Genomic mechanisms underlying the collapse and lack of recovery of Prince William Sound herring

By

JAMES ANTHONY GILL III
DISSERATION

Submitted in partial satisfaction of the requirements for the degree of

Doctor of Philosophy

in

Integrative Genetics and Genomics

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

_____

Andrew Whitehead

_____

Titus Brown

_____

Gary Cherr

Committee in Charge

2023

Without the love and support of my friends and family, none of this would have been possible.

Genomic mechanisms underlying the collapse and lack of recovery of Prince William Sound herring

James Anthony Gill III

University of California Davis

PhD

2023

ABSTRACT

The *Exxon Valdez* oil spill occurred March 24, 1989, when herring were preparing to spawn in Prince William Sound. The herring population experienced an unanticipated, abrupt decline three years later, due–in part–to a mortality from infectious and parasitic diseases. Linking the oil spill to subsequent population collapse remains controversial. A major insight from years of studying the spill is that embryonic herring are profoundly sensitive to crude oil; exposure to vanishingly low levels of oil over a brief time early in a herring's life-cycle can have long-lasting health effects, and oil exposure can disturb immune function. Could crude oil exposure during early life have compromised their immune system development, thereby increasing the risk of major disease outbreak in later life? To address this question, over the past few years we have sought to simulate the events surrounding the 1993 herring collapse using 1) experimental exposures to environmentally relevant levels of Alaska north slope crude oil, 2) fish from the Prince William Sound population and others, and 3) exposing fish in the laboratory to the same

pathogens that caused the disease outbreak. To facilitate our investigation into the molecular effects of oil exposure on a non-traditional model organism, we sequenced and annotated a reference transcriptome for the Pacific herring, and conducted extensive research into a cost-effective, high-throughput RNA-sequencing library construction method for our samples. To that end we compared the effectiveness of two methods for generating sequencing libraries for gene expression analysis: 3'-end sequencing and whole transcript sequencing. We found similar levels of precision and power for detecting differentially expressed genes with both methods, but whole transcript sequencing performed better in non-traditional model species. Next, embryonic herring sourced from Prince William Sound, AK were exposed to a range of crude oil and transcriptomically interrogated across a detailed time-course, paying particular attention to heart and immune system development. We found that crude oil exposure disrupted cardiogenesis and caused heart defects in the developing fish, as well as modulated the immune system, causing dysregulation of gene expression. Finally, three geographically distinct populations of embryonic herring were exposed to low levels of oil, left to recover and grow-up in clean seawater, then exposed to pathogens. We found that an overall response to crude oil exposure in Pacific herring was determined by geography and not population history, with geographical distance playing a large role in molecular and phenotypic response to oil exposure. In conclusion, the Exxon Valdez oil spill may have contributed to the subsequent collapse of the herring population in Prince William Sound through compromising heart development and function in early development, causing long-lasting health effects. The research here shows that oil exposure to even low levels of crude oil can disrupt heart and

potentially immune system development in embryonic herring, causing dysregulation in key genes involved in cardiogenesis, with the added insight that geographical distance plays a large role in the response to oil exposure.

Acknowledgements

This project represents a combined effort between The University of Calinfornia Davis, CA, The USGS Marrowstone Marine Field Station, WA, and NOAA and the Northwest Fisheries Science Center, WA. Without the input and support of many individuals—including colleagues, friends, and family—the research presented in this thesis would not have been possible.

First and foremost, I would like to thank my advisor, Andrew Whitehead, for his involvement in developing this successful collaborative project. He has been integral in my development as a researcher by not only providing scientific and career guidance, but also by allowing me the freedom to make many of my own decisions with the project (and sometimes mistakes). For all of his numerous academic successes, Andrew is down-to-earth, approachable, and optimistic, qualities I hope to emulate throughout my scientific career.

I would also like to thank my colleague and friend, James Cameron, for acting as a skilled and thoughtful colleague. His input was invaluable, and his encouragement and support kept me going even through the dark times.

Thank you to John Incardona and Nat Sholtz for their committed mentorship, both as my

previous supervisors at NWSF/NOAA, and in the direction of the project (as major collaborators). In the same spirit, I would also like to thank Paul Hershberger for his invaluable advice on Pacific herring fish pathology and infectious disease biology and in large part for welcoming me to his research facility on Marrowstone Island, WA. Thank you to Titus Brown for bioinformatics guidance and graduate school advice, and for agreeing to act as a committee member for this thesis. Thank you to Gary Cherr for bringing decades of Pacific herring biology and oil spill aquatic toxicology experience to bear at my committee meetings and agreeing to act as a committee member for this thesis, and continuing to do so after becoming emeritus professor.

I would like to thank all of my friends and colleagues in the Whitehead lab. Specifically, thank you to Jen Roach for welcoming me to the Whitehead lab, for outstanding technical assistance, for offering thoughtful insight, and lending her expertise in troubleshooting tricky wet-lab experiments. Thank you to Tiffany Linbo, NWFSC/NOAA, and Jake Gregg, USGS, for their hands-on expertise in microscopy and imaging and the day-to-day organization of the fish husbandry and exposure experiments. Thank you to the rest of my Whitehead lab friends and colleagues, including Elias Oziolor, Jane Park, and Nicole McNabb, for their helpful discussion, advice, and general fun times.

Thank you to my parents, Jim and Kris, even though they never bought me a pet shark. You two

are my biggest cheerleaders. Also, thank you to my sister Emily for putting up with all the strange wildlife I did manage to bring indoors during our childhood, and, as adults, for always being there for me.

A final thanks to my brilliant, funny, giant-hearted, super-hot wife, Megan.

Table of Contents

List of Tables

List of Figures

Introduction

Genomic mechanisms underlying the collapse and lack of recovery of Prince William Sound herring

The causes of the herring fishery collapse in Prince William Sound, AK (AK-P) in the early 1990's are complicated and controversial, as is the relationship of the population collapse in 1989 to one of the largest environmental disasters of its time, the Exxon Valdez oil spill (Pearson et al. 1999; Carls, Marty, and Hose 2002; Thorne and Thomas 2008). Prior to the spill, the AK-P population was fluctuating in concert with a nearby population from Sitka Sound. Since the spill, the two populations remain out of sync, suggesting there are other stressors that are dominant to regional oceanographic factors. The AK-P Pacific herring population has been affected by two major events in the past 25 years: (1) the Exxon Valdez oil spill in AK-P occurred just before spawning in 1989; and (2) a major VHSV outbreak in the winter of 1993- 1994 (Brown, Norcross, and Short 1996b) and subsequent population collapse. The oil spill severely damaged herring embryos in 1989 with likely sublethal effects continuing into 1990 (Carls et al. 2001). However, no detectable oil remained after 1990 leaving a three- to four-year gap between the last of the contaminating oil and the collapse. By the time of the collapse, starting in the spring of 1994, VHSV infection in AK-P herring was prominent and remains a limiting factor to this day (Marty, Ii, et al. 2003); however, the cause of its persistence remains unknown.

VHSV is found in northern marine and some freshwater environments worldwide and constitutes one of the world's most serious fish diseases, producing frequent epidemics that are associated with mortality that can reach 100% (International Committee on Taxonomy of Viruses et al. 2000). Notably, the North American strain of VHSV was first isolated in Pacific herring from the AK-P in the spring of 1994, where barely 20% of expected herring returned to spawn that year. These fish were lethargic and had many external hemorrhages. The unexpectedly low return marked the collapse of the fishery (Meyers et al. 1994). A multi-year study (1994-2002) concluded that VHSV was the primary cause in the collapse and lack of recovery of the population (Marty et al. 1998b). Although this virus has been linked to high mortality in other populations of herring along the eastern North Pacific, it has not been associated with a collapse or slow recovery, suggesting that the latent infection and lack of population recovery is unique to the AK-P population and putatively associated with the oil spill. Recent studies have demonstrated that herring previously naive to VHSV are highly susceptible; however, susceptibility decreases as temperature increases (Hershberger et al. 2013). This inverse relationship between temperature and VHSV severity is likely mediated by an enhanced innate immune response at warmer temperatures (Arkush et al. 2006). In addition to temperature-dependent resistance to virus susceptibility, variability in resistance to VHSV in rainbow trout appears to be a heritable quantitative trait, suggesting genetic differences in VHSV susceptibility could exist between individuals, populations and species (Henryon et al. 2002; Dorson et al. 1995; Quillet et al. 2001). Recently targeted Sanger sequencing showed significant allelic differences between AK-P and nearby populations at a locus containing an important virus response gene (Roberts et al. 2012), although the cause and functional

outcome of this divergence remains untested. Comparative genome-wide gene expression analysis may detect further differences in vector competence, but currently this comparison is limited by the absence of a high quality reference transcriptome for Pacific herring.

In the years following the oil spill most research has focused on the direct effects of crude oil on herring development; however, less is known about how oil interacts with other stressors such as exposure to pathogens and disease normally encountered by herring in the wild (Roberts et al. 2012; Whitehead 2013). A combination of stressors unique to AK-P could have contributed to the collapse of the herring fishery following the oil spill. Support for this hypothesis is found in recent discoveries on the negative effects of crude oil on herring development—even at very low concentrations (Fig3). Embryonic exposure to contaminating oil at trace levels leads to permanently malformed hearts by disrupting intracellular calcium levels in herring heart muscle cells (Incardona et al. 2015). In line with this, components of crude oil directly modulate calcium homeostasis in immune cells in common carp, suppressing immune function (Reynaud, Duchiron, and Deschaux 2003). Crude oil is also a well-known inducer of the aryl hydrocarbon (AH) receptor signaling pathway (Nebert et al. 2004). This receptor is a transcription factor, highly conserved among vertebrates, that mediates defensive metabolic responses but also the toxic activity of many environmental xenobiotics, including soluble two- and three-ring polycyclic aromatic hydrocarbons (PAHs) abundant in crude oil (Billiard et al. 2002). While the AH receptor is integral in mediating toxicant exposure responses, that is not its only function. Studies over the past decade have revealed that the receptor is important for cardiogenesis,

hematopoiesis, and regulation of immune responses (Schmidt and Bradfield 1996; Julliard, Fechner, and Mezrich 2014). In fact, the AH receptor is tightly regulated in cells. Activation of this receptor dampens Type 1 Interferon (IFN-I) innate response to viral infection in human primary monocytes (Yamada et al. 2016). The IFN system, a rapid and powerful antiviral defense mechanism in vertebrates (Samuel 1991), induces expression in host cells of a number of proteins that inhibit the translation of viral RNA. AHR activation in response to PAHs offers a plausible mechanism by which exposure to oil modulates an innate immune response to virus. However, this receptor's role in immune response during viral infection is poorly understood in fish. I posit that exposure to contaminating oil from the 1989 spill had sub-lethal effects on developing fish that subsequently made survivors uniquely susceptible to common epizootics in their environment. If true, this would explain how a combination of stressors unique to AK-P (oil and virus) could have contributed to the collapse of the herring fishery but not necessarily the lack of recovery.

The other outstanding mystery in the AK-P herring population is their unexpected lack of recovery following the 1993-1994 collapse. While herring populations can greatly fluctuate in size over time, only the AK-P population and another in Hokkaido have failed to recover two decades after a crash (unpublished work by John Trochta, University of Washington). What sets the AK-P herring apart from the Japan population is the Exxon Valdez oil spill and the VHSV outbreak. Pollutants can act as powerful selective forces, driving rapid adaptation, often with concomitant physiological costs (Monosson 2012; Hendry, Farrugia, and Kinnison 2008;

Gassmann, Carrière, and Tabashnik 2009). For example, populations of killifish within an environment contaminated with creosote, which is chemically and toxicologically similar to crude oil, have rapidly evolved resistance to the toxic effects of PAHs and related chemicals. This resistance is caused by adaptive desensitization of AHR signaling, which is associated with a rapid increase in the frequency of a large deletion in the AHR coding region (Reid et al. 2016). This may seem paradoxical as PAHs are considered part of the classical adaptive receptor response, by which they are eliminated from tissues; however, induction of AHR during early development can lead to altered gene expression, developmental deformities, and decreased life-expectancy. Some high-molecular weight PAHs found in crude oil are converted to carcinogenic intermediates by CYP1A (Phillips 1983). Conversely, AH receptor's function in other physiological processes suggests that global suppression of the receptor in fish from polluted sites may have imposed a physiological cost. As evidence of this, the same killifish study identified additional population genomic outliers in pollution resistant fish that included regions harboring immune response genes (Reid et al. 2016). However, putative selection by PAHs for resilience to contaminants in developing embryos has yet unknown impacts on immune function later in life. Little is known of the receptor's role in fish immune function but numerous hematopoietic defects have been described in AHR null mouse models, including reduced lymphocyte numbers in the spleen, altered hematopoiesis and limited immune response to infectious disease (Julliard, Fechner, and Mezrich 2014). It is possible that a similar mechanism seen in killifish is at play in the slow recovery of the AK-P herring population, suggesting a significant population-level change in immune response. For example, exposure to oil during 1989-1990 may have selected for genetic variants with impaired AHR function in

developing embryos. Alternatively, (or in addition), exposure to oil in 1989-1990 may have impaired immune function, such that the frequency of individuals with impaired immune function may have increased within the AK-P population. By 1993, when fish from the 1998-year class were recruiting into the fishery, this may have enabled an epizootic that contributed to the population collapse. Recovery is slow in WA-P because compensatory adaptations for immune function are slow, possibly because the collapse significantly eroded the genetic diversity that would have been available for subsequent adaptive evolution of immune system genes.

Using a combined toolkit of genomics and experimental physiology, I systematically characterized the molecular mechanisms related to the collapse of the AK-P herring following the Exxon Valdez oil spill and subsequent difficulties in recovery of the population. A large portion of the data leveraged to test and generate hypotheses for this dissertation is whole transcriptome differential gene expression analysis. For Chapter 1 I comprehensively compared two popular methods for generating sequencing libraries using Pacific herring as a representative of a non-traditional model organism with limited genomic resources. In Chapter 2 I created a reference transcriptome for Pacific herring and characterized the molecular impact of oil exposure on cardiac and immune function across developmental stages in AK-P herring. In Chapter 3 I explored population differences in herring response to oil and virus exposures to test hypotheses about evolved differences between populations, and hypotheses about the

interactive effects of oil and virus exposures, that could offer insight into the causes and consequences of the AK-P collapse.

Chapter 1

3’ RNA sequencing does not increase power or reduce costs for gene expression analysis

1.1 Abstract

Sequencing RNA transcripts for gene expression profiling is a popular and important technique with broad utility in biological sciences. We set out to comprehensively compare the two most popular methods for generating sequencing libraries for differential gene expression analysis: 3-end sequencing, which generates libraries from the 3’ end of an RNA transcript; and traditional RNA sequencing, which generates libraries from whole RNA transcripts. We include three species in our experiment to test whether our findings replicate across genomes and genome assemblies. We found similar levels of precision and power to detect differentially expressed genes between the two methods. Notably, whole transcript RNA-seq performed better in the non-traditional model species included in our study. Overall, we recommended whole transcript RNA sequencing for the added benefits of alternative splicing detection, and gene-body variant detection.

1.2 Introduction

High-throughput RNA sequencing is widely used in biological research as a powerful tool for quantifying gene expression. The transcriptome reflects the state of gene expression at a cellular level, and is highly dynamic and responsive to external perturbations, which makes it an ideal quantitative molecular phenotype for a variety of biological questions. For example, how does expression change in response to environmental stressors? What cellular processes are important for acclimating to seasonal weather patterns? What genes are involved in the shift between vegetative and reproductive growth? Many basic and applied research hypotheses can be tested and validated using RNAseq, however, many questions and assumptions about RNAseq have not been adequately explored. One of the first questions a researcher needs to answer is: what type of RNAseq library construction should I choose?

The first step in RNAseq library construction is converting mRNA into cDNA libraries compatible for high-throughput sequencing platforms. Reverse transcription is primed with the addition of either random single-stranded hexamers–hybridizing along the length of the mRNA molecule–or through oligo d(T) primers–targeting the 3' polyadenylated tail of the mRNA. This step is called first strand synthesis, where the single-stranded mRNA molecule is paired with a complementary DNA (cDNA) strand. For the purposes of this dissertation, I refer to the two priming methods as whole transcript (WT) for random primers, and 3' for libraries constructed with an oligo d(T) primer.

While both WT and 3' libraries can accurately identify actively expressed genes (using strand-specific priming methods), their outputs permit and require different downstream applications.

WT libraries sequence across the full-length of a transcript, making it possible to quantify alternative splicing of transcripts at a single locus. 3' libraries can only distinguish among transcript families that have different 3'UTRs. However, expression-level estimates from WT libraries show a transcript length bias where longer transcripts are sequenced at a higher read count relative to their expression. This can lead to biases for downstream analyses, e.g., Gene Ontology (GO) enrichment that need a length-bias correction (Mandelboum et al. 2019). 3' libraries do not have a transcript-length bias because priming with a poly-A primer guarantees that only a single library molecule is made per transcript molecule.

Various authors have reasoned that either WT or 3' libraries are inherently superior for differential gene expression analysis, based on arguments of cost, statistical power, and ease of down-stream analysis (Lohman, Weber, and Bolnick 2016; Xiong et al. 2017; Tandonnet and Torres 2017; Ma et al. 2019). Previous studies have found conflicting results when testing the difference in power between WT and 3' (Table 1.1). For example, Tandonnet and Torres 2017 tested two treatments in a non-traditional model insect species (New World screw-worm fly) and detected 150 more DE contigs with 3' than with WT (Tandonnet and Torres 2017). However, they were unable to confidently annotate most of these contigs to gene models because they arose from 3' UTR regions. The authors suggested that 3' may have more power than WT for the same sequencing depth/cost but would be most appropriate for organisms with high-quality reference genomes. Xiong et al. 2017 tested three treatments (including untreated control) in a human commercial cell line and detected 15 percent more genes with WT libraries than 3' libraries (Xiong et al. 2017). However they stated that WT libraries were

more expensive and so experiments using 3' could afford greater numbers of biological replicates. Ma et al. 2019 tested two treatments in a single system and detected between 58-72 percent more DE genes with WT libraries than 3' libraries depending on read depth (Ma et al. 2019). While WT libraries did suffer from a length bias, they stated that WT libraries were likely better for non-model organisms.

Author Year	Species	Tissue	3' Reads	WT Reads	3' DE	WT DE
Ma 2019	Lab mice	Liver	NA	NA	1157	1982
Tandonnet & Torres 2017	Cochliomyia hominivorax	L2 instar larvae	40e6	16e6	148	140
Xiong 2017	Homo sapiens	Cardiomyocyte	4e6	6e6	4017	6819

Table 1.1: Summary of previous peer-reviewed research explicitly comparing RNA-seq priming methods.

