., Feyereisen, R. and Galbraith, D. W. (2001) Microarray-based analysis of gene expression in very large gene families: the cytochrome P450 gene superfamily of Arabidopsis thaliana. *Gene* **272**(1-2), 61-74.

139. Ton, C., Hwang, D. M., Dempsey, A. A., Tang, H.-C., Yoon, J., Lim, M., Mably, J. D., Fishman, M. C. and Liew, C.-C. (2000) Identification, characterization, and mapping of expressed sequence tags from an embryonic zebrafish heart cDNA library. *Genome Res* **10**, 1915-1927.

140. Jenssen, T. K., Langaas, M., Kuo, W. P., Smith-Sorensen, B., Myklebost, O. and Hovig, E. (2002) Analysis of repeatability in spotted cDNA microarrays. *Nucleic Acids Res* **30**(14), 3235-44.

141. Bober, M., Wiehe, K., Yung, C., Suzek, T. O., Lin, M., Baumgartner, W. J. and Winslow, R. The Cardiac Gene Expression Knowledgebase (CaGE), Vol. 2003.

142. *Handbook of fluorescent probes and research products*. Molecular Probes.

143. Battaglia, C., Salani, G., Consolandi, C., Bernardi, L. R. and De Bellis, G. (2000) Analysis of DNA microarrays by non-destructive fluorescent staining using SYBR green II. *Biotechniques* **29**(1), 78-81.

144. Hessner, M. J., Wang, X., Khan, S., Meyer, L., Schlicht, M., Tackes, J., Datta, M. W., Jacob, H. J. and Ghosh, S. (2003) Use of a three-color cDNA microarray platform to measure and control support-bound probe for improved data quality and reproducibility. *Nucleic Acids Res* **31**(11), e60.

145. Hessner, M. J., Wang, X., Hulse, K., Meyer, L., Wu, Y., Nye, S., Guo, S. W. and Ghosh, S. (2003) Three color cDNA microarrays: quantitative assessment through the use of fluorescein-labeled probes. *Nucleic Acids Res* **31**(4), e14.

146. Worley, J., Bechtol, K., Penn, S., Roach, D., Hanzel, D., Trounstine, M. and Barker, D. (2000) Technology and Applications of Gene Expression Microarrays. In *Microarray Biochip Technology* (ed. M. Schena), pp. 65-85. Eaton Publishing.

147. Dobbin, K., Shih, J. H. and Simon, R. (2003) Statistical design of reverse dye microarrays. *Bioinformatics* **19**(7), 803-810.

148. Tseng, G. C., Oh, M., Rohlin, L., Liao, J. and Wong, W. H. (2001) Issues in cDNA microarray analysis: quality filtering, channel normalization, models of variations and assessment of gene effects. *Nucleic Acids Res* **29**, 2549-2557.

149. Pfaffl, M. W. (2001) A new mathematical model for relative quantification in real-time RT- PCR. *Nucleic Acids Res* **29**(9), e45.

150. Broughton, R. E., Milam, J. E. and Roe, B. A. (2001) The complete sequence of the zebrafish (Danio rerio) mitochondrial genome and evolutionary patterns in vertebrate mitochondrial DNA. *Genome Res* **11**(11), 1958-67.

151. Calabrese, E. J. and Baldwin, L. A. (2003) The hormetic dose-response model is more common than the threshold model in toxicology. *Toxicol Sci* **71**, 246-250.

152. Nebert, D. W., Puga, A. and Vasiliou, V. (1993) Role of the Ah receptor and the dioxin-inducible [Ah] gene battery in toxicity, cancer, and signal transduction. *Ann N Y Acad Sci* **685**, 624-40.

153. Nebert, D. W., Roe, A. L., Dieter, M. Z., Solis, W. A., Yang, Y. and Dalton, T. P. (2000) Role of the aromatic hydrocarbon receptor and [Ah] gene battery in the oxidative stress response, cell cycle control, and apoptosis. *Biochem Pharmacol* **59**(1), 65-85.

154. Brake, P. B., Zhang, L. and Jefcoate, C. R. (1998) Aryl hydrocarbon receptor regulation of cytochrome P4501B1 in rat mammary fibroblasts: evidence for transcriptional repression by glucocorticoids. *Mol Pharmacol* **54**(5), 825-33.

155. Christou, M., Savas, U., Schroeder, S., Shen, X., Thompson, T., Gould, M. N. and Jefcoate, C. R. (1995) Cytochromes CYP1A1 and CYP1B1 in the rat mammary gland: cell-specific expression and regulation by polycyclic aromatic hydrocarbons and hormones. *Mol Cell Endocrinol* **115**(1), 41-50.