We propose that many of the arguments in these papers are based on misconceptions about the relative advantages of the two priming methods for RNAseq: 1) Library construction costs: Several authors have stated that 3' libraries are less expensive (Lohman, Weber, and Bolnick 2016; Xiong et al. 2017). 2) Sequencing costs: authors have argued that 3' libraries can be sequenced to lower depth without sacrificing precision or statistical power, thus reducing costs (Lohman, Weber, and Bolnick 2016). Neither of these issues actually favors 3' libraries in general. While price differences certainly exist between commercial library construction kits, the difference between WT and 3' library protocols is the type of oligo used to initiate reverse transcription, which has little to no difference in material costs (Townsley et al. 2015).

The issue of sequencing costs differences between WT and 3' libraries is a more subtle topic. Sequencing mRNA using WT can result in multiple library molecules primed and sequenced from the same mRNA molecule. Therefore, it is possible that the same mRNA molecule will be counted multiple times (e.g. two fragments from the same mRNA are sequenced). The two reads from the same mRNA molecule do not provide independent estimates of the expression of the corresponding transcript, so the second read is useless for quantification and a waste of resources. If quantification software is unable to determine that the second read is actually from the same mRNA molecule as the first and double-counts the molecule, this is effectively pseudoreplication and will lead to elevated false positive rates under models that assume a Poisson process (e.g., *edgeR*, *DESeq2*). The 3' procedure is not subject to this issue because only a single library molecule can be reverse transcribed per transcript so no reads are wasted; every new read counts a new mRNA molecule. However, this "double-counting" of mRNA molecules is actually extremely uncommon in most WT experiments, unless the input RNA quantities are exceptionally small, and sequencing depth is exceptionally high. For example, a typical RNAseq library preparation protocol recommends starting with approximately 50ng of mRNA. If the average gene length of a study organism is 2kb, such as in Eukaryotes (Lynch and Marinov 2015), then a single sample would contain 46 billion mRNA molecules (The average molecular weight of a RNA base is 321.47 g/mol, so an average transcript will have a molecular weight of 643 Kg/mol. 50ng / 643Kg/mol * 6.022e23 = 46 billion molecules, calculations from: https://nebiocalculator.neb.com/#!/ssrnaamt, accessed September 1, 2021). RNAseq libraries are generally sequenced to a depth of 1-20 million reads. Since 3' libraries sequence either 0 or 1 read per mRNA molecule, with 20 million sequenced reads, ≈ 0.04% of the original mRNA

molecules will be sequenced. With WT libraries, if mRNA molecules are broken into 300nt fragments, there will be approximately 300 billion fragments, of which ≈ 0.007% will be sequenced. For any given mRNA molecule that has one read sequenced, the chance of a second being sequenced is also ≈ 0.007% for an mRNA molecule broken into 2 fragments, or up to ≈ 0.07% for an mRNA molecule broken into 11 fragments. This may well happen a few times in an RNAseq sample, but the vast majority of mRNA molecules will be represented by only a single sequenced read in WT libraries. Therefore, the effect of pseudoreplication in WT RNAseq is extremely low, and unlikely to affect analytical pipelines.

It is true that longer transcripts will have a higher probability of being counted (at least 1 time) than shorter transcripts in a WT library, meaning that longer transcripts will be measured with more precision by WT libraries and shorter transcripts will be measured with more precision with 3' libraries. For the shorter transcripts, 3' libraries could be sequenced to lower depth than WT libraries and achieve equal precision. For example, a transcript of 50% the average transcript length (formally, a weighted average where transcript lengths are weighted by their expression) could be measured equally by a 3' library sequenced to 50% the depth of an equivalent WT library. However this would come at the cost of much less precision for longer transcripts in the same library. A transcript of 2x the average transcript length would then receive only 25% of the read counts as the equivalent WT library. Also, the lack of length-bias does not mean that 3' libraries measure each transcript with equal precision – higher expressed transcripts will still be measured with more precision by either protocol, and transcripts vary by many orders of magnitude in expression level but only by several fold in length, meaning that

the contribution to heteroskedasticity of transcript length in WT libraries is not especially important relative to other factors.

While the analytical results above suggest that WT should not have higher costs or lower statistical power than 3' libraries, we aimed to test this result experimentally. One possible confounding factor could be the relative abilities to correctly map and annotate RNAseq reads from each method, which depends on genome size and annotation quality, and varies among species. We therefore designed a study to carefully measure the differences in statistical power between experiments using WT vs 3' libraries across three species with qualitatively different genomic and transcriptomic resources. Notably, unlike previous studies comparing WT and 3' sequencing methods, we isolated the technical variation inherent to the RNAseq measurement process from the biological variation in order to make the comparisons more precise. We find no basis in the common assumption that WT libraries generate pseudoreplication of fragments, increasing type 1 error. Furthermore, 3' libraries do not suffer from the same length bias as WT libraries which are more likely to over-represent longer transcripts. Finally, we find a small advantage in WT library's ability to detect DE genes, most prominent in our non-traditional model system, suggesting WT is a better option for non-traditional model organisms that are perhaps lacking in genome assembly and annotation quality and completeness compared to more traditional model species.

1.3 Methods

1.3.1 Experimental Design

We designed an experiment to quantify the effect of each library priming method on our ability to detect differentially expressed genes in 3 species. In the simplest experiment with two treatment groups, differential expression is tested by comparing a t-statistic, calculated as the ratio between an estimated change in expression divided by its standard error, to a null distribution. The library priming method could affect the power to detect differentially expressed genes by altering either the numerator (estimated effect size) or the denominator (within-group variance) of this ratio. If effect sizes are measured as log-fold-changes, it is unlikely that the two priming methods will differ in the estimated effect sizes (beyond sampling error) because longer genes should have proportionally more reads in both treatment groups with WT libraries. However, if pseudoreplication occurs in WT libraries, the within-group variance may be larger, leading to lower power. Therefore, our experiment was designed to compare the within-group variances of expression estimates between WT and 3' libraries.

In a typical RNAseq experiment, biological replicates of each treatment group are used to test the effect of a treatment on expression by measuring the within-group variance. Previous experimental comparisons of WT and 3' libraries have compared the results of experiments where some biological replicates are measured with WT libraries and others with 3' libraries (Lohman, Weber, and Bolnick 2016; Xiong et al. 2017; Tandonnet and Torres 2017; Ma et al. 2019). However, in such experiments, testing whether the within-group variation (denominator of the t-statistics) is greater for WT than for 3' libraries is under-powered because within-group variance is composed of both biological variation in mRNA levels among samples and measurement error in estimating mRNA levels inherent to the RNAseq process itself (including

Poisson sampling error, any random error introduced during the library preparation process itself, and any random error in read mapping or counting).

We therefore used an experimental design to directly compare the magnitudes of measurement error inherent to WT and 3' libraries. Rather than collecting multiple biological replicates from two treatment groups, we created a single sample from each of two treatments by pooling multiple samples together to get sufficient mRNA. We then split each pool into multiple identical RNA samples and used each to build either a WT or a 3' library (Figure 1.1). Therefore, all variation in expression estimates from libraries created from the same pool is entirely technical in origin and we can directly compare the magnitudes of the among-sample (within-pool) variance for each gene between libraries made with each priming method.

Below, we describe the detection of "significant" differences in expression between the two biological samples of each species as "differentially expressed genes", or DEG. Note that we are not claiming that these lists of genes are biologically interesting in any sense outside of this experiment because the treatments applied to each species are not biologically replicated. Nevertheless, the analysis of differences between the two groups of technical replicates is exactly the same as if they had been biological replicates, and since all genes will differ to some extent in expression between the two pools, we can use this design to evaluate the effect of priming method on statistical power.

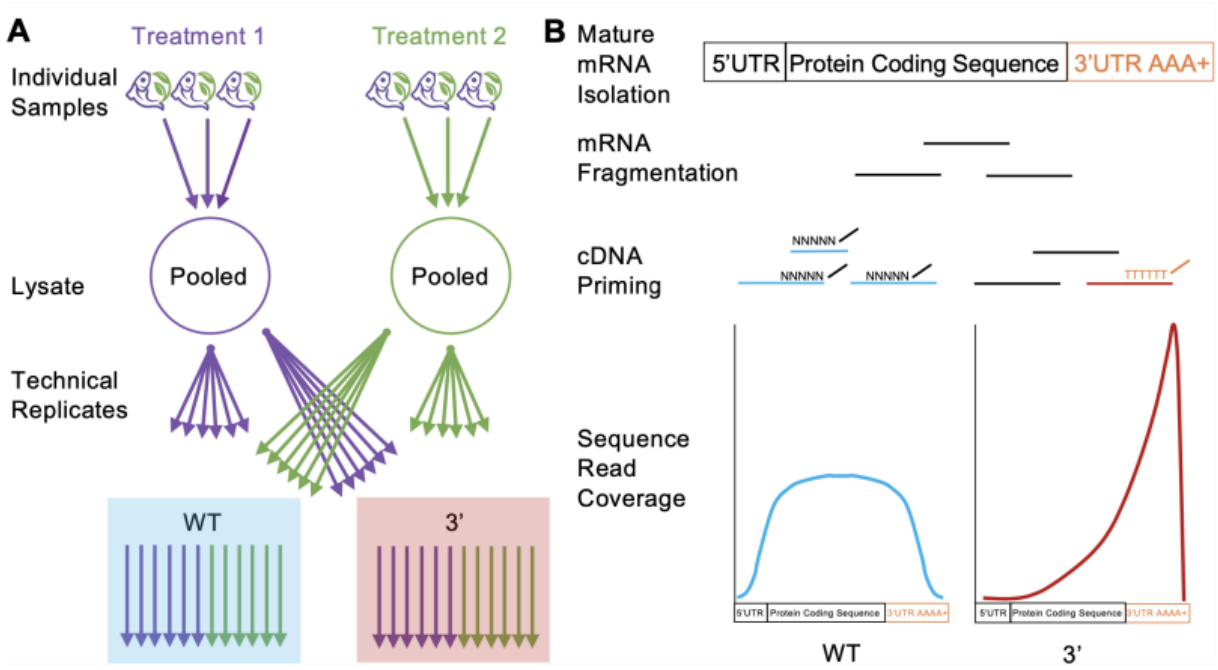

Figure 1.1: Graphical representation of experimental design to compare the magnitudes of technical noise in 3' and WT libraries. A) 3 biological replicates per species were combined into a common pool then split into 6 equimolar technical replicates per library construction method (WT or 3'). B) Each technical replicate was individually fragmented by high temperature (95◦C) and magnesium (Mg) ions and then first strand cDNA was synthesized either with poly-A (3') or random hexamer (WT) primers. 3' libraries are therefore 3' biased and generate sequence only in the 3' UTR or 3' CDS. WT libraries are less biased along the mRNA molecule

1.3.2 Biological Samples

Three species were selected for testing RNAseq library priming methods: Arabidopsis (*Arabidopsis thaliana*), Maize (*Zea mays spp. mays*), and Pacific herring (*Clupea pallasii*). We used genetically homozygous genotypes of Arabidopsis (Col-0) and Maize (B73), and samples from a single wild population (Puget Sound, WA) of Pacific herring. The treatment groups for each species were as follows: Arabidopsis samples were collected at two different times of day; Maize samples were taken from two different growing temperatures; and Pacific herring samples were immediately late-stage (pre-hatch) embryos that had been exposed to crude oil and control (no oil) treatments during embryogenesis. Male and female adult Pacific herring

gonads were collected from Puget Sound, WA by the Washington Department of Fish and Wildlife in 2017, and gametes were combined in the lab to form embryos for exposure experiments.

1.3.3 mRNA extraction and creating technical replicates

All samples were stored at -80 ◦C. Tissue samples were homogenized using stainless steel beads in the SPEX Geno/Grinder (Metuchen, NJ, USA) at 1200 RPM for 1 minute. Pacific herring samples required a second round of homogenization at the same settings. Ground lysates were suspended in 200ul of Lysis binding buffer (Townsley et al. 2015), incubated at room temperature for 10 minutes, centrifuged for 10 minutes at 13,000rpm, and the supernatant, referred to as cleared lysate, was retained.

The cleared lysate for each treatment group was combined to create a single mRNA pool with sufficient mRNA content to be subsequently divided into 6 200µl identical samples of mRNA to create 6 technical replicates per treatment group. mRNA was extracted separately from each technical replicate using oligo (dT)25 beads (DYNABEADS directTM) to enrich for polyadenylated mRNA.

1.3.4 RNA-seq library prep

We prepared strand specific RNA-seq libraries using the BRaD-seq protocol (Townsley et al. 2015), and used 14 cycles of PCR at the enrichment phase. Library preparations for all species and for both priming methods were identical, except for the primer used for first strand cDNA

synthesis. In this study we use a 3' priming method (3') to refer to libraries made with an oligo dT primer–which primes mRNA transcripts from the 3' end–and whole transcript (WT) to refer to libraries made using a random hexamer to prime fragments from along the length of the transcript. For 3' libraries, we used the primer: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTTTTTTTTTTTTTTTTTTT, and for WT libraries, we used GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNN, where N represents a random nucleotide as suggested in (Townsley et al. 2015). Finished libraries were quantified using the Quant-iTTM PicoGreen dsDNA high sensitivity kit, and normalized to 1ng/ul. We took 2ng per library and multiplexed the 36 samples for sequencing. The libraries were sequenced in 2 lanes of an Illumina HiSeq X platform, generating a mean of 6.1 million paired reads per sample. Raw reads were quality checked with FastQC v.0.11.5 (Andrews 2010). Low quality reads (q<20), adapters, and reads less than 25-bp were removed using Trimmomatic v.0.36 (Bolger, Lohse, and Usadel 2014).

1.3.5 Genomic resources

Assembled reference genomes, and gene transfer format (GTF) files were downloaded from NCBI (https://www.ncbi.nlm.nih.gov/sra) (O'Leary et al. 2016). We used the Arabidopsis accession number GCF 000001735.4 version T AIR10.1; the maize accession number GCF 000005005.2 version B73 RefGen v4. Pacific herring does not yet have an assembled genome, therefore we used the reference genome from a closely related species, Clupea harengus (Atlantic herring), accession number GCA 900700415.1 version Ch v2.0.2. The genomes varied in size and assembly and annotation quality. Arabidopsis has the highest quality assembly and

annotation and smallest genome, maize has the largest genome, and herring has the lowest quality assembly and annotation (Table 1.2).

	Metric	Atlantic herring	Arabidopsis	Maize
1	Scaffold N50	1,842,407	23,459,830	10,679,170
2	Contig N50	21,338	11,194,537	1,279,870
3	Percent complete	84.7	N/A	N/A
4	Genome coverage	72.0x	N/A	65.0x
5	Total size	808,340,743	119,668,634	2,134,373,047
6	Release date	2019	2015	2017

Table 1.2: Summary data of reference genome quality for each species used in this study.
Source of the reference genomes is NCBI (ncbi.nlm.nih.gov\sra).

1.3.6 RNA read mapping, quantification

Trimmed reads were mapped to reference genomes with *Hisat2* (Kim, Langmead, and Salzberg 2015) using default settings. The *featureCounts* function from Subread (Liao, Smyth, and Shi 2013) was used to count transcript features from all samples, and sum the counts to the gene level.

1.3.7 Differential expression pipeline

We developed a pipeline to detect differential expression that would account for variation in read depth across individuals, and bootstrapped the analyses to determine whether the results were consistent. Our pipeline was designed to reflect common experimental methods used in RNA-seq projects.

Samples with low total read counts (less than 100k reads), were removed from analysis. Since the remaining samples still varied in read counts, we randomly downsampled without replacement to the sample with the lowest number of mapped reads. All analyses were repeated 5 times with different downsampled datasets. After down-sampling, WT and 3' libraries of each species were processed independently through our differential expression pipeline. The downsampled gene count data were normalized using the weighted trimmed mean of M-values (TMM), using the *calcNormFactors* function in *edgeR* (Robinson, McCarthy, and Smyth 2010). The *voom* function (Robinson, McCarthy, and Smyth 2010; Law et al. 2014) in the *limma* package (Ritchie et al. 2015) was used to convert read counts to log2-counts per million (log2CPM), and estimate mean-variance trends to account for heteroscedasticity when testing for differential expression between the mRNA pools. The *lmFit* function from *limma* was used to fit a model using the pool ID as a predictor variable. The ebayes function was used to perform an empirical bayes shrinkage, and estimate an adjusted P-value after correcting for multiple testing using Benjamini-Hochberg correction.

1.4 Results

We sequenced the multiplexed library on two lanes of Illumina HiSeq X platform using 150 bp paired-end reads. Previous work has shown that read length (50, 100, 150 bp) has no effect on the detection of differentially expressed genes (Chhangawala et al. 2015). The mean number of reads, trimming results, and mapping rate to the reference genome varied by species and library prep method (Table 1.3). Overall, 3' libraries had more raw reads prior to trimming in

Arabidopsis and Maize, but fewer in Pacific Herring. Since this effect would necessarily go away if all libraries were of the same type (because the total number of reads per lane is constant), we excluded variation in raw read counts in downstream analysis by down-sampling all libraries to the same total read number. For all three species, an average of 27% 3' raw reads were trimmed (removed due to low quality), while 7% were trimmed in WT. Expression quantified using featureCounts resulted in approximately 15k, 18k, and 8k genes with greater than 5 reads in more than half of individuals identified in maize, arabidopsis, and Pacific herring respectively (Table 1.4).

	Species	Priming	Raw reads	Trimmed	Reads remain	Mapping
1	arabidopsis	3'	10,187,619	24.00	7,708,045	98.00
2	arabidopsis	WT	6,095,645	7.00	5,733,226	99.00
3	herring	3'	3,598,875	27.00	2,588,210	95.00
4	herring	WT	3,798,496	7.00	3,562,483	90.00
5	maize	3'	17,585,885	31.00	11,992,257	93.00
6	maize	WT	11,342,424	7.00	10,628,689	96.00

Table 1.3: Sequencing results for the number of raw, processed, and percent of uniquely mapped reads for each RNA-seq method and sample condition.

To explore if WT or 3' libraries differed significantly in either estimates of treatment effects, or precision (within-group standard errors of estimated expression), we first assessed whether the estimates of treatment effects differed between WT and 3' libraries. The difference in estimated logFC was not biased above or below zero for either low or high expressed genes (Figure 2A), or for short of long genes (Figure 2B). We next assessed whether the standard error

of effect size estimates was larger for WT libraries than 3' libraries. Standard errors trended lower for higher expressed genes with both library types, but there was no significant difference between library types in standard error at any expression level (Figure 1.2C). However, standard errors were lower for WT libraries for longer transcripts and higher for shorter transcripts (Figure 1.2D). This corresponded with greater read counts for transcripts with lengths ≈ 1800bp in WT libraries, and lower read counts for shorter transcripts (Figure 1.2E). Putting this all together we found no average difference in statistical power for discovering differentially expressed genes between WT and 3' libraries as a function of expression level (Figure 1.3A). Instead, power was higher for longer genes and lower for shorter genes, as expected (Figure 1.3B). This means that the distribution of transcript lengths as a function of expression levels determines whether WT or 3' libraries are likely to discover more differentially expressed genes. For all three species that we studied, WT libraries found 1-5% more differentially expressed genes than 3' libraries (Table 1.4), with the advantage of WT over 3' most evident in herring.

	Species	Primer	Total	DE	DE (%)
1	arabidopsis	3'	18,688	3,398	18.18
2	arabidopsis	WT	18,142	3,459	19.06
3	herring	3'	7,783	388	4.98
4	herring	WT	9,035	811	8.98
5	maize	3'	15,410	5,373	34.87
6	maize	WT	15,361	5,814	37.87

Table 1.4: Summary of results from analysis for each species and primer. Total number of analyzable genes (genes with 5 or more reads in half of individuals), DE = total number of differentially expressed

genes and %DE = the percent of total differentially expressed genes divided by the total number of analyzable genes. Differential expression threshold was genes below an FDR level of 0.05.

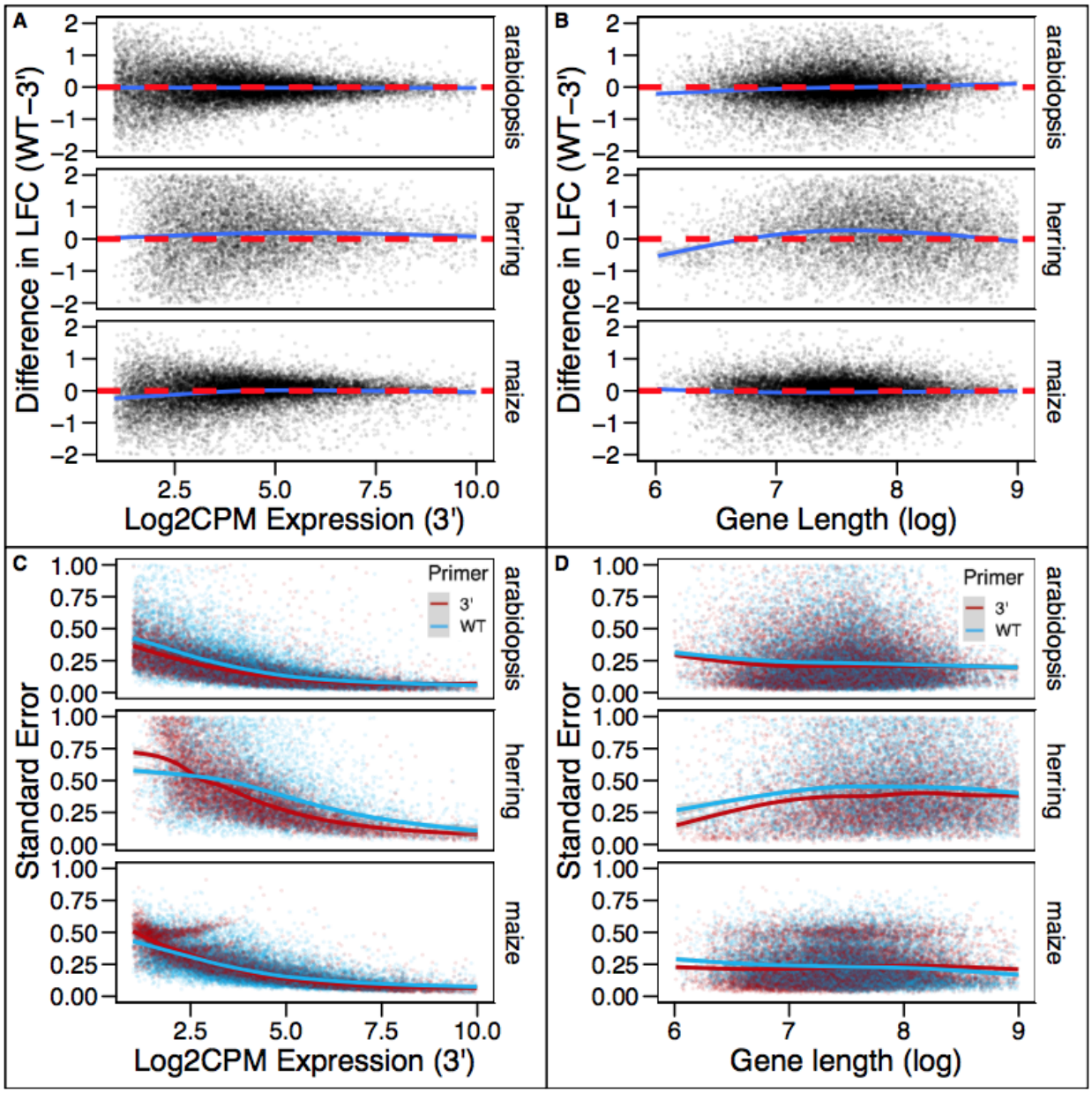

Figure 1.2: Standard errors but not effect sizes differ between WT and 3' libraries as a function of gene length but not gene expression. A,B) Estimated log-foldchanges (LFC) were similar between WT and 3' libraries across genes as a function of the A) gene's expression (measured as log2 counts per million, log2 CPM, by the 3' library), B) or transcript length. Estimated standard errors were similar for the two library types as a function of C) expression level, log2 CPM), but lower for WT libraries for D) longer genes and higher for shorter genes. Each analysis was done separately for each species. Each point in each plot is a single gene, and the smooth curves were generated by the geom smooth function of *ggplot2*. X and Y-axes were truncated for visualization.

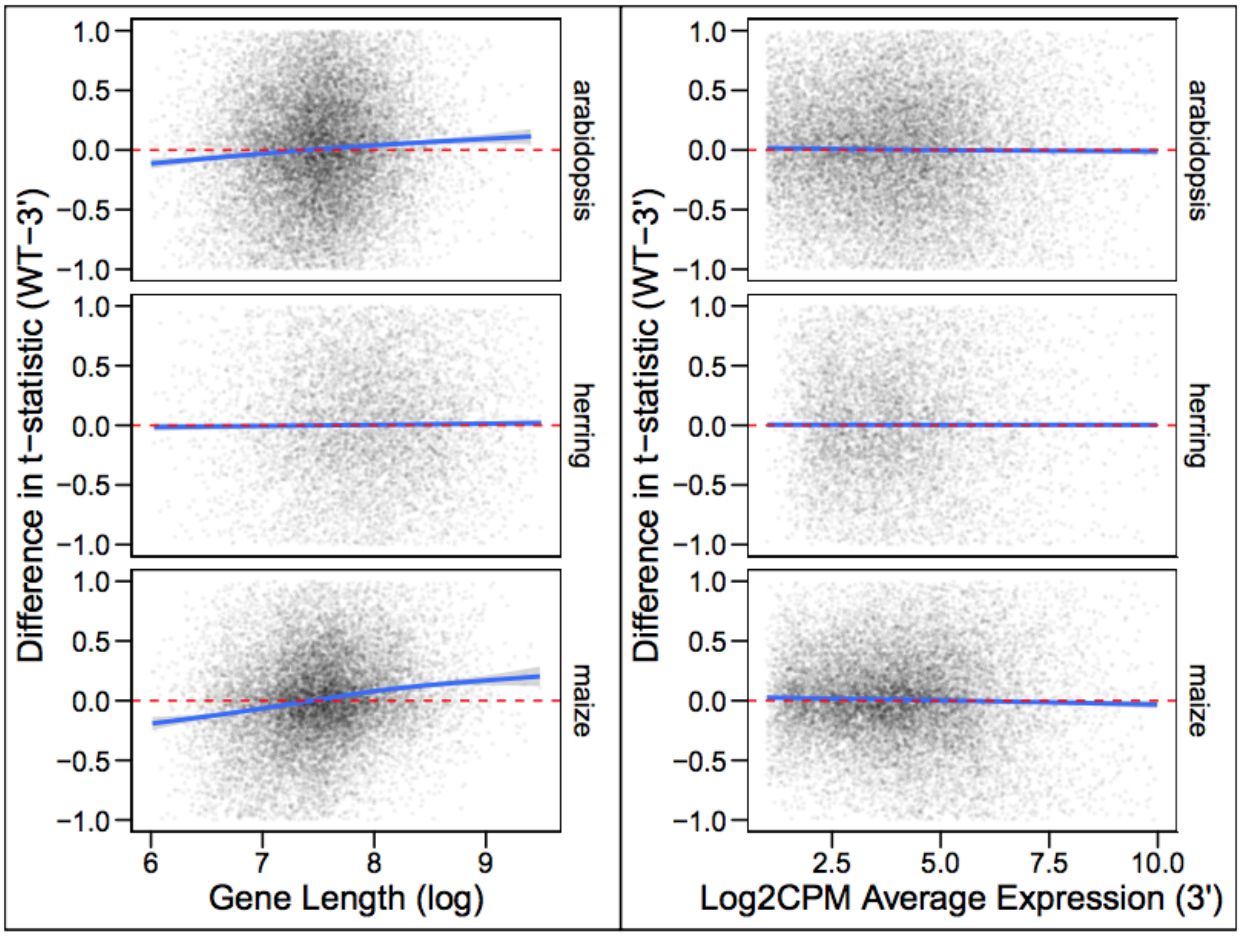

Figure 1.3: Differential expression analyses of WT libraries have higher power for longer genes, but not for higher-expressed genes than 3' libraries. A) Absolute values of t-statistics were higher on average for longer genes and lower for shorter genes using WT libraries, but did not differ between library priming methods as a function of the B) expression of each gene (measured as log2 counts per million, log2 CPM, by the 3' library. Each point in each plot is a single gene, and the smooth curves were generated by the geom smooth function of ggplot2. X and Y-axes were truncated for visualization.

1.5 Discussion

Our analysis and experimental results show that there is no overall benefit of 3' over WT libraries for expression quantification by RNAseq. WT libraries need not be more expensive than 3' libraries if comparable kits are available, and WT libraries do not tend to suffer from pseudoreplication or require more sequencing depth to achieve the same precision in expression quantification on average. However, length bias does impact the discovery of

differentially expressed genes, such that longer transcripts are more likely to be discovered by WT libraries, and shorter transcripts are more likely to be discovered by 3' libraries. If researchers knew beforehand which transcripts were of most interest, this could factor into the decision between WT and 3' libraries. But in general, if all transcripts are equally of interest, there is no clear winner between WT and 3' for differential expression analyses. However, WT libraries do have several advantages over 3' libraries beyond the simple assessment of differential expression at each gene locus. First, mRNA fragments are sequenced from across the entire exonic region of a transcript, which increases the chances of annotating reads (especially in non-traditional model species with low reference genome annotation and assembly quality), as well as alternative splicing and variant detection. The main advantage of 3' libraries is that they are not subject to the length bias of WT, which can impact downstream analysis such as gene set enrichment (Oshlack and Wakefield 2009; Mandelboum et al. 2019; "[No Title]" n.d.; Mi et al. 2012; Siepel et al. 2005). However, expression level also impacts the probability of detecting differential expression for both priming methods, and likely causes a much greater variation in power across genes than variation in length, so this issue is not limited to WT libraries.

Why then have most studies found that WT libraries have more statistical power to detect differentially expressed genes? We also found a similar result, although the difference is small (≈ 1%−5% more genes). The largest difference in numbers of differential gene expression between priming methods in our study was observed in herring. Pacific Herring is the only non-traditional model species in this study. With limited genomic resources, we mapped our

RNAseq reads to the publicly available Atlantic herring genome assembly. While mapping rates were slightly improved for herring 3' reads, herring WT libraries quantified over 1,000 additional expressed genes relative to 3' libraries (Table 1.3). This may be partially explained by the observation that 3' UTRs, where priming occurs for 3' library generation, are less conserved at the sequence level than protein coding sequences (Siepel et al. 2005). A common alternative for species without a reference is to assemble a transcriptome from the same RNAseq reads used for the differential expression analysis. However, this decision precludes using 3' priming methods which only retain the 3' end of gene transcripts (Tandonnet and Torres 2017). Therefore, WT is a better option for non-traditional model organisms lacking genomic resources. Although further experiments and analysis are necessary to pinpoint the source of increased statistical power inherent in WT libraries, our analysis demystifies the issue of psuedoreplication in WT priming methods and informs a general strategy for applying the wide-ranging utility of RNAseq in biological research.

Chapter 2

Uncovering the molecular mechanisms of trace levels of crude oil exposure on heart development in Pacific herring embryos

2.1 Abstract

Exposure to polycyclic aromatic hydrocarbons (PAHs) present in crude oil has been linked to cardiotoxicity in developing fish embryos. While the impact of PAHs on the developing heart has been well documented, the effects of sublethal concentrations and trace amounts of crude oil on heart development in embryos remain unclear. The aim of this study was to examine the short-term acute and long-term delayed toxicity of trace amounts of crude oil to Pacific herring embryos. To accomplish this, the morphology of the developing ventricle was used as a sensitive concentration-dependent phenotype. We detected significant cardiac injury at tissue concentrations at and above 64.5 ng/g TPAH. We detected oil-induced gene expression perturbations at concentrations as low as 13.4 ng/g, and at developmental times as early as 2 days post fertilization. The study systematically pinpointed dose-dependent genes and pathways involved in cardiogenesis and ventricle function from transcriptomic data, including a detailed transcriptomic time-series of early herring development. Though exposure ended three days prior to hatch, the expression of cardiac genes continued to be perturbed by

exposure for more than two months post-hatch. The study also expanded the scope of oil-induced developmental toxicity by characterizing the impacts of PAHs on immune system molecular pathways. The results of this study confirm that crude oil-induced cardiotoxicity is the canonical embryonic injury phenotype in Pacific herring. Perturbations in key cardiac-related genes and pathways responding to trace oil levels and time-points preceding organogenesis indicate early embryonic oil exposure sets the stage for larval cardiotoxicity. Notably, we observed significant dysregulation of *wnt11*, an inducer of cardiomyocyte specification, as well as heart proliferative genes, *bmp2b*, *bmp4,* in oil-exposed two day old embryos. We identify perturbed gene expression related to intracellular calcium homeostasis, inflammation and the heart injury response—initiating events that are upstream of the disrupted ventricular development that lingers after oil exposure. Overall, this study provides new insights into the molecular mechanisms whereby oil exposures perturb early life development in fish.

2.2 Introduction

The industrialized world is currently dependent on fossil fuel hydrocarbons. To meet global demand, exploration, extraction, transport, refinement and consumption of crude oil in offshore and nearshore marine environments increases the risks of accidental spills (National Academies of Sciences, Engineering, and Medicine. Oil in the Sea IV: Inputs, Fates, and Effects 2022). The effects of crude oil (petroleum) spills into the marine environment include habitat degradation and fouling, lethal and sublethal toxicity across all trophic levels and closure to fisheries. The long-term effects of oil spills on the environment from accidents such as the

Exxon Valdez are well studied, and more recent oil spill accidents such as the *Deepwater Horizon* explosion in 2010 have led to a surge of new research and literature on oil spill science (W. H. Pearson et al. 1999; Dubansky et al. 2013; Kujawinski et al. 2020).

Our understanding of the impacts of oil spills into marine environments is informed by experimental laboratory and field studies (Anderson et al. 1974). While crude oil is a complex mixture composed primarily of hydrocarbons, it has been shown that polycyclic aromatic hydrocarbons (PAHs), a minor fraction of hydrocarbons, have high water solubility and can bioconcentrate in animals exposed to crude oil (Meador et al. 1995). While the vulnerability of organisms depends on the species, life-stage, and toxic mechanisms involved, exposure to PAHs during embryogenesis across a wide range of teleost species results in similar cardiac developmental abnormalities, with severity of these impairments increasing with increasing PAH concentration (Incardona et al. 2013).

Oil exposure during a sensitive developmental window may result in delayed or persistent physiological responses or even mortality. Mechanisms that underlie perturbation during vulnerable early life stages are poorly understood. Additionally, molecular initiating events that link PAH exposure to cardiac organogenesis also remain poorly understood. As herring are exquisitely sensitive to oil-induced cardiac injury during embryogenesis, even in trace amounts, subsequently the lowest exposure with no effect has not been determined (Incardona et al. 2013, 2015).

Beyond the canonical embryo heart injury syndrome, less is understood about persistent physiological effects in novel pathways affected by early-life oil exposure. Oil exposure may impact multiple systems—including skeletal, immune, energy, neural, and hepatorenal (Xu et al. 2016; Khursigara, Johansen, and Esbaugh 2018; Bonatesta et al. 2022; Arkoosh et al. 1998). Some of these effects may be initiated if exposure coincides with the development of the corresponding organ. Transcriptomic approaches are useful for providing hypotheses for mechanisms that underlie these exposure outcomes, and as-yet-undiscovered perturbations to health and performance. For instance, AHR ligands, such as PAHs, modulate immune expression in fish (Segner et al. 2021). However it is unclear whether immunotoxicity is directly mediated through activation of the AHR pathway or through other signaling pathways. While the molecular and morphological events that underlie teleost cardiogenesis are well understood, little is known about how the immune system develops and matures in Pacific herring, and it is not clear how exposure to oil might disrupt this development and increase susceptibility to environmental pathogens later in life.

The present study aims to investigate the effects of crude oil exposure on Pacific herring development. Herring spawn in intertidal and shallow subtidal marine habitats, nearshore breeding grounds that are susceptible to oil marine spills (Hay 1985). They serve as an ideal model system due to their high sensitivity to contaminating crude oil during development and prominence in the literature on hydrocarbon toxicology (Hose et al. 1996; Brown, Norcross, and Short 1996a). Our study design involves examining the short-term acute and long-term delayed toxicity in Pacific herring embryos exposed to trace amounts of crude oil along a concentration

response gradient through embryonic, larval, and early juvenile stages of development. The objectives of the study focus on cardiac morphogenic processes in the developing ventricle, assembling a high-quality transcriptome to understand molecular mechanisms and identify potential toxicity phenotypes, quantifying and characterizing gene responses in cardiogenesis and ventricle function in embryos exposed to field relevant, very low-dose range of crude oil, and defining the transcriptional arc of innate immune system development.