156. Bofinger, D. P., Feng, L., Chi, L. H., Love, J., Stephen, F. D., Sutter, T. R., Osteen, K. G., Costich, T. G., Batt, R. E., Koury, S. T. and Olson, J. R. (2001) Effect of TCDD exposure on CYP1A1 and CYP1B1 expression in explant cultures of human endometrium. *Toxicol Sci* **62**(2), 299-314.

157. Sanderson, J. T., Slobbe, L., Lansbergen, G. W., Safe, S. and van den Berg, M. (2001) 2,3,7,8-Tetrachlorodibenzo-p-dioxin and diindolylmethanes differentially induce cytochrome P450 1A1, 1B1, and 19 in H295R human adrenocortical carcinoma cells. *Toxicol Sci* **61**(1), 40-8.

158. Jana, N. R., Sarkar, S., Ishizuka, M., Yonemoto, J., Tohyama, C. and Sone, H. (2000) Comparative effects of 2,3,7,8-tetrachlorodibenzo-p-dioxin on MCF-7, RL95-2, and LNCaP cells: role of target steroid hormones in cellular responsiveness to CYP1A1 induction. *Mol Cell Biol Res Commun* **4**(3), 174-80.

159. Badawi, A. F., Cavalieri, E. L. and Rogan, E. G. (2000) Effect of chlorinated hydrocarbons on expression of cytochrome P450 1A1, 1A2 and 1B1 and 2- and 4-hydroxylation of 17beta-estradiol in female Sprague-Dawley rats. *Carcinogenesis* **21**(8), 1593-9.

160. Bard, S. M., Woodin, B. R. and Stegeman, J. J. (2002) Expression of P-glycoprotein and cytochrome p450 1A in intertidal fish (Anoplarchus purpurescens) exposed to environmental contaminants. *Aquat Toxicol* **60**(1-2), 17-32.

161. Bard, S. M., Bello, S. M., Hahn, M. E. and Stegeman, J. J. (2002) Expression of P-glycoprotein in killifish (Fundulus heteroclitus) exposed to environmental xenobiotics. *Aquat Toxicol* **59**(3-4), 237-51.

162. Hwang, J. J., Allen, P. D., Tseng, G. C., Lam, C. W., Fananapazir, L., Dzau, V. J. and Liew, C. C. (2002) Microarray gene expression profiles in dilated and hypertrophic cardiomyopathic end-stage heart failure. *Physiol Genomics* **10**(1), 31-44.

163. Senft, A. P., Dalton, T. P., Nebert, D. W., Genter, M. B., Hutchinson, R. J. and Shertzer, H. G. (2002) Dioxin increases reactive oxygen production in mouse liver mitochondria. *Toxicol Appl Pharmacol* **178**(1), 15-21.

164. Senft, A. P., Dalton, T. P., Nebert, D. W., Genter, M. B., Puga, A., Hutchinson, R. J., Kerzee, J. K., Uno, S. and Shertzer, H. G. (2002) Mitochondrial reactive oxygen production is dependent on the aromatic hydrocarbon receptor. *Free Radic Biol Med* **33**(9), 1268-78.

165. Nohl, H., de Silva, D. and Summer, K. H. (1989) 2,3,7,8, tetrachlorodibenzo-p-dioxin induces oxygen activation associated with cell respiration. *Free Radic Biol Med* **6**(4), 369-74.

166. Takeda, N. (2003) Cardiomyopathy: Molecular and immunological aspects (Review). *Int J Mol Med* **11**(1), 13-6.

167. Dzeja, P. P., Redfield, M. M., Burnett, J. C. and Terzic, A. (2000) Failing energetics in failing hearts. *Curr Cardiol Rep* **2**(3), 212-7.

168. Dzeja, P. P., Pucar, D., Redfield, M. M., Burnett, J. C. and Terzic, A. (1999) Reduced activity of enzymes coupling ATP-generating with ATP-consuming processes in the failing myocardium. *Mol Cell Biochem* **201**(1-2), 33-40.

169. Seidman, J. G. and Seidman, C. (2001) The genetic basis for cardiomyopathy: from mutation identification to mechanistic paradigms. *Cell* **104**, 557-567.

170. Haase, D., Lehmann, M. H., Korner, M. M., Korfer, R., Sigusch, H. H. and Figulla, H. R. (2002) Identification and validation of selective upregulation of ventricular myosin light chain type 2 mRNA in idiopathic dilated cardiomyopathy. *Eur J Heart Fail* **4**(1), 23-31.