2.3 Methods

2.3.1 Animal Care

Pacific herring (*Clupea pallasii*) were collected near Cedar Bay in Prince William Sound, Alaska. Collection of ripe herring by gill net occurred on May 9th in 2018. Ovaries and testes were dissected and stored in humidified whirl packs or petri dishes prior to transport to the Marrowstone Field Station (USGS Western Fisheries Research Center, Nordland, WA). Fertilizations were performed in water tables at the Marrowstone Field Station as described elsewhere (Griffin et al. 1998). Briefly, polyvinyl alcohol with Ca^{2+} and Mg^{2+} -free seawater was used to prevent clumping of eggs which were evenly distributed onto 24 sheets of 1 mm nylon mesh, each of which could hold 10-20 grams of embryos. At least 5 males were used for fertilization and gametes were evenly distributed amongst the egg-covered sheets. Fertilized eggs were left undisturbed for 1 hour and then transferred to 1.2 m diameter tanks with filtered seawater overnight for 8-12 hours. Fertilization rates were then assessed microscopically and oil exposure occurred on batches with rates above 75%. Crude oil exposure occurred between 1- and 10-days post fertilization (dpf) in filtered seawater at ambient

temperature. Temperature was recorded twice an hour in a downstream tank from the exposure system by a Hobo Water Temp Pro v2 sensor (Onset). Animal care and use was evaluated and approved by the Western Fisheries Research Center IACUC (protocol# 2008-51).

2.3.2 Oil Exposure

Prior to exposures, Alaskan North Slope crude oil (ANSCO) was weathered by heating to 60°C in a water bath until the volume decreased by 10%. The exposure system (SINTEF, Trondheim, Norway) produced a continuous flow-through of dispersed oil droplets as described in (Nordtug et al. 2011). It consisted of 24 10-L tanks with individual flow-through of filtered seawater (360 µL/min flow rate) at ambient temperature and ANSCO effluent (60 µL/hr injection rate). ANSCO effluent entered the system through a dispersion generator which produced microdroplets of oil which were pumped into a 2 L reservoir that was gravity-fed to a series of solenoid valves which controlled the ratios of effluent to filtered seawater into the exposure tanks. Six concentrations were generated ranging from 0.01 (no oil) to 3.5 µg/L ∑PAHs. Exposure experiments were conducted with four replicate exposure tanks per oil concentration. Fertilized eggs adhered to Nitrex sheets were distributed amongst the tanks (1 per tank) and monitored several times daily until 10 dpf at which time oil injection stopped.

2.3.3 Analysis of PAHs

Water samples were collected mid-water column and analyzed for 58 PAH analytes as described previously (Sloan et al. 2005). Briefly, PAH separations were conducted on a 60 m DB-

5 chromatography (GC) capillary column before detection on an electron impact mass spectrometer (MS) in selected-ion monitoring mode. Concentrations below the detection limit were reported as <LOQ (less than the limit of quantification). TPAH values were calculated from detected values only. For body burdens of TPAHs, 100 embryos at 1.5, 2, 3, 4, 10 dpf (n=4) were collected from each Nitrex sheet and stored at -80°C. Samples were analyzed for 49 PAH analytes utilizing liquid extraction prior to normal phase solid phase extraction (LE-SPE) as described elsewhere (Sørensen et al. 2016). Blank corrections were conducted for tissue PAHs when the blank value was higher than the detection limit (included: naphthalene, methylnaphthalenes, C3-BT, C2-PHE, and C2-PYR). Available PAH concentrations for water and tissue were screened for near-zero and near-LOQ values. PAHs were excluded from further analysis when high dose treatments (3.42 µg/L TPAHs) had >75% of detections below the LOQ.

2.3.4 Cardiovascular Imaging

At 10 dpf and hatch (12-14 dpf), 20 embryos/larvae from each tank (n=4) were immobilized with 0.01% MS 0.222 and mounted laterally in petri dishes with methyl cellulose or 1.5% agarose gels with slots molded by glass capillary tubes. Seawater was maintained between 9°C by a microscope stage chiller. For each larva, 10 second videos focusing on the atrium and ventricle were taken with Unibrain Fire-i400 1394 cameras (San Ramon, CA) at the highest magnification (6-8x) on Nikon SMZ-800 stereomicroscopes (Irvine, CA). BTV Pro software (Bensoftware.com) was used for imaging. To avoid confounding time-of-day with treatment effects, all treatment replicates were imaged evenly across an 8-hour period. Heart videos were

blinded for treatment and randomly analyzed for function and morphology using ImageJ software (NIH). Prior to measurements, video was assessed for appropriate quality and contrast and excluded if morphology of interest was not visible. At 10 dpf pericardial area, an indicator of edema, was measured by drawing a perimeter around the pericardium with the yolk as the posterior end. For larvae at hatch, posterior ventricle outgrowth was measured. Posterior ventricle outgrowth was measured by drawing a perimeter within the ventricle at diastole starting at the posterior side of the atrioventricular (AV) valve to the apex of the ventricle to assess ballooning of the ventricle beyond the AV valve (Figure 2.3). Linear regressions were performed to compare posterior ventricle outgrowth at hatch with increasing tissue concentration of TPAHs. A two-way ANOVA using R Studio (base package; version 1.4.1717) was used to determine the effect of body burden TPAHs with ventricle outgrowth area (mm2). When oil effects were present, a Tukey's post hoc test was used to compare means between treatments.

2.3.5 Transcriptome Assembly and Annotation

Currently, no complete reference transcriptome exists for Pacific herring. We built a high-quality, annotated reference transcriptome using Pacific Biosciences Iso-Seq technology to provide a template to test hypotheses related to molecularimpacts of oil exposure. Whole embryo and larva RNA from 5 developmental stages (4 dpf, 8 dpf, 33 dph, 61 dph) and seven tissues (gill, brain, heart, kidney, liver, gonad, skeletal muscle) from a single adult female were isolated with the RNeasy Plus Micro Kit (Qiagen), quantified by fluorometry with a Qubit (RNA

Broad Range Assay Kit, Molecular Probes) and quality assessed with a 2100 Agilent Bioanalyser (RNA Nanochip, Agilent Technologies). First-strand cDNA synthesis was performed with the TeloPrime Full-Length cDNA Amplification Kit (Lexogen) using the cap-dependent linker ligation method. cDNAs were quantified for each sample and equimolar concentrations were combined into a single pool prior to library prep. Single Molecule Real Time (SMRT) bell libraries were prepared by UC Davis DNA Technologies Core with the SMRTbell Express Template Prep Kit 2.0 (Pacific Biosciences). Sequencing was completed on the Sequel II System on a single SMRT Cell 8M. Polished Circular Consensus reads were demultiplexed, refined, and clustered into high quality reads representing isoform-specific transcripts using the Pacific Biosciences IsoSeq v3.1.1 default parameters. For de novo transcript annotation, we employed the dammit pipeline (Scott et al. 2019). Briefly, gene models are built with Transdecoder (Haas, https://github.com/TransDecoder/TransDecoder), which were annotated based on evidence from four protein databases (Pfam-A, Rfam, OrthoDB, Uniref90) and the reference genome from *Clupea harengus*, accession number GCA 900700415.1 version Ch v2.0.2. We then used the Benchmarking Universal Single-Copy Orthologs (BUSCO v3) software to compare the gene content in our *de novo* transcriptome with a lineage-specific data set (*Actinopterygii*; (Simão et al. 2015)). The complete isoform-aware assembled transcriptome consists of a total of 180,175 transcripts, a contig N50 of 3,663, BUSCO completeness score of 95%, and 75% of the transcripts annotated with gene IDs.

2.3.6 RNA Sequencing

Two pools of embryos (~30 embryos per pool) were sampled from each of 4 replicate tanks per treatment, beginning at 2 dpf and every other day until 10 dpf, for 5 embryonic time-points. After oil exposure, three treatment groups (Control, Medium Low, and Medium High) were followed for long-term assessment. Included treatment tanks (n=4) were combined and randomly split into 2 grow-out tanks with clean seawater. Larva were sampled at time of hatch (1 dph) and followed up at 50, 64, and 78 dph. Similar to embryos, 3 pools (~30 larva per pool) of larva were sampled from each of 2 replicate tanks at 1 dph (n=30) per treatment. Single larva were used for subsequent post-hatch samples. All post-hatch samples were sampled in sets of 3 per treatment tank (n=2 x 3). Pacific herring samples were stored at -80°C prior to homogenization using 1.4 mm ceramic beads (Genaxxon Bioscience) on a TissueLyser (Qiagen). Ground lysates were suspended in 200 µl of Lysis binding buffer (Townsley et al. 2015), incubated at room temperature for 10 minutes, centrifuged for 10 minutes at 13,000 rpm, and the supernatant was retained. mRNA was extracted separately from each sample using oligo (dT) 25 beads (DYNABEADS directTM) to enrich for polyadenylated mRNA. Strand specific RNA-seq libraries were prepared using the BRaD-seq protocol (Townsley et al. 2015), and used 14 cycles of PCR at the enrichment phase. Library preparations for all populations were identical. A random hexamer was used to prime fragments from along the length of the transcript:

GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNN, where *N* represents a random nucleotide (Townsley et al., 2015). Finished libraries were quantified using the Quant-iTTM PicoGreen dsDNA high sensitivity kit (ThermoFisher), and normalized to 1ng/µl. We used 2ng per library and multiplexed up to 96 samples for sequencing. The libraries were sequenced in 4 lanes of an Illumina HiSeq X platform (Illumina, Inc.), generating an average of 10.8 million

paired reads per sample. Raw reads were quality checked with FastQC v.0.11.5 (Andrews 2010). Low quality reads (q<20), adapters, and reads less than 25-bp were removed using Trimmomatic v.0.36 (Bolger, Lohse, and Usadel 2014).

2.3.7 Read Mapping and Quantification

Trimmed reads were mapped to a reference transcriptome and mapped reads quantified with Salmon v 1.3.0 (Patro et al. 2017) using the reference transcriptome (section 2.3.5) as an index in mapping-based mode. Average mapping rate for all paired-end read samples was 70.92%. The length-scaled TPM function in *tximport* was used to sum the transcript counts to the gene level (Soneson, Love, and Robinson 2015). The gene count data were normalized using the weighted trimmed mean of M-values (TMM) and converted to log2-counts per million (log2CPM) using the limma R package (Law et al. 2014; Ritchie et al. 2015).

2.3.8 Differential Expression Analysis

We transformed read counts to stabilize the variance so that it was independent of the mean (where the variance neither increases systematically nor decreases systematically), according to Rocke et al. (Rocke et al., 2015). This was accomplished by regressing the gene-and-treatment variance on the gene-and-treatment mean for all samples. Once the mean variance relationship was defined, we estimated a transformation to the counts for which the slope was near zero (0.09e-8) which, practically speaking, stabilizes the variance, and applied this transformation across all gene counts. Low-abundance transcripts were filtered using the function *filterByExpr*

where the minimum number of counts per transcript per day per treatment was set to 35 and the minimum sum of counts within a treatment was 500 (Robinson and Oshlack 2010). The set minimum number of counts per transcript was determined by the smallest value that best normalized total read counts (library sizes) across all samples. Counts were further normalized using *CalcNormFactors*, method=TMM. Normalized counts by sample were first graphed using multidimensional scaling (MDS) to visually identify primary sources of variance. The 96 well plate the sample was on for library prep (batch effect) was found to be a significant source of variance across the. Due to this plate effect we chose to run a linear model in R with plate as the random effect. *lm(Count ~ Time * Oil + (1 | plate)).* The linear model was set up with embryonic developmental time (2, 4, 6, 8, 10 dpf) as a factor, oil treatment a cofactor (expressed as mean total PAHs per dry weight at day 10 per treatment), while plate effect was a random effect in the model. All p-values were corrected with the False Discovery rate (FDR) algorithm. Using this output we were able to select gene sets that have a significant developmental time effect, oil dose effect, and time by dose interaction effect. Differentially expressed genes in oil-exposed embryos, relative to unoiled control embryos, were defined as having higher or lower levels of expression (FDR <0.1). Individual genes were categorized across development based on their dose-response at each time-point. This was achieved by calculating the slope of the regression at each time-point for genes with a significant oil-dose by time interaction effect (FDR<0.1), comparing absolute slope values across time-points within each gene, then assigning a gene to a time-point based on the slope with the greatest value (i.e., largest dose response). Limma's function *removeBatchEffect* was used to remove the library plate effect for all visual representation of the data (Limma; Ritchie et al., 2015). Further

visualizations utilized the *pheatmap* package (Kolde and Vilo 2015) in R to generate heat maps, the *factoextra* package (Kassambara and Mundt, 2020) to generate PCA plots, and the *ggballoonplot* package v. 0.6.0 to generate figure 2.5A and figure 2.7.

2.3.9 Gene Coexpression Network Analysis

Weighted-gene correlation network analysis (WGCNA) was used to identify highly co-expressed clusters of genes (modules) within the post-hatch (1, 50, 64, 78 dph) herring whole-transcriptome data (Langfelder and Horvath 2008). Normalized counts were passed on to the WGCNA (v1.71) package in R. Standard WGCNA analysis was performed on each dataset following the steps outlined in the WGCNA tutorials with recommended parameter settings. The power-law soft-threshold was estimated using the *pickSoftThreshold* function in the online FAQ, as has become the standard in most studies. To find modules highly correlated to the oil dose response, the correlation coefficient for all module eigengenes was calculated using Pearson's correlation. Modules with adjusted p values less than 0.1 were kept for enrichment analysis.

2.3.10 Functional Gene Enrichment Analysis

Gene functional enrichment analysis was performed with *ShinyGO V0.77 (Ge, Jung, and Yao 2020)* using the Gene Ontology (GO) resource, and the Kyoto Encyclopedia Genes and Genomes (KEGG) pathways database. Terms with FDR-corrected p values of < 0.1 were considered significantly enriched within modules.

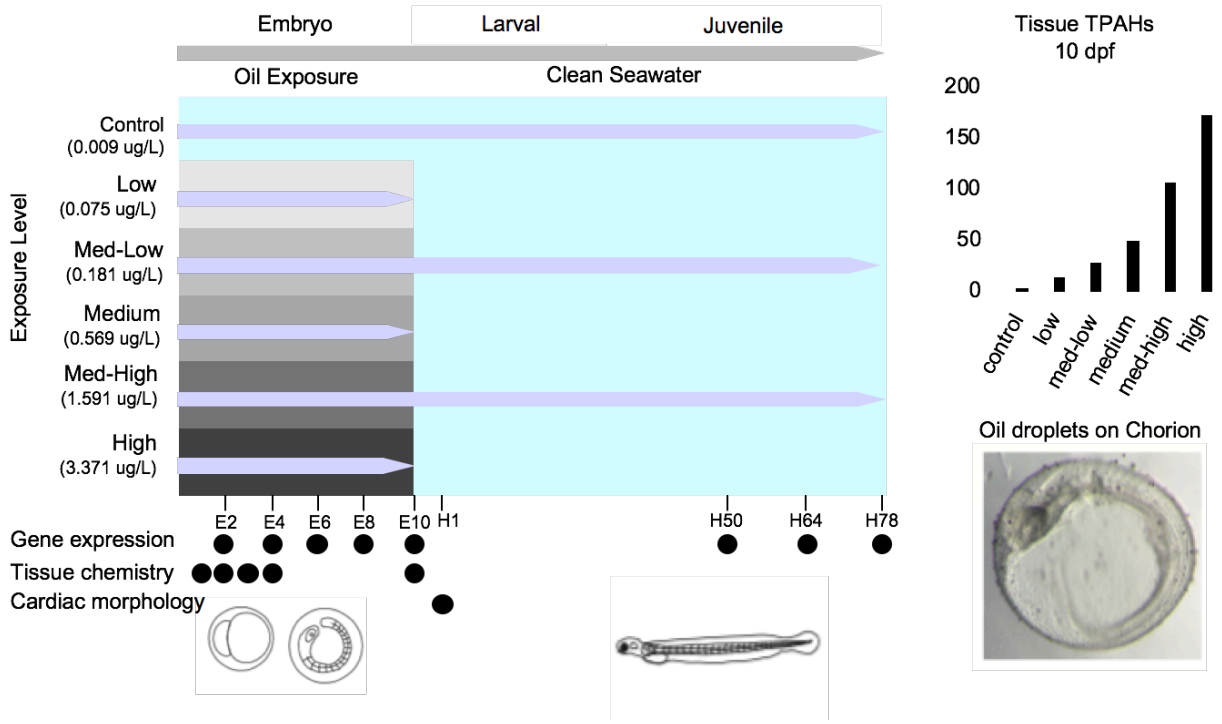

Figure 2.1: Comparison of crude oil concentrations across molecular and morphometric toxicity indicators during Pacific herring development. A) Schematic depicting experimental design showing length of exposure, time of hatching, and period of imaging/tissue collection for oil exposure experiment. E2-E10 represents embryo developmental time-points included in this study, and H1-H78 represents post-hatch time-points. Experimental design consisted of 6 treatment groups (Control-High) with corresponding water PAH concentrations at the onset of herring exposures (ug/L). Three treatment groups (Control, Med-Low, Med-High) were carried through to juvenile stage. B) The uptake of TPAH tissue concentrations across oil levels measured at the final day of exposure (E10). C) Evidence of oil microdroplets adhering to herring chorion, imaged at 5 dpf.

2.4 Results and Discussion

2.4.1 Measured PAH concentrations in exposure water and herring tissue

The *Exxon Valdez* oil spill released 37,000 metric tons of Alaskan North Slope crude oil (ANSCO) into Prince William Sound, AK in 1989 during the annual Pacific herring spawn event. Water column exposure shortly after the spill ranged from 0-42 ppb (µg/L). In this study, Pacific herring embryos were continuously exposed at the low end of environmentally relevant

concentrations of ANSCO across a serial dose range of TPAH concentrations of 0.06 ± 0.004 µg/L (low dose), 0.15 ± 0.006 µg/L (medium low dose), 0.53 ± 0.02 µg/L (medium dose), 1.42 ± 0.09 µg/L (medium high dose), and 3.42 ± 0.05 µg/L (high dose), with clean seawater controls at 0.01 ± 0.0009 µg/L (Table 2.1). The embryonic exposure began at 1 dpf and ended shortly after the end of organogenesis at 10 dpf, just prior to hatch (Figure 1.1). By the end of exposure (day 10), embryos accumulated 13.4 ± 1.6, 27.5 ± 2.0, 238 ± 27, 64.5 ± 3.0, 105.0 ± 11.9, and 171.9 ± 15.0 ng/g TPAHs across low to high dose, respectively (Table 2.1). After oil exposure, all embryos hatched and were grown out in clean seawater for 78 days post-hatch (dph) to monitor lingering physiological effects of the oil. This dose range captures concentrations in PWS nearshore herring spawning habitats in 1989 and 1990. PAH body burdens were measured within the highest crude oil concentration treatment at 6 time-points throughout embryogenesis. TPAHs in fish tissue increased to a maximum of 696.2 ng/g at 4 dpf before decreasing to 171.9 ng/g at 10 dpf (Figure 2.2). This suggests effective PAH metabolism and elimination is initiated after 4 dpf, similar to what has been observed in other species such as Atlantic cod and haddock (Sørensen et al. 2017).

Oil Dose	Water	Embryo Tissue
Control	0.01 ± 0.0009	2.7 ± 0.7
Low	0.06 ± 0.004	13.4 ± 1.6
Medium Low	0.15 ± 0.006	27.5 ± 2.0

Medium	0.53 ± 0.02	64.5 ± 3.0
Medium High	1.42 ± 0.09	105.0 ± 11.9
High	3.42 ± 0.05	171.9 ± 15.0

Table 2.1: TPAH concentrations of aqueous (µg/L) and embryo concentrations (ng/g wet wt.) at 10 dpf for Pacific herring. Data presented as mean ± SEM (n=4).

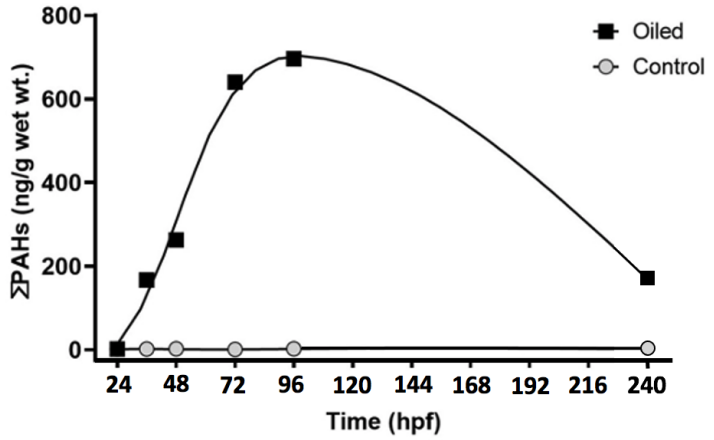

Figure 2.2: The uptake of TPAHs over the course of exposure (high dose only). Between 24 and 96 hours post fertilization (hpf) TPAHs rapidly accumulate, peak at 696.2 ng/g at 96 hpf, then decline to 171.9 ng/g by 240 hpf.

2.4.2 Morphological phenotypes associated with crude oil exposure

The developing heart is the primary target organ for crude oil toxicity in the developing fish embryo. Even sublethal low-dose concentrations can disrupt cardiogenesis and lead to heart failure. Pericardial edema is routinely used to assess cardiotoxicity during embryonic crude oil exposures of fish. The trace oil exposures used in this study, even at the highest cumulative tissue concentrations (171.9 ng/g), were too low to induce significant edema. Developmentally

downstream of edema incidence is an emerging target of crude oil toxicity in the developing heart, ventricular ballooning, the posterior outgrowth of the ventricle during larvagenesis. Anatomically, the ventricle begins to protrude ventrally to the atrium as an outgrowth or "ballooning". In contrast to edema, we observed a post-oil exposure dose-dependent decrease in ventricular ballooning at 1 dph, beginning with the medium dose (Figure 2.3). From the lowest to highest dose, percent reduction in the ventricular outgrowth was 3.7, 2.8, 15.1, 17.1, 45.9%, respectively. Ventricle ballooning was responsive to oil exposure, with significant reduction beginning at the medium dose relative to no oil control. Recent research has shown that oil perturbed ventricular ballooning recovers in later larval stages but transitions to dose-dependent trabeculation with indications of abnormal growth of cardiomyocytes in juveniles (Incardona et al. 2021). These morphological disruptions may be causally connected to the later-in-life cardiac hypertrophy and reduction in cardio-respiratory performance (Incardona et al. 2015).

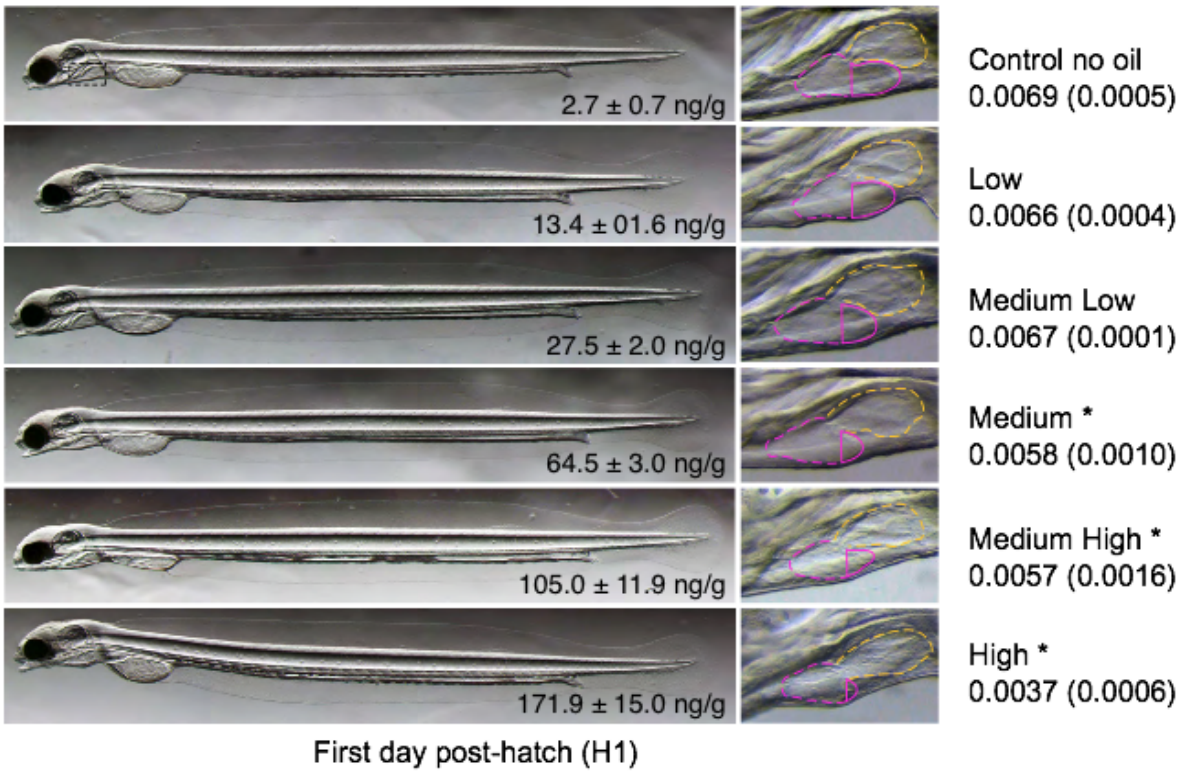

Figure 2.3: Posterior ventricular outgrowth, or ballooning, at hatch for oil-exposed Pacific herring. Oil effect by exposure level is depicted as a reduction of ventricular ballooning. Lateral images (right panel) were taken for whole body morphology at 2x magnification. Right insets present ballooning (solid magenta line) within the context of a Pacific herring atrium (yellow hash line) and ventricle (magenta hash line) at 8x magnification. Ventricular ballooning varied and was negatively correlated in an increasing dose-dependent manner (significant oil effect marked by asterisk if $p<0.05$). Ventricular outgrowth (ballooning) measurements listed below doses were averaged (mm2) across 4 replicates accompanied by standard deviation in parenthesis.

2.4.3 Developmental Atlas of Gene Expression

To understand the molecular mechanisms underlying the oil injury response in Pacific herring embryos and larvae, we used RNA sequencing to evaluate whole animal transcriptional response. In order to delineate the effects of oil perturbation on gene expression patterns throughout developmental stage and oil dose, we established a developmental framework for studying the temporal dynamics of gene expression (Figure 2.4). Across 10 timepoints spanning embryogenesis we detected 23,353 genes with a significant main effect of developmental time

on expression (FDR<0.05). These developmental genes represent 74.7% of the annotated genes that were included in the analysis. The heatmap in Figure 2.4A visualizes changes in gene expression throughout development. Gene expression was organized by time of peak expression. The bimodal distribution (Figure 2.4A top panel) reflects developmental genes with pronounced expression differences between early and late stage embryos. Genes that have highest expression in embryogenesis (peak expression at 2 dpf: Module A in Figure 2.4A) are enriched for the KEGG pathway *RNA transport and protein synthesis*. Genes in mid to early embryogenesis (peak expression at 4 dpf: Module B in Figure 2.4A) are enriched for KEGG pathways *DNA replication*. Genes in middle-to-late embryogenesis (peak expression at 6 dpf: Module C, Figure 2.4A) are enriched for KEGG pathways involved in *Cell differentiation* and *Energy production*. Dynamic organ growth continues into 8 dpf (peak expression: Module D, Figure 2.4A) with genes enriched for the KEGG pathway *Branched-chain amino acid metabolism*, which is commonly metabolized in skeletal and cardiac muscle (Lu et al. 2007). In addition, genes enriching the KEGG *Extracellular matrix receptor signaling* pathway were observed at 8 dpf. The extracellular matrix is composed of macromolecules acting as a scaffold during organ morphogenesis, particularly the growing heart (Garcia-Puig et al. 2019). The final timepoint in this analysis is 10 dpf (peak expression at 10 dpf: Module E, Figure 2.4A), two days before hatch. Gene that are enriched for the KEGG pathway *Oxidative phosphorylation* define this late embryo period. The Ox-Phos pathway includes the production of ATP by oxidation of metabolic fuels in the mitochondria, particularly to maintain the highly oxidative metabolic capabilities of muscle tissues in the heart. This Atlas sets a baseline of normal gene expression

patterns across time that provide context for understanding how oil exposure may perturb development during these sensitive early life stages.

2.4.4 Effects of oil on developmental gene expression

Considering the 5 core temporal modules of developmental gene expression (modules A, B, C, D, and E; Figure 2.4A), we visualized trajectories of expression change through developmental time for each module (through principal component space), and tested for perturbations by oil exposure. An oil effect was considered for genes showing a significant oil by developmental time interaction (FDR <0.1). Module C was significantly enriched for genes with a significant oil x time interaction (p-value=0.019, Fold Enrichment=1.32), where perturbation by oil was most apparent at 4 dpf, especially at the highest doses (Figure 2.4B). Module C genes were particularly perturbed by oil exposure during the earliest developmental times (2 dpf and 4 dpf; Figure 2.4B). Module C genes were enriched for a number of GO terms, including *Angiogenesis*, and *Compound catabolic processes*, which is involved in the process that breaks down PAHs. However, overrepresentation analysis of functional pathways showed the strongest enrichment for the GO Biological Process Cardiac muscle myosin thick filament assembly (Figure 2.4C), which correlates to early-stage circulatory-dependent morphogenesis, particularly cardiogenesis (England and Loughna 2013). This supports previous findings on an enrichment of cardiac related genes at 6 dpf in oil exposed herring embryos (Incardona et al. 2021). The current observation strongly suggests that a sensitive developmental time for cardiogenesis occurs between 4 and 6 dpf in herring. Genes in Module C that were significantly differentially expressed starting at the medium dose relative to controls ($p<0.05$) include those associated

with xenobiotic response, angiogenesis, and innate immune function *cyp1a, cyp1b1, si:ch73-334d15.2, pdap1*, respectively (Hsu et al. 2021; Kramer et al. 2020).

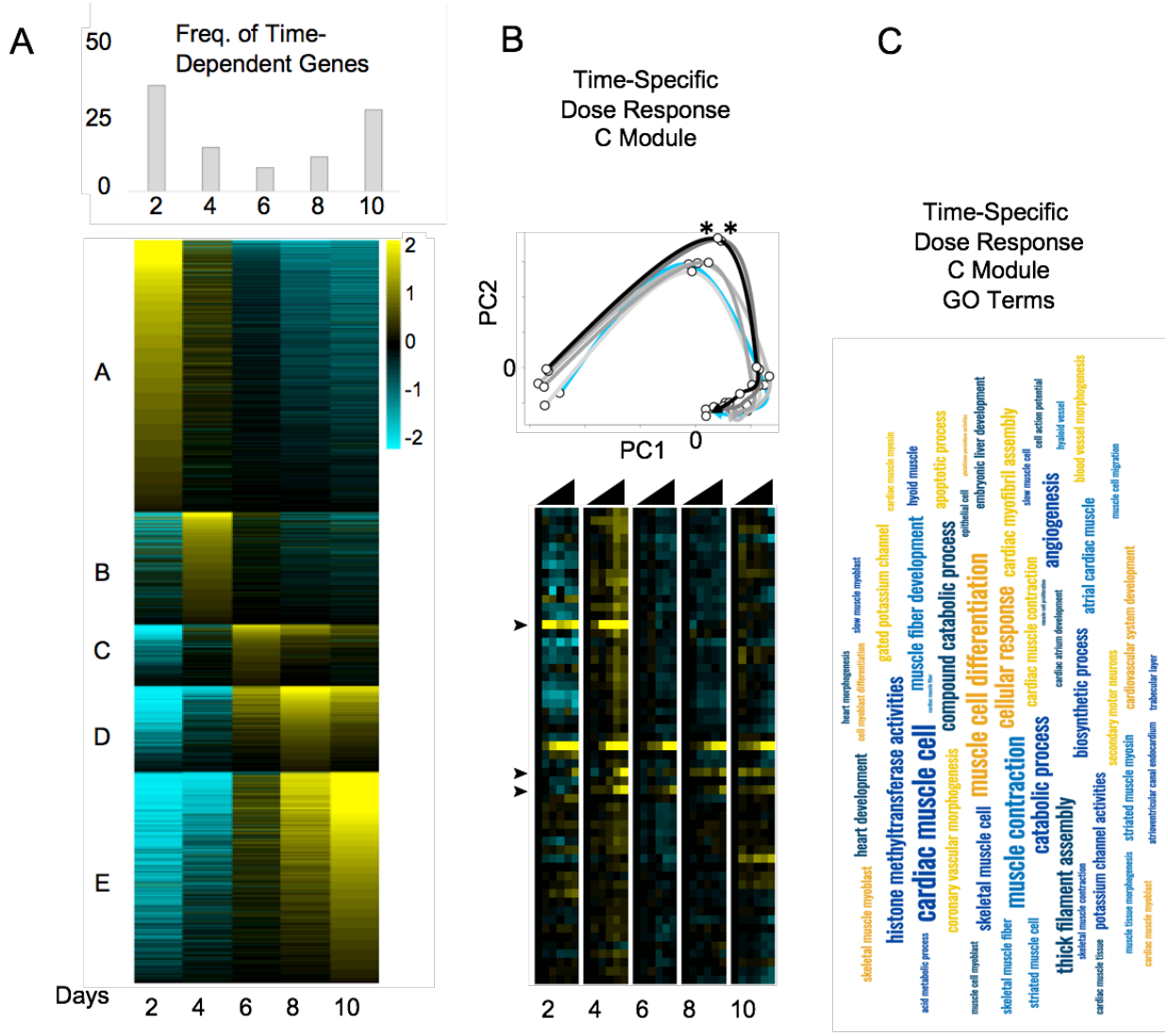

Figure 2.4: Development and oil response genes involved in Pacific herring embryogenesis. A) Top panel represents frequency of genes clustering by developmental time-point: 2-10 dpf. Heatmap visualizes changes in gene expression throughout development. The mean TMM was calculated from 6-8 biological replicates and was mean-centered normalized (yellow is upregulated, teal is downregulated). Rows were ordered by peak expression for 23,353 genes with a significant main effect of time on expression (FDR<0.05) obtained from five time-points (2,4,6,8,10 dpf). The time course is divided into five modules (A-E) corresponding to days of peak expression (e.g., module A is composed of genes showing peak expression at 2 dpf). B) Principal Component Analysis (PCA) of the subset of Module C genes (55) from Figure 2.3A heatmap with a significant oil dose response

by time interaction effect. Arrows indicate direction of developmental time. No oil controls is blue line, followed by darkening gray for increasing doses. Gene expression differed between medium-high and high dose at day 4 only ($p<0.05$; ANOVA). Heatmap visualizes the same Module C genes and expression levels across oil exposure from PCA, normalized to no oil controls and ordered by their baseline peak expression from Figure 2.3 A. Arrows indicates those genes with a significant dose-response at day 4 in the medium-high and high dose (from top: *pdap1, cyp1a,* c*yp1b1, si:ch73-334d15.2*). C) Word cloud of Gene functional pathway analysis from Module C genes to identify meaningful functions related to oil-impacted developmental processes.