171. Barrans, J. D., Allen, P. D., Stamatiou, D., Dzau, V. J. and Liew, C. C. (2002) Global gene expression profiling of end-stage dilated cardiomyopathy using a human cardiovascular-based cDNA microarray. *Am J Pathol* **160**(6), 2035-43.

172. Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., *et al.* (2001) Initial sequencing and analysis of the human genome. *Nature* **409**(6822), 860-921.

173. Waterston, R. H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J. F., Agarwal, P., Agarwala, R., Ainscough, R., Alexandersson, M., An, P., *et al.* (2002) Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**(6915), 520-62.

174. Deininger, P. L. and Batzer, M. A. (2002) Mammalian retroelements. *Genome Res* **12**(10), 1455-65.

175. McCarthy, E. M., Liu, J., Lizhi, G. and McDonald, J. F. (2002) Long terminal repeat retrotransposons of Oryza sativa. *Genome Biol* **3**(10), RESEARCH0053.

176. McCarthy, E. M. and McDonald, J. F. (2003) LTR_STRUC: a novel search and identification program for LTR retrotransposons. *Bioinformatics* **19**(3), 362-7.

177. Duncan, L., Bouckaert, K., Yeh, F. and Kirk, D. L. (2002) kangaroo, a Mobile Element From Volvox carteri, Is a Member of a Newly Recognized Third Class of Retrotransposons. *Genetics* **162**(4), 1617-30.

178. Goodwin, T. J. and Poulter, R. T. (2001) The DIRS1 group of retrotransposons. *Mol Biol Evol* **18**(11), 2067-82.

179. Vogel, A. M. and Gerster, T. (1999) Promoter activity of the zebrafish bhikhari retroelement requires an intact activin signaling pathway. *Mech Dev* **85**(1-2), 133-46.

180. Westerfield, M. (2000) *The Zebrafish Book: A guide for the laboratory use of zebrafish (Danio rerio).* University of Oregon Press.

181. Geisler, R., Rauch, G. J., Baier, H., van Bebber, F., Brobeta, L., Dekens, M. P., Finger, K., Fricke, C., Gates, M. A., Geiger, H., *et al.* (1999) A radiation hybrid map of the zebrafish genome. *Nat Genet* **23**(1), 86-89.

182. Quandt, K., Frech, K., Karas, H., Wingender, E. and Werner, T. (1995) MatInd and MatInspector: new fast and versatile tools for detection of consensus matches in nucleotide sequence data. *Nucleic Acids Res* **23**(23), 4878-84.

183. Curcio, M. J. and Garfinkel, D. J. (1994) Heterogeneous functional Ty1 elements are abundant in the Saccharomyces cerevisiae genome. *Genetics* **136**(4), 1245-59.

184. Goncalves, I., Duret, L. and Mouchiroud, D. (2000) Nature and structure of human genes that generate retropseudogenes. *Genome Res* **10**(5), 672-8.

185. Chen, J. N. and Fishman, M. C. (1996) Zebrafish tinman homolog demarcates the heart field and initiates myocardial differentiation. *Development* **122**(12), 3809-16.

186. Skerjanc, I. S., Petropoulos, H., Ridgeway, A. G. and Wilton, S. (1998) Myocyte enhancer factor 2C and Nkx2-5 up-regulate each other's expression and initiate cardiomyogenesis in P19 cells. *J Biol Chem* **273**(52), 34904-10.

187. Yao, Y., Hoffer, A., Chang, C. Y. and Puga, A. (1995) Dioxin activates HIV-1 gene expression by an oxidative stress pathway requiring a functional cytochrome P450 CYP1A1 enzyme. *Environ Health Perspect* **103**(4), 366-71.

188. Jankun, J., Matsumura, F., Kaneko, H., Trosko, J. E., Pellicer, A. and Greenberg, A. H. (1989) Plasmid-aided insertion of MMTV-LTR and ras DNAs to NIH 3T3 fibroblast cells makes them responsive to 2,3,7,8-TCDD causing overexpression of p21ras and down-regulation of EGF receptor. *Mol Toxicol* **2**(3), 177-86.

189. Abbott, B. D., Perdew, G. H., Buckalew, A. R. and Birnbaum, L. S. (1994) Interactive regulation of Ah and glucocorticoid receptors in the synergistic

induction of cleft palate by 2,3,7,8-tetrachlorodibenzo-p-dioxin and hydrocortisone. *Toxicol Appl Pharmacol* **128**(1), 138-50.