2.4.5 Oil-induced changes in cardiac gene expression during embryonic and larval development

To further illustrate how oil exposure perturbs early development, we categorized all dose-response genes (2306 genes, FDR<0.1) by the developmental time-point at which it had the largest estimated dose response from the linear regression model (see methods). While some genes had similarly large dose responses across serial time-points (e.g., *cyp1a),* due to the dynamic temporal effect of embryogenesis, most genes have a singularly prominent dose response at a single time-point. KEGG pathway enrichment analysis was performed on these stage-specific dose-response genes (Figure 2.5A). We also manually curated three sets of genes for functional enrichment analysis: genes involved in cardiovascular system development and function, genes involved in innate immune system development and function, and genes involved in PAH metabolism (listed as Cardiogenesis*, Innate Immunogenic*, and Xenobiotic Metabolism*, in Figure 2.5A)

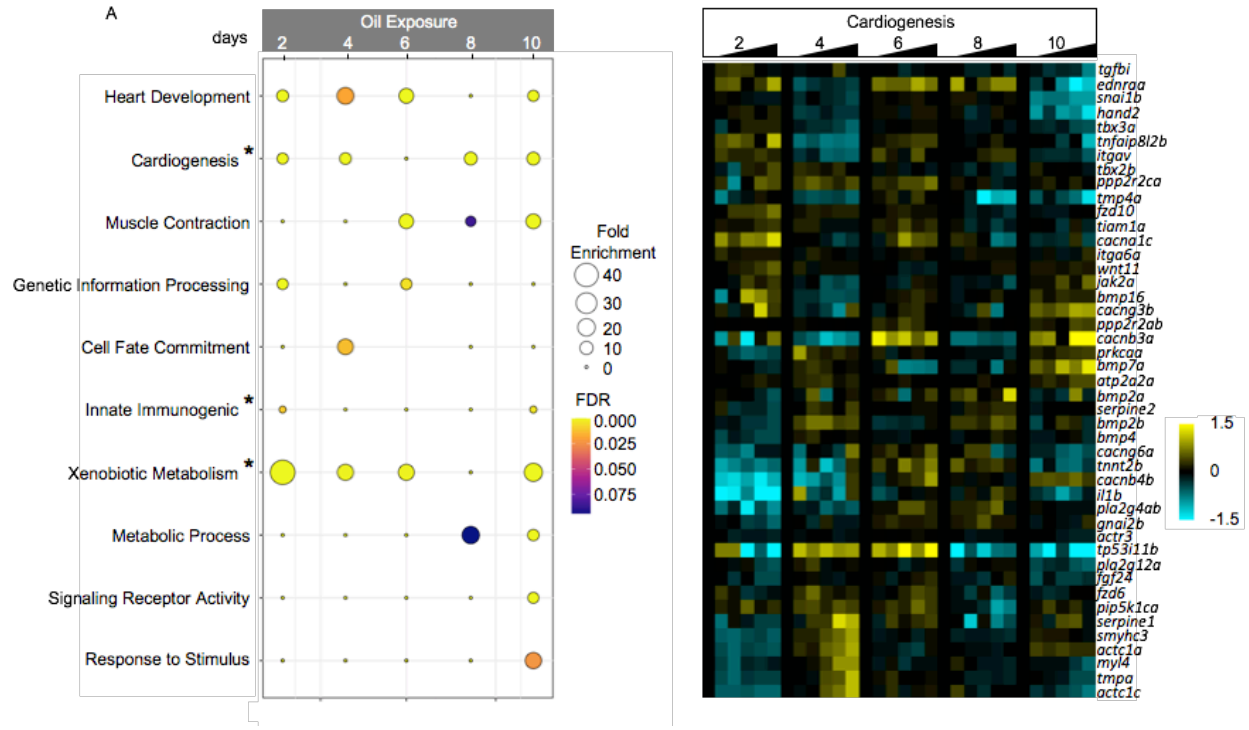

Figure 2.5: Enrichment of functional pathways for embryonic stage-specific oil dose response genes with a focus on cardiogenesis gene set. A) Dose response genes grouped into developmental time were tested for overrepresentation of KEGG pathways. Included are the seven functional domains, including three manually curated gene panels (Cardiogenesis*, Innate Immunogenic*, Xenobiotic*), that are most significantly enriched among all dose response genes (2,306). The x-axis presents the five developmental time-points during oil exposure, and the y-axis shows the name of KEGG pathways. The size of circles corresponds to the number of dose response genes with a given module (Fold Enrichment), and the different colors indicate significance levels (Enrichment FDR). B) Heatmap visualization of the significant dose response genes in the manually curated Cardiogenesis* gene panel. Expression represents fold change (yellow upregulated, teal downregulated) relative to no oil controls within each time-point. Gene rows are ordered by time of peak expression.

Notably, prior to organogenesis in the fish at day 2, there are a number of genes involved in early cardiogenesis with a significant dose response. Among genes with a significant and prominent oil dose-response at 2 dpf are those involved with cardiomyocyte proliferation, cardiac hypertrophy and fibrosis, and angiogenesis, including *itga6a, fbxo32, serpine2, ror2, tp53illb* and *fzd6* (Kramer et al. 2020; Zhou et al. 2020; Dogra et al. 2017). Within the manually curated list labeled Cardiogenesis, a subset of genes were significantly ($p<0.05$) differentially upregulated at 2 dpf (*cacna1c, wnt11, tbx3a,* and *fzd10*) (Figure 2.5B). Both *wnt11* and fzd10

form a complex involved in the Planar Cell Polarity Signaling pathway , which is crucial for cardiomyocyte migration and heart shape during ventricular outgrowth (Merks et al. 2018; Panáková, Werdich, and Macrae 2010; Galli et al. 2014). Of particular interest are expression patterns which reflect the negative concentration-dependent reduction in posterior ventricle outgrowth, beginning at the medium dose (64.5 ng/g). Cardiac-related genes significantly downregulated beginning at the same medium dose at 2 dpf include *actc1c, smad7, sox9, gnai2b, serpine2, pla2g4ab, bmp2b, bmp4,* and *il1b* (Figure 2.5B). *actc1c* is cardiac-specific and mutations in its human ortholog are involved in hypertrophic cardiomyopathy, linking reduced expression with the observed reduced ventricular outgrowth. While *pla2g4ab* and *gnai2b* encode phospholipase Cβ enzymes, which promote the release of intracellular calcium, playing a prominent role in cardiovascular signaling (Thai et al. 2010; Lyon and Tesmer 2013; Zuberi, Birnbaumer, and Tinker 2008). At the medium high and high dose, day 2 is notable for differentially downregulated genes involved in endocardial cushion morphogenesis, *bmp2b, bmp4, smad7, sox9,* and *tbx3a* (Asai et al. 2010; de Pater et al. 2012; Pogoda and Meyer 2002). Bmp signaling from the myocardium mediates endocardial proliferation. Under normal development, endocardial cells acquire chamber- and region-specific cellular shapes contributing to constriction of cardiac cushions at the atrioventricular canal, chamber ballooning, and eventually, two distinct cardiac chambers: a ventricle and atrium (de Pater et al. 2012). What is observed in the manually curated list is supported by the pathway in Figure 2.5A labeled Heart Development, representing, among others, the Heart Primordium Pathway (Zebrafish Anatomy Ontology, TAO:0000028). The heart primordium consists of a bilateral field of myocardial precursor cells that eventually form a heart cone (Tu and Chi 2012). Together,

these genes and pathways suggest that oil exposure prior to a functioning heart may still contribute to the embryonic cardiac injury phenotype and specifically reduced posterior ventricular outgrowth (Figure 2.3).

Day 4 represents the early stages of cardiomyocyte proliferation and heart formation in herring. We explored the different pathways and genes involved in heart development at the 4 dpf as well as the effects of oil exposure on gene expression and potential inflammation and injury response in cardiomyocytes. Indeed, genes in this still early time-point in cardiogenesis are enriched in two pathways, Cardiogenesis* (*ppp2r2ab, zgc:101810, zmp:0000001082*), and the Zebrafish Anatomy Ontology term Muscle (TAO:0005145), listed in Figure 2.5A under the heading Heart Development. Genes enriched in this pathway are the heart specific *nxk2.5*, *rbfox1l*, *myod1*, and *jam2a* (Lubbers and Mohler 2016; Veerkamp et al. 2013; Gunawan et al. 2019; Ye et al. 2015; Gao et al. 2016; Kobayashi et al. 2020; Jiao, Xu, and Du 2021). Linking the first two time-points in early development, *serpine2* was significantly differentially downregulated at the high oil dose at 2 dpf and differentially upregulated at 4 dpf. This gene, along with *serpine1*, plays a key role in fibrosis and the injury response pathway, particularly cardiac remodeling (Figure 2.5B & Figure 2.6) (Münch et al. 2017). Notably, *tnfaip8l2b* was a significantly down-regulated gene at 4 dpf across all doses (low-high). It is understudied in fish but is notable as both highly expressed in cardiomyocytes in mammals and as a regulator of inflammation and lipid biosynthesis (Münch et al. 2017; Li et al. 2018). *jak2a* was significantly repressed across all three doses at 4 dpf. Interestingly, JAK-STAT is a proinflammatory pathway that impairs the post-heart repair and reprogramming process, while *in vitro* pharmacological

inhibition of JAK2 enhances reprogramming efficiency (Hashimoto et al. 2019). Day four post fertilization embryos appear to be responding to inflammation and possible myocardial injury, as a result of oil exposure (Figure 2.6). This could result from PAHs binding to the ligand-activated transcription factor AHR, which could trigger oxidative stress via upregulation of the cytochrome p450 metabolizing enzymes and the release of superoxide, in turn evoking inflammation response molecules in cardiomyocytes. Similar to 2 dpf, endocardial cushion morphogenic genes were significantly differentially suppressed at 4 dpf starting at medium dose (*gata5, snai1b, tbx3a*, and *bmp16*). However, *actc1a, actc1c*, and *kcnma1a*, which are also involved in atrioventricular development (Gauvrit et al. 2022), were significantly upregulated. The expression of these genes is influenced by voltage-gated calcium channel genes, including *ryr1b* which was significantly upregulated at 4 dpf in the medium and high dose, and *camk1b* which disrupts RY1 protein receptor function (Gauvrit et al. 2022; Hirata et al. 2007; Berchtold et al. 2016) and is significantly downregulated at high dose.

Gene expression was measured in 6 dpf embryos at a point at which the heart in herring embryos has a regular beat. Genes associated with heart development unique to 6 dpf were identified in GO enrichment analysis under Cardiac Muscle Myosin Thick Filament Assembly (Muscle Contraction, Figure 2.5A). Gene expression perturbations that were sustained throughout the remainder of development were initiated at 6 dpf, including genes involved in Heart Development (GO enrichment of Angiogenesis Involved in Coronary Vascular Morphogenesis) and Muscle Contraction (GO enrichment of Voltage-gated Potassium Channel Activity). Two genes involved in regulation of voltage-gated calcium channel activity had

significant increased expression at medium dose relative to controls (*cacna1c, cacnb3a*). The dihydropyridine receptor, encoded by *cacna1c* and *cacnb3a*, is a calcium channel protein found in the membrane of cardiomyocytes. The dihydropyridine and ryanodine (encoded by *ryr1/2*) receptors are critical for excitation-contraction (E-C) coupling and cardiac muscle function (Tiitu and Vornanen 2003). Voltage-dependent activation of the dihydropyridine receptor proteins initiate calcium ion release from the sarcoplasmic reticulum via ryanodine receptor channels. Earlier studies with Pacific herring embryos showed that exposure to crude oil resulted in reduced contractility at 7 dpf with tissue TPAH about 3.5 times higher than the highest dose here (Incardona et al. 2009). There is a positive association with the number of dihydropyridine receptors and myocardial contractile strength, and while heart rates and receptors were not measured in these fish, the potential for increased contractile strength suggests a compensatory response to reduced heart rates (Rottbauer et al. 2001). In any event, the inhibitory effects of PAHs on calcium ion channels leads to abnormal excitation–contraction coupling as a result of altered calcium homeostasis.

Genes regulating contractility (KEGG pathway Adrenergic signaling in cardiomyocytes) were significantly enriched at 8dpf (Muscle Contraction, Figure 2.5A). Adrenergic signaling in cardiomyocytes KEGG pathway was continually significantly enriched into early larval-stage (10 dpf - 1 dph) as well (Figure 2.5A). Moreover, inspection of individual dose response genes associated with this pathway show genes involved in phosphatase regulatory and calcium channel activity (*ppp2r2bb, ppp2r2ca, cacna1c, atp1a3a, cacnb4b*). Chronic stimulation of these genes induces cardiomyocyte hypertrophy and apoptosis in humans (Rouillard et al. 2016).

Genes down-regulated starting at the medium dose in 8 dpf embryos included *cacng6a* which is involved in calcium flux, and *pip5k1ca* which is a lipid kinase; both of these affect cardiac contractility (Ameen et al. 2022; Shankar et al. 2021). *adamts14* was also significantly downregulated at medium-high and high doses. Though function is not specifically known in fish, paralogs in mammals (e.g., *Adamts9*) affect cardiac and aortic development. Our data suggest that these genes play a critical role in the complex process of herring embryonic heart development, and their dysregulation by oil exposures can lead to congenital heart defects.

On the final day of embryonic oil exposure we identified perturbation of genes associated with the Apelin Signaling pathway and their impact on the oil-induced phenotype, as well as genes and pathways associated with contractility and ventricular outgrowth. The Apelin Signaling pathway (Muscle Contraction, Figure 2.5A) is functionally relevant for cardiac hypertrophy and ventricular ballooning seen in early larval fish (Qi et al. 2022), and this pathway is enriched among oil responsive genes at 10 dpf, 2 days prior to the dose-dependent reduction in the degree of ventricular ballooning (observed at 1 dph) (Figure 2.3). A troponin gene (*tnnc2b*) from this pathway was significantly differentially downregulated at medium-high dose. This gene acts upstream of muscle conduction (Tsedeke et al. 2021). A p53 gene (*tp53i11b*) from this pathway was differentially downregulated medium-high and high dose at 10 dpf. Reduced downregulation of this ortholog in humans is associated with aortic valve dysfunction (Padang et al. 2015). Additional down-regulated genes at 10 dpf in at least one of the medium, medium-high and high doses include *hand2 and gata5*, both transcription factors that promote the expansion of cardiac ventricles, the latter of which functions downstream of *bmp2*

(Schoenebeck and Yelon 2007), also significantly repressed in 10 dpf embryos. *hand2* and *bmp2b* zebrafish mutants have reduced cardiomyocytes and in mammals produce ventricular hypoplasia (McFadden et al. 2005). Finally, *fgf24* is significantly differentially suppressed at the high dose. This fibroblast growth factor gene is believed to be functionally similar to *fgf8*, a gene in mutant zebrafish that leads to a severe reduction in cardiomyocyte production, particularly in ventricular tissue (Padang et al. 2015; Reifers et al. 2000). Throughout embryogenesis, genes perturbed by oil exposure enriched cardiac pathways associated with contraction and conduction. Notably, there was a trend toward a greater number of contractile genes becoming dysregulated as oil-exposed embryos advanced through development. This supports the idea that these genes are implicated in heart shape formation and a concentration-dependent reduction in ventricular outgrowth.

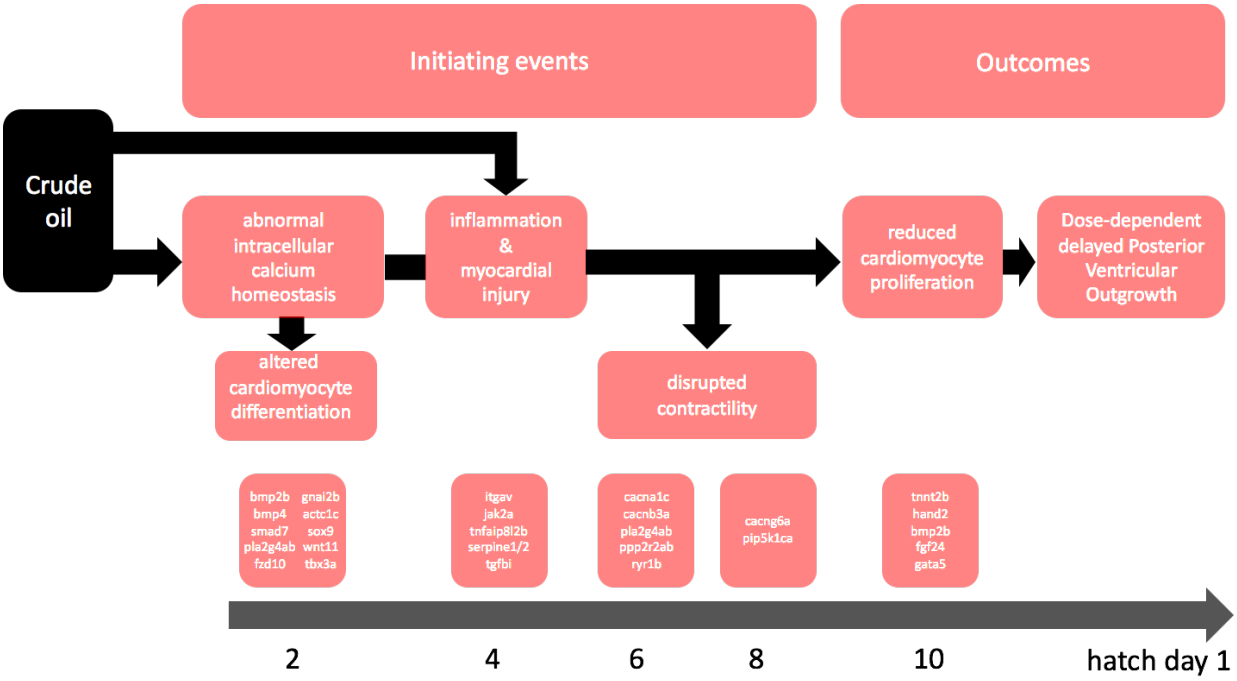

Figure 2.6: Model summarizing the pathway for low-concentration crude oil toxicity during herring early development. Molecular and cellular initiating events linked (black arrow) to cardiovascular

system developmental perturbations following exposure to trace concentrations of crude oil during embryogenesis, inferred from transcriptomic responses to oil exposure at multiple timepoints during development (timeline in days post-fertilization indicated by gray arrow).

We explored the persistent effects of crude oil exposure on larval herring heart development, identifying the molecular mechanisms linked to calcium signaling and contraction important for proper heart function and growth, and linking embryonic morphological defects to persistent molecular perturbations in the maturing heart. After hatching in clean seawater, and the metabolism and excretion of tissue PAHs, the ventricles of larval herring hearts exhibited a concentration-dependent reduction in growth (Figure 2.3). Cardiomyocyte proliferation is dependent on intracellular calcium signaling and contraction, which was likely inhibited during early cardiogenesis in the embryo by the bioconcentration of contaminating tissue PAHs. Two oil exposure levels of herring were grown out in clean seawater, no oil control treatment group and the medium high (105.0 ng/g) treatment group. To characterize the molecular mechanisms underpinning delayed ventricular ballooning, gene expression was measured at the time of the disrupted heart morphology assessment (Figure 2.3, day 1 post hatch). To further assess the persistent effects of crude oil exposure during embryogenesis, periodical sampling of larval transcriptomes was continued to the end of larvagenesis: 50, 64, and 78 days post-hatch (dph). Gene expression analysis for all larval fish was performed using weighted gene co-expression network analysis (WGCNA). Co-expressed gene modules that were significantly affected by oil exposure were functionally characterized by KEGG enrichment analysis (Figure 2.7). The Adrenergic Signaling in Cardiomyocytes KEGG pathway was first enriched among genes perturbed by oil exposure at 8 dpf, and this pathway continued to be significantly enriched among oil-perturbed gene modules in post hatch larval fish, underscoring the importance of

calcium signaling and contraction in PAH-induced cardiotoxicity (Circulatory System, Figure 2.7). Between 64 and 82 days after the initiation of oil exposure (50-68 dph), larva are approaching metamorphosis. Previous reports indicated that similarly late-stage larval herring exposed as embryos to 140 ng/g dose of tissue TPAHs exhibited abnormal trabeculation – the formation of spongy myocardium and a precursor to cardiac hypertrophy (overgrowth) (Incardona et al. 2021). At 50 dph, but not 64 dph, an oil-perturbed gene co-expression module was enriched for Vascular Smooth Muscle Contraction and ECM-receptor interaction KEGG pathways (Circulatory System, Figure 2.7). Similar to cardiomyocytes, vascular smooth muscle contraction is initiated with an influx of calcium. In addition, formation of fibrotic tissue, a sign of injury to the heart, can result in remodeling of the heart's extracellular matrix (Beffagna 2019). Larval herring approach metamorphosis into juveniles around 78 dph, the final time-point in this study. At this late stage, long after oil exposure, we find co-expression modules still perturbed from exposures that are enriched the heart-related pathway Cardiac Muscle Contraction (Circulatory System, Figure 2.7). Notably this pathway contains genes which regulate cardiomyocyte action potential, excitation-contraction coupling, and contractility in the heart. Previous studies, using much higher doses of crude oil (620 ng/g TPAHs), resulted in bradycardia (slow heart rate) in herring embryos, and reduced swim performance in juveniles (Incardona et al., 2009). The time series transcriptome data presented here link embryonic morphological defects from this and previous research to persistent molecular signals in the maturing heart, and at a fraction of the dose. Remarkably, the impact of a fleeting exposure during embryogenesis on cardiac function is supported by perturbed gene expression that persists until at least late-larval stages. This finding is reinforced by previous studies linking embryonic and larval exposure to reduced

survival and fitness in adults across fish species, including zebrafish, pink salmon and mahi larvae (Mager et al. 2014; Heintz 2007; Hicken et al. 2011).

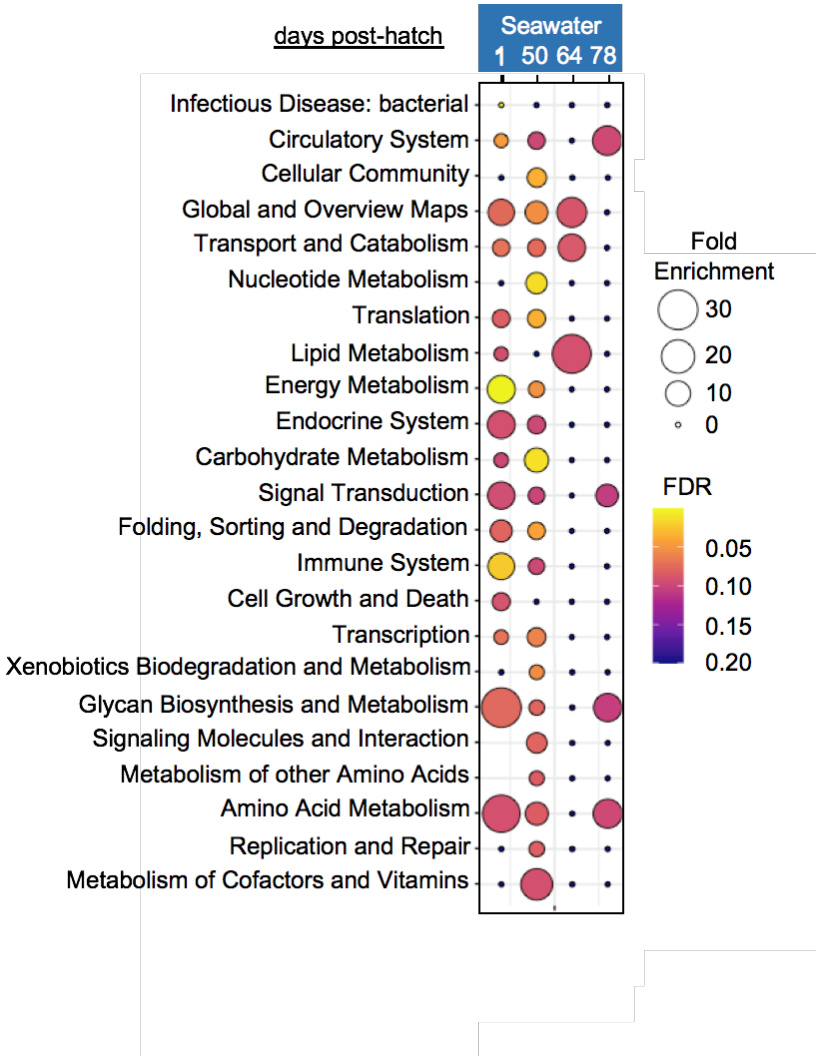

Figure 2.7: Biological function of oil-dose response genes post oil exposure. KEGG pathway analysis was performed on gene modules identified through WGCNA (p value <0.1) to identify functions related to long-term perturbation of developmental processes. The x-axis spans herring larvagenesis from post-hatch day 1 to early juvenile stage h78, and the y-axis shows the name of the top 24 enriched KEGG pathways. The size of circles corresponds to the number of dose response genes with a given module (Fold Enrichment), and the different colors indicate significance levels (Enrichment FDR).

2.4.6 The Arc of Innate Immune Development and Oil-induced changes in gene expression

We used a manually curated set of genes to characterize the temporal dynamics of immune system development and function through embryogenesis and larval stages. Within this context we then tested how oil exposure may perturb early life immune function. To this end we establish a robust framework to study innate immune system development by focusing on a manually curated set of 1,744 immunogenic genes and characterizing their expression from embryogenesis to hatch and through larvagenesis up until early stage juveniles (from 2 dpf - 90 dpf). After characterizing the transcriptional arc of innate immune system development in Pacific herring, we pinpoint the modulatory impacts of PAHs on these pathways. Transcriptional analysis of the initiation and development of the innate immune system was determined by ordering genes by their peak expression across nine time-points (Figure 2.8). Samples were taken at four time-points after hatch to provide information about transcriptional changes up until when the immune system is expected to have matured (larval through to metamorphosis stages). The heatmap reveals two major developmental clusters representing immune system genes that peak in expression during embryogenesis (2-10 dpf) and those that peak after hatch during larval stages (1-78 dph) (Figure 2.8). Principal Component Analysis (PCA) of immune system gene expression also clearly distinguished embryo and larval stages (Figure 2.8, upper panel). The differences between early and late immune-related genes can be explained by the onset of hatch, as a number of critical transitions occur during embryo and larva development. In teleost fishes the first line of defense against invading pathogens is the physical barrier of the chorion. A week after shedding the protective chorion herring begin feeding, further intensifying interactions between the developing immune system and the environment. Among immune system genes with peak expression in early embryogenesis, overrepresentation

analysis of KEGG pathways showed the strongest enrichment at 2 and 4 dpf for the *Wnt, FoxO, MAPK, and MTOR* signaling pathways, which are non-specific multi-functional pathways involved in normal development and organismal growth. In contrast, among immune system genes with peak expression post-hatch, significantly enriched KEGG pathways include *Bacterial Infection, MAPK Signaling, NOD-like Receptor Signaling*, and the *C-type Lectin Receptor Signaling*, pointing to the emergence of the innate immune system after hatch. Together, these patterns suggest that much of the maturation of innate immune function occurs at hatch and continues into late-larval/early juvenile stage fish.

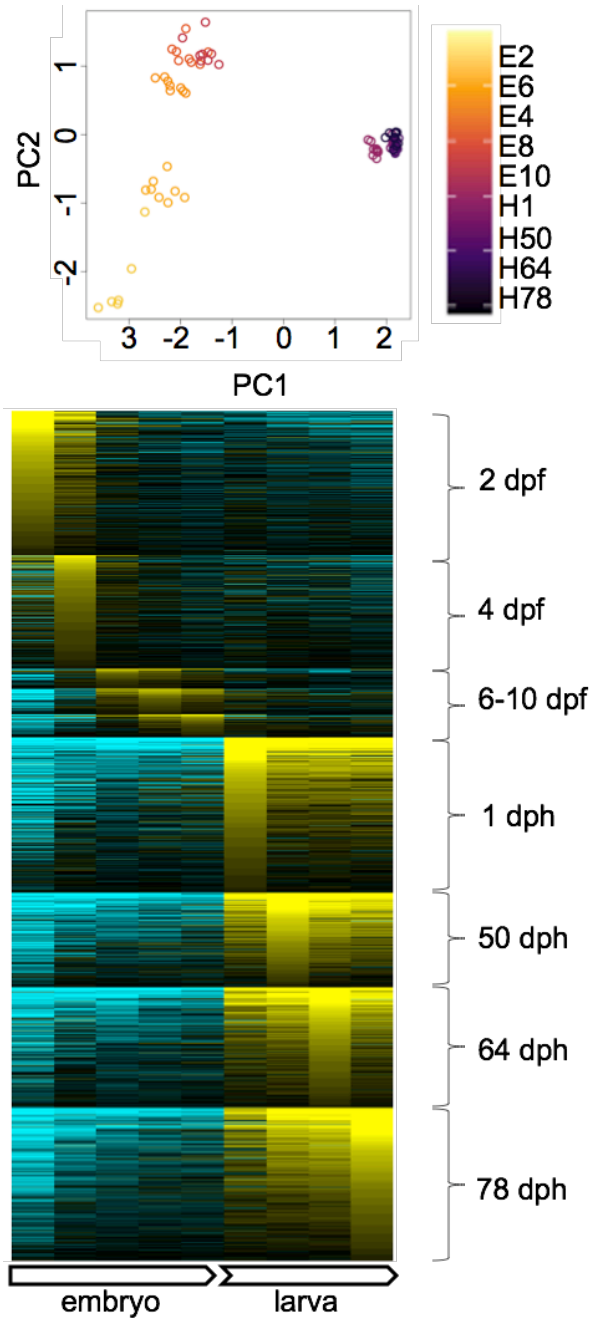

Figure 2.8: Innate Immune system temporal expression during herring development. Top panel PCA represents the variance of gene expression for the manually curated panel of innate immune genes (1,744). Colors represent developmental time-points for individual no-oil control samples. PC1 captures variance between embryos and larval fish. Bottom panel heatmap visualizes individual genes in the innate immune gene panel throughout development. The mean TMM was calculated from 6-8 biological replicates for no-oil controls and was mean-centered normalized. Rows were ordered by peak expression across nine developmental time-points.

2.4.7 Oil-induced changes in innate immune gene expression during embryonic and larval development

Patterns of transcriptional response to PAHs point to an initial and late embryonic stage immune response to contaminating oil. The manually curated gene set, labeled *Innate Immunogenic ** in figure 2.5A, was significantly enriched among genes perturbed by oil exposure at 2 dpf and 10 dpf, bracketing early and late embryogenesis. During development post-hatch, co-expressed gene modules that were perturbed by oil exposure at 1 dph were enriched for Pathogen (*Infectious Disease: Bacterial*) and *Immune System* KEGG pathways, with the Immune System pathway also significantly enriched among oil-perturbed genes at 50 dph (late-larval stage). Genes enriched in the *Innate Immunogenic* pathway at 2 dpf were associated with Wnt and BMP signaling, and protein biogenesis, which are all implicated in growth and development. Interestingly, the majority of the genes enriched in the *Innate Immunogenic* pathway at 10 dpf were also associated with the cardiovascular system–particularly the Apelin Signaling pathway, suggesting a mechanistic link between oil-induced perturbation of cardiac and immune development.

Much of the innate immune system oil-induced perturbation occurs at, which is coincidentally after oil exposure (Figure 2.8). At 1 dph genes are enriched in two parent KEGG pathways, Pathogen (*Infectious Disease: Bacterial*) and Immune System (*Toll-, RIG-, and NOD-like Receptor Signaling, and Cytosolic DNA-sensing* pathways). Immune system development is a dynamic process, and as early as 4 dpf, NOD-like receptor genes, *map1lc3c* reach peak expression in normally developing fish. However, as pathway enrichment analysis suggests, *Toll-Rig*, and

NOD-like receptor signaling pathways, hallmarks of innate immune function, are not fully developed until after hatch. While persistent molecular perturbations in cardiac genes lasted until at least 78 dph, the molecular immunological response to PAH exposure did not last after 1 dph. We can speculate that this was largely due to a mismatch between the embryo oil exposure period (embryo) and the early larva critical window of development of innate immune system development. The immune system is a dynamic and highly integrated organ with the ability to discriminate 'self' and 'non-self', attack pathogens, and engage in organ development and tissue homeostasis and repair. It is becoming increasingly apparent that there is significant cross-talk between the innate immune and cardiovascular systems. For example, an auto-immunological component to cardiac conduction drives some rhythm pathologies in humans (Lazzerini et al. 2018). Whether this cross-talk is common across vertebrates or if there is a link between cardiotoxicity and immunotoxicity in crude oil exposure remains largely unknown.

2.4.8 Oil-induced changes in xenobiotic metabolism gene expression during embryonic and larval development

We examined the induction of cytochrome P4501A and other xenobiotic response genes during the embryonic development of Pacific herring in response to oil exposure, finding a dose-dependent increase in gene expression at 2-4 dpf and a varied expression of Phase 1 and 2 genes across developmental time-points. Expression of the *cyp1a* enzyme is induced as part of a broader activation of xenobiotic metabolism pathways initiated by the binding of PAHs to the AHR. Induction of *cyp1a* in response to oil first detected as early as 2 dpf, with expression peaking at 4 dpf. This induction correlated with increasing PAH tissue concentrations (Figure

2.2). To further characterize the defensive xenobiotic response in developing herring, we evaluated the transcriptional changes of 34 xenobiotic response genes, including *cyp1a*, with a significant dose response during embryogenesis. The same methods were used to categorize this gene set into developmental time-points as was done for the Cardiogenesis and Immunogenic curated pathways. (Figure 2.5A, Xenobiotic Metabolism*). Overall, differentially expressed genes were primarily Phase1 metabolism genes (e.g., cytochrome P450s, aldehyde dehydrogenases) or those that control Phase 1 xenobiotic metabolism (e.g., *ahrr* and *aip*). The Xenobiotic Metabolism pathway was significantly enriched among genes perturbed by oil exposure at 2, 4, 6, and 10 dpf. Xenobiotic gene expression varied between developmental time points (Figure 2.9), with more genes eliciting a linear dose response to increasing concentrations of TPAHs in the middle (4 dpf) of embryonic development and at the same time that PAH uptake was rapidly increasing (Figure 2.3). By late embryogenesis (8, 10 dpf), xenobiotic metabolism gene induction was less pronounced, presumably as the embryo's capacity to metabolize and excrete xenobiotics increases, leading to lower TPAH concentrations by 10 dpf (Figure 2.9). The largest set of oil-responsive genes enriched for Xenobiotic Metabolism* was at 2 dpf, which is very early in development and only 1 day after initiation of oil exposure (Figure 2.5A). Xenobiotic Metabolism* genes that were oil-responsive at 2 dpf were primarily composed of Phase 1 and two Phase 2 genes sulfation genes, *sult6b1, sult1st6*. Seven of the nine genes that were oil-responsive by 4 dpf are members of the cytochrome P450 superfamily (*cyp26c1, cyp1a, cyp2y3, cyp2r1, cyp1a1, cyp2g1, cyp1c1*) with the two exceptions being *aldh1a3*, an aldehyde dehydrogenase, and *aip*, a regulator of xenobiotic metabolism. Xenobiotic Metabolism* genes that were oil-responsive at 6 dpf were primarily phase 2 genes:

three aldehyde dehydrogenases (*aldh7a1, akr1a1b, aldh1a3*), two methyltransferase genes (*comtd1, hnmt*) and a sulfation gene (*sqor*). In later embryogenesis and the final day of oil exposure, oil-responsive genes at 10 dpf included both phase 1 and 2 xenobiotic metabolism genes, and two regulators of xenobiotic metabolism, *aip* and *ahrra*. A fuller representation of genes in the xenobiotic pathway by 10 dpf suggests that metabolism has matured relative to earlier time-points. By hatch, Pacific herring have the ability to induce phase 1 and 2 enzymatic reactions and metabolize xenobiotics like PAHs. This results in more water-soluble metabolites for excretion.

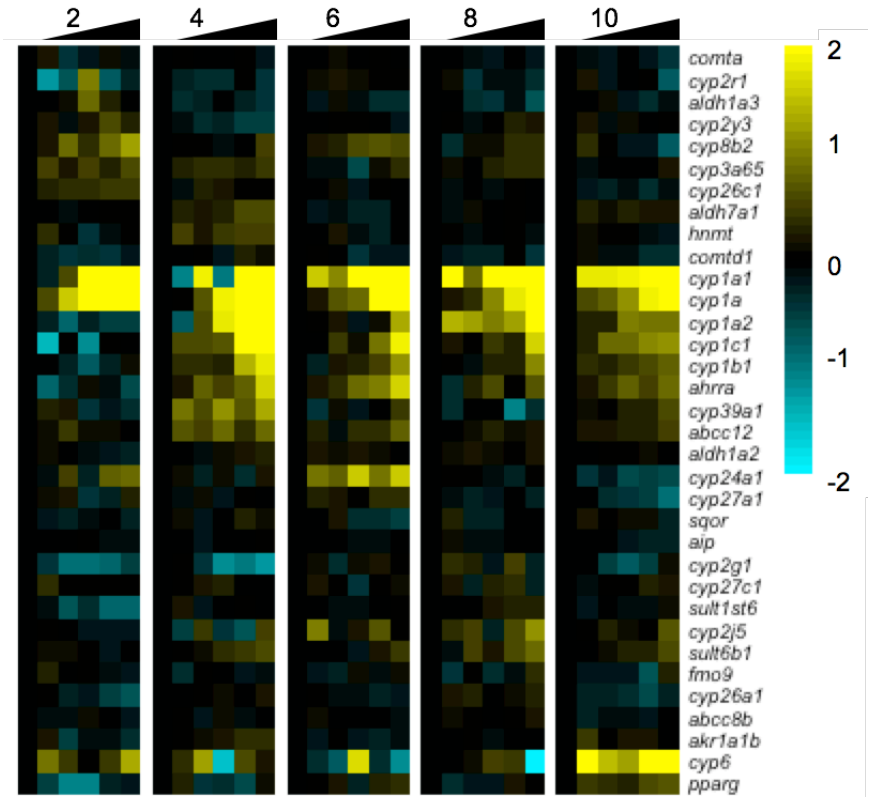

Figure 2.9: Temporal dynamics of fold change in mRNA expression for xenobiotic response genes. A) Heatmap of the xenobiotic response gene panel with a significant oil dose by developmental time interaction at five developmental time points (2, 4, 6, 8, and 10 dpf). Gene expression is normalized to within day controls. Within each time point concentrations of TPAHs increase from left to right.