190. Puga, A., Barnes, S. J., Chang, C., Zhu, H., Nephew, K. P., Khan, S. A. and Shertzer, H. G. (2000) Activation of transcription factors activator protein-1 and nuclear factor-kappaB by 2,3,7,8-tetrachlorodibenzo-p-dioxin. *Biochem Pharmacol* **59**(8), 997-1005.

191. Puga, A., Nebert, D. W. and Carrier, F. (1992) Dioxin induces expression of c-fos and c-jun proto-oncogenes and a large increase in transcription factor AP-1. *DNA Cell Biol* **11**(4), 269-81.

192. Ashida, H. and Matsumura, F. (1998) Effect of in vivo administered 2,3,7,8-tetrachlorodibenzo-p-dioxin on DNA-binding activities of nuclear transcription factors in liver of guinea pigs. *J Biochem Mol Toxicol* **12**(4), 191-204.

193. Hoffer, A., Chang, C. Y. and Puga, A. (1996) Dioxin induces transcription of fos and jun genes by Ah receptor- dependent and -independent pathways. *Toxicol Appl Pharmacol* **141**(1), 238-47.

194. Enan, E. and Matsumura, F. (1995) Regulation by 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD) of the DNA binding activity of transcriptional factors via nuclear protein phosphorylation in guinea pig adipose tissue. *Biochem Pharmacol* **50**(8), 1199-206.

195. Kigami, D., Minami, N., Takayama, H. and Imai, H. (2003) MuERV-L Is One of the Earliest Transcribed Genes in Mouse One-Cell Embryos. *Biol Reprod* **68**(2), 651-4.

196. Perkins, E. (2003) *Proceedings of PRIMO 12.*

197. Lu, K. P., Hallberg, L. M., Tomlinson, J. and Ramos, K. S. (2000) Benzo(a)pyrene activates L1Md retrotransposon and inhibits DNA repair in vascular smooth muscle cells. *Mutat Res* **454**(1-2), 35-44.

198. Bing, O. H., Sirokman, G. and Humphries, D. E. (1998) Hypothesis: link between endogenous retroviruses and cardiovascular disease. *J Mol Cell Cardiol* **30**(7), 1257-62.

199. Sirokman, G., Humphries, D. E. and Bing, O. H. (1997) Endogenous retroviral transcripts in myocytes from spontaneously hypertensive rats. *Hypertension* **30**(1 Pt 1), 88-93.

200. Nowak, J., Januszkiewicz, D., Pernak, M., Liwen, I., Zawada, M., Rembowska, J., Nowicka, K., Lewandowski, K., Hertmanowska, H. and Wender, M. (2003) Multiple sclerosis-associated virus-related pol sequences found both in multiple sclerosis and healthy donors are more frequently expressed in multiple sclerosis patients. *J Neurovirol* **9**(1), 112-7.

201. Ide, T., Tsutsui, H., Kinugawa, S., Utsumi, H., Kang, D., Hattori, N., Uchida, K., Arimura, K., Egashira, K. and Takeshita, A. (1999) Mitochondrial electron transport complex I is a potential source of oxygen free radicals in the failing myocardium. *Circ Res* **85**, 357-363.

202. Dhalla, N. S., Temsah, R. M. and Netticadan, T. (2000) Role of oxidative stress in cardiovascular disease. *J Hypertens* **18**(6), 655-673.

203. Sorescu, D. and Griendling, K. K. (2002) Reactive oxygen species, mitochondria, and NAD(P)H oxidases in the development and progression of heart failure. *Congest Heart Fail* **8**(3), 132-140.

204. Lee, P. J. and Choi, A. M. K. (2003) Pathways of cell signaling in hyperoxia. *Free Rad Biol Med* **35**(4), 341-350.

205. Pelster, B., Sanger, A. M., Siegele, M. and Schwerte, T. (2003) Influence of swim training on cardiac activity, tissue capillarization, and mitochondrial density in muscle tissue of zebrafish larvae. *Am J Physiol Regul Integr Comp Physiol* **285**(2), R339-47.

206. Guhaniyogi, J. and Brewer, G. (2001) Regulation of mRNA stability in mammalian cells. *Gene* **265**, 11-23.

207. Brown, B. S. V., Chi, T. B. and Williams, N. (2001) The Trypanosoma brucei mitochondrial ATP synthase is developmentally regulated at the level of transcript stability. *Mol Biochem Parasitol* **115**(2), 177-187.

208. Thames, E. L., Newton, D. A., Black, S. A. J. and Bowman, L. H. (2000) Role of mRNA stability and translation in the expression of cytochrome c oxidase during mouse myoblast differentiation: instability of the mRNA for the liver isoform of subunit VIa. *Biochem J* **351**, 133-142.

209. Butow, R. A., Zhu, H., Perlman, P. and Conrad-Webb, H. (1989) The role of a conserved dodecamer sequence in yeast mitochondrial gene expression. *Genome* **31**(2), 757-760.

210. Flamme, I., von Reutern, M., Drexler, H. C. A., Syed-Ali, S. and Risau, W. (1995) Overexpression of vascular endothelial growth factor in the avian embryo

induces hypervascularization and increased vascular permeability without alterations of embryonic pattern formation. *Dev Biol* **171**, 399-414.

211. Forsythe, J. A., Jiang, B.-H., Iyer, N. V., Agani, F., Leung, S. W., Koos, R. D. and Semenza, G. L. (1996) Activation of vascular endothelial growth factor gene transcription by hypoxia-inducible factor 1. *Mol Cell Biol* **16**, 4604-4613.

212. Ema, M., Taya, S., Yokotani, N., Sogawa, K., Matsuda, Y. and Fujii-Kuriyama, Y. (1997) A novel bHLH-PAS factor with close sequence similarity to hypoxia-inducible factor 1a regulates the VEGF expression and is potentially invvd in lung and vascular development. *Proc Natl Acad Sci* **94**, 4273-4278.

213. Gradin, K., McGuire, J., Wenger, R. H., Kvietikova, I., Fhitelaw, M. L., Toftgard, R., Tora, L., Gassmann, M. and Poellinger, L. (1996) Functional interference between hypoxia and dioxin signal transduction pathways: competition for recruitment of the ARNT transcription factor. *Mol Cell Biol* **16**(10), 5221-5231.

214. Dohr, O., Vogel, C. and Abel, J. (1994) Modulation of growth factor expression by 2,3,7,8-tetrachlorodibenzo-*p*-dioxin. *Exp Clin Immuogene* **11**, 142-148.

215. Chua, C. C., Hamdy, R. C. and Chua, B. H. L. (1998) Upregulation of vascular endothelial growth factor by H2O2 in rat heart endothelial cells. *Free Rad Biol Med* **25**(8), 891-897.

216. Kuroki, M., Voest, E. E., Amano, S., Beerepoot, L. V., Takashima, S., Tolentino, M., Kim, R. Y., Rohan, R. M., Colby, K. A., Yeo, K.-T. and Adamis, A. P. (1996) Reactive oxygen intermediates increase vascular endothelial growth factor expression in vitro and in vivo. *J Clin Invest* **98**(7), 1667-1675.

217. Chismar, J. D., Mondala, T., Fox, H. S., Roberts, E., Langford, D., Masliah, E., Salomon, D. R. and Head, S. R. (2002) Analysis of result variability from high-density oligonucleotide arrays comparing same-species and cross-species hybridizations. *Biotechniques* **33**(3), 516-518.

218. Moody, D. E., Zou, Z. and McIntyre, L. (2002) Cross-species hybridization of pig RNA to human nylon microarrays. *BMC Genomics* **3**, 27-39.

Document Library

Distribution List for Technical Report Exchange—August 2002

University of California, San Diego
SIO Library 0175C
9500 Gilman Drive
La Jolla, CA 92093-0175

Hancock Library of Biology & Oceanography
Alan Hancock Laboratory
University of Southern California
University Park
Los Angeles, CA 90089-0371

Gifts & Exchanges
Library
Bedford Institute of Oceanography
P.O. Box 1006
Dartmouth, NS B2Y 4 A2
CANADA

NOAA/EDIS Miami Library Center
4301 Rickenbacker Causeway
Miami, FL 33149

Research Library
U.S. Army Corps of Engineers
Waterways Experiment Station
3909 Halls Ferry Road
Vicksburg, MS 39180-6199

Institute of Geophysics
University of Hawaii
Library Room 252
2525 Correa Road
Honolulu, HI 96822

Marine Resources Information Center
Building E38-320
MIT
Cambridge, MA 02139

Library
Lamont-Doherty Geological Observatory
Columbia University
Palisades, NY 10964

Library
Serials Department
Oregon State University
Corvallis, OR 97331

Pell Marine Science Library
University of Rhode Island
Narragansett Bay Campus
Narragansett, RI 02882

Working Collection
Texas A&M University
Dept. of Oceanography
College Station, TX 77843

Fisheries-Oceanography Library
151 Oceanography Teaching Bldg.
University of Washington
Seattle, WA 98195

Library
R.S.M.A.S.
University of Miami
4600 Rickenbacker Causeway
Miami, FL 33149

Maury Oceanographic Library
Naval Oceanographic Office
Building 1003 South
1002 Balch Blvd.
Stennis Space Center, MS 39522-5001

Library
Institute of Ocean Sciences
P.O. Box 6000
Sidney, B.C. V8L 4B2
CANADA

National Oceanographic Library
Southampton Oceanography Centre
European Way
Southampton SO14 3ZH
UK

The Librarian
CSIRO Marine Laboratories
G.P.O. Box 1538
Hobart, Tasmania
AUSTRALIA 7001

Library
Proudman Oceanographic Laboratory
Bidston Observatory
Birkenhead
Merseyside L43 7 RA
UK

IFREMER
Centre de Brest
Service Documentation–Publications
BP 70 29280 PLOUZANE
FRANCE

50272-101

REPORT DOCUMENTATION PAGE	1. REPORT NO. MIT/WHOI 2003-19	2.	3. Recipient's Accession No.

4. Title and Subtitle
Zebrafish Cardiovascular cDNA Microarrays: Expression Profiling and Gene Discovery in Embryos Exposed to 2,3,7,8–Tetrachlorodibenzo-*p*-dioxin

5. Report Date September 2003

6.

7. Author(s) Heather Martin Handley

8. Performing Organization Rept. No.

9. Performing Organization Name and Address
MIT/WHOI Joint Program in Oceanography/Applied Ocean Science & Engineering

10. Project/Task/Work Unit No. MIT/WHOI 2003-19

11. Contract(C) or Grant(G) No.
(C) NIH 5-P42-ES07381;
(G) EPA R827102-01-0;

12. Sponsoring Organization Name and Address
National Institutes of Health Superfund Basic Research Program
Environmental Protection Agency
National Science Foundation
Woods Hole Oceanographic Institution—Ocean Life Institute and Coastal Ocean Institute

13. Type of Report & Period Covered Ph.D. Thesis

14.

15. Supplementary Notes
This thesis should be cited as: Heather Martin Handley, 2003. Zebrafish Cardiovascular cDNA Microarrays: Expression Profiling and Gene Discovery in Embryos Exposed to 2,3,7,8–Tetrachlorodibenzo-p-dioxin. Ph.D. Thesis. MIT/WHOI, 2003-19.

16. Abstract (Limit: 200 words)

2,3,7,8-Tetrachlorodibenzo-*p*-dioxin (TCDD) is a potent teratogen that impacts cardiovascular development in fish. The goal of this thesis work was to identify genes likely to be involved in embryotoxicity. We constructed microarrays using cDNA libraries derived from zebrafish embryonic and adult heart tissue. Embryonic heart arrays were used for protocol development. The resulting workflow was employed in production of adult heart microarrays containing ~2800 unique genes. These arrays were used to characterize gene expression in TCDD-treated zebrafish embryos. Overall, 44 genes or ESTs were differentially expressed ≥1.8-fold (p-values ≤ 5×10^{-4}). Transcriptional responses were highly dose-dependent, and adaptive responses were a prevalent feature of TCDD expression profiles. TCDD-responsive genes fell into three major functional classes – xenobiotic detoxification, sarcomere structure, and energy transfer. Induction of mitochondrial electron transfer genes was variable and modest; such induction may contribute to TCDD-induced reactive oxygen generation. Altered expression of cardiac myosins and troponin T2 suggest TCDD-induced cardiomyopathy in zebrafish embryos. Investigation of differentially expressed EST sequences led to the discovery of a novel, unorthodox retroelement, EZR1. Putative regulatory elements in LTR sequences may account for constitutive expression and TCDD induction. The function, if any, of EZR1 remains open to speculation.

17. Document Analysis a. Descriptors
zebrafish
microarray
dioxin

b. Identifiers/Open-Ended Terms

c. COSATI Field/Group

18. Availability Statement
Approved for publication; distribution unlimited.

19. Security Class (This Report) UNCLASSIFIED

20. Security Class (This Page)

21. No. of Pages 288

22. Price

(See ANSI-Z39.18) *See Instructions on Reverse* OPTIONAL FORM 272 (4-77) (Formerly NTIS-35) Department of Commerce