2.4.9 Conclusion

It is apparent that sublethal concentrations of PAHs cause subtle changes in heart morphology with potential downstream physiological consequences for cardiorespiratory performance. The aim of this study was to examine the short-term acute and long-term delayed toxicity of embryos exposed to even lower trace amounts of crude oil than have previously been published. To accomplish this, we focused on the developing ventricle as an exquisitely sensitive concentration-dependent phenotype. We systematically pinpointed genes and pathways involved in cardiogenesis and ventricle function from transcriptomic data, including a detailed time-series. We expanded the scope of oil-induced developmental toxicity by characterizing the modulatory impacts of PAHs on immune system functional pathways and identified a potential critical and largely unexplored window of innate immune development in Pacific herring. Finally, we detailed the full xenobiotic response to contaminating PAHs during herring embryogenesis. Our analysis and findings confirm that crude oil-induced cardiotoxicity is the canonical embryonic injury phenotype in Pacific herring. Perturbations in key cardiac-related genes responding to trace oil levels and time-points preceding organogenesis indicate early embryonic oil exposure sets the stage for larval cardiotoxicity. Future oil exposure research should aim to bridge the understudied developmental gap at hatch between late embryogenesis and early larvagenesis.

Chapter 3

Population contrasts in global gene expression during oil exposure and subsequent virus challenge

3.1 Abstract

The Pacific herring (*Clupea pallasi*) fishery in Prince William Sound (PWS) collapsed three years after the 1989 Exxon Valdez Oil Spill and has failed to rebound. While disease epizootics were likely major contributing factors to the PWS collapse, and possibly to the lack of recovery, nearby herring populations unaffected by the spill have remained relatively healthy despite the presence of disease vectors. This suggests that the oil spill was a contributing factor to the PWS collapse, perhaps through interactive effects between oil and virus. Herring larvae are exquisitely sensitive to contaminating crude oil, but linking the oil spill to a population-level collapse and lack of recovery remains controversial. Recent research into modeling multi-stressor impacts have enriched our understanding of the complex interactions between ecological factors that may govern herring population dynamics. A multiple environmental stressor scenario where oil acted as an interactive contributor to the collapse and slow recovery is plausible, where oil exposure could link to delayed mortality and reduced recruitment at the population level. We posit that oil may have acted as a strong genotypic selective agent with tradeoffs on immune system function. Alternatively, sublethal embryonic exposure to oil at the time of the spill may have impaired immune system development and amplified later-in-life

susceptibility to disease in PWS herring. To test these hypotheses, we evaluated how extremely low levels of crude oil affects the heart and immune system in three populations of Pacific herring, including fish from PWS (descendants of the Exxon Valdez oil spill), a nearby reference population in Sitka Sound (AK) without an oil spill history, and a more geographically distant reference population (from Puget Sound, WA). Fish were serially exposed to crude oil as embryos, raised in clean seawater until metamorphosis, then challenged with viral hemorrhagic septicemia virus (VHSV). We measured larval cardiac morphology and analyzed global gene expression in the embryos during oil exposure, and examined gene expression throughout the subsequent virus challenge. Oil at very low doses caused impairment of cardiovascular system development, where the Alaska populations were more sensitive than the WA population. Transcriptomics provided further insight into population variation in gene expression, and population variation in their transcriptional response to oil. Notably, a significant set of oil-responsive genes unique to the geographically distant Washington population were enriched for immunoglobulin A as well as cardiac contraction pathways. Brief embryo exposure to oil did not affect the lethal effects of acute VHSV exposure in later life, but populations varied in their molecular immune response to virus. Although populations varied in their response to oil, our results do not support the hypothesis of a unique evolved response to oil or virus exposure in PWS fish.

3.2 Introduction

Oil spills have major ecological consequences and lead to marine ecosystem disruption. The Exxon Valdez oil spill (EVOS) was the largest single oil spill of its time in US coastal waters, and

occurred about 2 weeks before the peak of herring spawning in Prince William Sound, AK (AK-P) (Pearson et al. 1999). This unprecedented disaster, coupled with the collapse of the fishery three years later, fueled research into the long-term and sublethal effects of oil on herring populations. Prior to the Exxon Valdez, it was thought that herring chorions acted as a physical barrier to the developing embryo, limiting impacts on the population from crude oil to acute mortalities resulting from direct contact. We now know that developing herring are vulnerable and highly sensitive to crude oil, and that PAHs act as developmental toxins (Cherr, Fairbairn, and Whitehead 2017). Furthermore, we have learned that while exposure to oil in early development may not be immediately fatal to herring, even vanishingly small amounts impedes growth and development and may have lifelong fitness consequences (Incardona et al. 2015). Finally, advances in modeling multi-stressors and both direct and indirect biotic interactions have complicated our understanding of the impacts of interactive effects on herring populations (Muradian et al. 2017).

The causes of the AK-P herring fishery collapse remain controversial 30 years later (Pearson et al. 2012). Unexpectedly, the fishery declined 75% from 1993 to 1994, strongly suggesting external factors. In recent years, a disease epizootic has been advanced as a direct contributor to the collapse (Marty, Quinn, et al. 2003; Marty et al. 1998a). Additional drivers potentially acting in concert include poor nutrition, and over-harvesting of the fishery. At the event of the collapse, crude oil was no longer detectable in AK-P waters. However, the timing of the oil spill is conspicuous, and its role in a slow recovery of the AK-P fishery remains unexplored. Forage fish undergo natural fluctuations in population size over time, but the prolonged depressed

state is a unique feature of the AK-P herring fishery, including no large recruitment in the past three decades. Predator pressure, resource competition, trace exposure to lingering oil and continued disease cycles, similar to the epizootic that initiated the collapse, are plausible explanations for why the AK-P herring population has failed to recover, but the root cause is unknown.

Pollutants can act as powerful selective agents, inducing rapid adaptation that in turn can incur fitness trade-offs (Whitehead et al. 2017). In particular, previous work by has shown how exposure to hydrocarbons, similar to those seen in crude oil, acted as a powerful selective force for survivorship during early development in a nearshore forage fish (Reid et al. 2016). Favored were genetic variants with impaired function of the aryl hydrocarbon receptor (AHR) signaling pathway. Normal AHR signaling interacts with immune system signaling, such that impairments in AHR function could come at the cost of normal immune system function. Indeed, the authors found evidence for subsequent selection on immune system genes in this population. We posit that contaminating oil in 1989, and perhaps lingering oil in subsequent years, acted as a strong selective agent on developing herring, with trade-offs related to impaired immune function, the consequences of which may have contributed to their subsequent decline and/or lack of recovery. We test two hypotheses 1) Oil exposure caused genotype-selective mortality, where surviving adults and their offspring had compromised immune function (adaptive selection hypothesis) thereby increasing vulnerability to epizootic disease. 2) Oil exposure during early life impaired fish immune system development such that adults were more susceptible to disease (physiological/developmental perturbation hypothesis).

We compared three populations so that we could distinguish the influence of neutral genetic drift from non-neutral processes (e.g., natural selection) that may have contributed to population differences, particularly for Prince William Sound fish. In addition to the population with an oil spill history, our focal population (AK-P), we chose two reference populations, one geographically nearby population (AK-S) and one geographically distant (WA-P), to assess variation due to neutral genetic drift (Figure 3.1). Both reference populations were without a history of oil spill, epizootic disease, or population collapse, allowing us to distinguish non-neutral drivers unique to the AK-P population. In a series of laboratory-based experiments, we exposed herring embryos from all three populations to the same source crude oil from the Exxon Valdez during embryogenesis, transferred the survivors to clean water pre-hatch and reared them through metamorphosis to juveniles, and exposed juveniles to a viral pathogen that is endemic to the Pacific northwest (viral hemorrhagic septicemia virus; VHSV). Cardiotoxic effects of crude oil during embryo exposure occur for every fish species tested and are particularly well-studied in Pacific herring (Incardona et al. 2013). We quantified larval cardiac morphology and function (heart chamber measurements) following embryonic oil exposure, and compared sensitivity between populations. We then tested responses of juvenile herring to VHSV challenge and compared those between populations and between fish that had been exposed to oil or not during embryogenesis. We tracked genome-wide gene expression through development, and in response to embryonic oil exposure, for each population to test for conserved and evolutionarily diverged molecular responses to oil, cardiotoxic responses, and responses that would indicate perturbation of immune system development.

Outcomes from laboratory-based experiments will differentiate the following hypotheses and their alternatives and are as follows. For adaptive hypotheses, our null hypothesis is that population varaition in responses to oil are governed by random-neutral differentiation, in which case we would expect the WA population would be most distinct from the two AK populations. (H1.0: null expectation). Our alternative hypothesis is that natural selection following the EVOS oil spill and/or the VHSV epizootic caused population differentiation in sensitivityi to oil and/or virus, in which case we would expect the AK-P population to be most distinct in their responses compared to AK-S and WA fish. This scenario ties the oil spill to the collapse and continued susceptibility of AK-P herring to disease epizootics (H1.1 adaptive selection). Our second hypothesis is that oil exposure during development has persistent impacts on disease susceptibility in later life, where our null expectation is that embryonic exposure to oil has no effect on sensitivity to later-in-life VHSV challenge (H2.0 null expectation). If we observe a similar oil-induced response to virus challenge across all three populations relative to no oil controls, this would indicate that oil exposures during development modulate herring immune system maturation and function; this could have contributed to the disease-associated collapse of the AK-P fishery (H2.1: physiological impact).

3.3 Methods

3.3.1 Animal Care

Pacific herring (*Clupea pallasii*) were collected near Krestof Island in Sitka Sound, Alaska (AK-S), Cedar Bay in Prince William Sound, Alaska (AK-P), and Cherry Point in Puget Sound, Washington

(WA-P) (Figure 3.1, top right panel). Collection of ripe herring by gill net occurred on April 9th for AK-S, April 12th for AK-P, and May 9th for WA-P in 2018. Ovaries and testes were dissected and stored in humidified whirl packs or petri dishes prior to transport to the Marrowstone Field Station (USGS Western Fisheries Research Center, Nordland, WA) for the Alaska sites and within an hour after transport for CP (# Females used: SS-16, PWS-11, CP-10).

Fertilizations were performed in water tables at the Marrowstone Field Station as described elsewhere (Griffin et al., 1998). Eggs were evenly distributed onto 24 sheets of 1 mm nylon mesh, each of which could hold 10-20 grams of embryos. At least 5 males were used for fertilization and gametes were evenly distributed amongst the egg-covered sheets. Fertilized eggs were left undisturbed for 1 hour and then transferred to 1.2 m diameter tanks with filtered seawater overnight for 8-12 hours. Fertilization rates were then assessed microscopically and oil exposure occurred on batches with rates above 75%. For the experiments reported here, fertilization rates were 85%, 85%, and 96% for AK-S, AK-P, and WA-P fish, respectively. Fertilizations were performed the day after gonad dissection for each population, and crude oil exposure was initiated 1 day post fertilization (dpf) and continued until 10 dpf. Exposures were in filtered seawater at ambient temperatures which were 9.2 ± 0.2, 9.3 ± 0.2, and 10.8 ± 0.3°C during the AK-S, AK-P, and WA-P exposures, respectively. Temperature was recorded twice an hour in a downstream tank from the exposure system by a Hobo Water Temp Pro v2 sensor (Onset). Animal care and use was evaluated and approved by the Western Fisheries Research Center IACUC (protocol# 2008-51).

3.3.2 Oil Exposure

Prior to exposures, Alaskan North Slope crude oil (ANSCO) was weathered by heating to 60°C in a water bath until the volume decreased by 10%. The exposure system (SINTEF, Trondheim, Norway) produced a continuous flow-through of dispersed oil droplets as described in Nordtug et al., (2011), which allowed for precisely controlled and highly comparable exposures for each population. The system consisted of 24 10-L tanks with individual flow-through of filtered seawater (360 µL/min flow rate) at ambient temperature and ANSCO effluent (60 µL/hr injection rate). ANSCO effluent entered the system through a dispersion generator that produced microdroplets of oil which were pumped into a 2 L reservoir that was gravity-fed to a series of solenoid valves which controlled the ratios of effluent to filtered seawater into the exposure tanks. Six concentrations were generated ranging from 0.01 (no oil) to 3.5 µg/L ∑PAHs. Exposure experiments were conducted with four replicate exposure tanks per oil concentration. Fertilized eggs adhered to Nitrex sheets were distributed amongst the tanks (1 per tank) and monitored several times daily until 10 dpf at which time oil injection stopped (Figure 3.1). Exposure dates for each population were 4/10/18-4/20/18 for AK-S, 4/13/18-4/23/18 for AK-P, and 5/11/18-5/21/18 for WA-P. During the overlapping exposure days for AK-S and AK-P (4/13/18 – 4/20/18), there were two Nitrex sheets with adhered eggs per tank.

3.3.3 Analysis of PAHs

Water samples were collected mid-water column by siphoning 100-mL into 125-mL amber bottles with 20 mL of dichloromethane for stabilization (HPLC-grade, Burdick & Jackson, Muskegon, MI). Samples were analyzed for 58 PAH analytes as described elsewhere (Sloan et

al., 2014). Briefly, samples were extracted with methylene chloride using an accelerated solvent extractor for tissue or separatory funnels for water samples. Highly polar compounds and lipids were removed from PAH extracts with silica/aluminum columns and size-exclusion high-performance liquid chromatography, respectively. PAH separations were conducted on a 60 m DB-5 chromatography (GC) capillary column before detection on an electron impact mass spectrometer (MS) in selected-ion monitoring mode. Concentrations below the detection limit were reported as <LOQ (less than the limit of quantification). TPAH values were calculated from detected values only. For body burdens of TPAHs, four replicate pools of 100 embryos each were collected from each oil exposure treatment at 10 dpf and stored in 2 mL cryotubes at -80°C. Samples were analyzed for 49 PAH analytes utilizing liquid extraction prior to normal phase solid phase extraction (LE-SPE) as described elsewhere (Sørensen et al., 2016) (Figure 3.1, bottom right panel).

3.3.4 Cardiovascular Imaging

At hatch (12-14 dpf), 20 larvae from each of four replicate tanks were immobilized with 0.01% MS 0.222 and mounted laterally in petri dishes with methyl cellulose or 1.5% agarose gels with slots molded by glass capillary tubes. Seawater was maintained between 9-11.5°C depending on current ambient temperature of source tanks by a microscope stage chiller. Ten second videos focusing on the atrium and ventricle were taken with Unibrain Fire-i400 1394 cameras (San Ramon, CA) at the highest magnification (6-8x) on Nikon SMZ-800 stereomicroscopes (Irvine, CA). BTV Pro software (Bensoftware.com) was used for imaging. Treatments were imaged randomly throughout an 8-hour period.

Heart videos were blinded for treatment and randomly analyzed for function and morphology using ImageJ software (NIH). Prior to measurements, video was assessed for appropriate quality and contrast and excluded if morphology of interest was not visible. For larvae at hatch, atrium and ventricle areas during systole and diastole were measured. Atrial and ventricle chamber areas were measured by drawing around the perimeter during systole and diastole three times for each embryo and larva.

3.3.5 VHSV Challenge and mx1 gene expression

Juvenile herring exposed to VHSV from all three populations (AK-S, AK-P, WA-P) were the survivors of embryonic exposure to oil at each of two concentrations: (un-oiled controls - 0.01 µg/L, and medium-high oil - 1.42-1.67 µg/L). The relative susceptibilities to VHSV caused by each of these treatments, including a no virus control for each oil treatment, were evaluated using triplicate tanks, with each tank containing 70 herring. 3 mL of stock VHSV (batch 2019BB, titer 2.4x108 PFU/mL) were added to the treatment tanks to realize 2.66x 103 PFU/mL. Supply water was shut off and the air turned up during the 2 hr VHSV exposure period. Unexposed control treatments were exposed to an analogous volume of carrier solution in lieu of VHSV. VHSV tissue titer was measured from freshly dead fish (from kidney and spleen).

For mx1 gene expression analysis, four replicate fish per treatment were subsampled from the triplicate tanks on days 0 and 5 post VHSV exposure (n= 4 per time point). Total RNA was extracted by homogenizing samples in TRIzol (5% v/v, ThermoFisher Scientific) and then

purified with an RNeasy 96 Kit (Qiagen). RNA concentrations and purities were measured fluorometrically using a Qubit™ RNA Broad Range kit (ThermoFisher Scientific). Superscript IV (Thermofisher) with oligo dt(20) primers was used to synthesize cDNA. Reverse transcriptase quantitative polymerase chain reaction (qPCR) was performed on a CFX96 Touch Real-Time PCR Detection System (BioRad) with SYBR® Green master mix (Life Technologies). Gene-specific RT-qPCR primers were designed using Primer3 (http://bioinfo.ut.ee/primer3/) and synthesized by Integrated DNA Technologies, Inc. Herring-specific target sequences for primer design were identified by BLAST searches based on sequences from either an Atlantic haddock (*Melanogrammus aeglefinus*) embryonic transcriptome (Sørhus et al., 2017, 2016) or the coding region of zebrafish genes (National Center for Biotechnology Information online database). The herring genes identified by this approach were annotated using BLAST-based sequence alignments. Target (*mx1*) and reference genes (*ef1a, mtm1,* and *wdtc1*) were selected based on a lack of oil exposure treatment effect from the previously published Atlantic haddock embryonic and larval transcriptomes (Sørhus et al., 2017). For *mx1* expression, normalized Ct values (dCt) and relative fold-change values were calculated using the Comparative Ct method using the geometric mean of the three reference genes (Schmittgen and Livak, 2008). A two-way ANOVA using R Studio (base package; version 1.4.1717) was used to evaluate the effect of oil, population, virus, and all pairwise and three-way interaction terms. When significant main or interaction effects were detected, a Tukey's post hoc test was used to compare means between treatments.

3.3.6 RNA Sequencing

Two pools of embryos (~30 embryos per pool) were sampled from each of 4 replicate tanks per treatment, at 4, 6, and 10 dpf for three embryonic time-points. Pacific herring samples were stored at -80°C prior to homogenization using 1.4 mm ceramic beads (Genaxxon Bioscience) on a TissueLyser (Qiagen). Ground lysates were suspended in 200 µl of Lysis binding buffer (Townsley et al., 2015), incubated at room temperature for 10 minutes, centrifuged for 10 minutes at 13,000 rpm, and the supernatant was retained. mRNA was extracted separately from each sample using oligo (dT) 25 beads (DYNABEADS directTM) to enrich polyadenylated mRNA. Strand specific RNA-seq libraries were prepared using the BRaD-seq protocol (Townsley et al., 2015), and used 14 cycles of PCR at the enrichment phase. Library preparations for all populations were identical. A random hexamer was used to prime fragments from along the length of the transcript:

GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNN,

where *N* represents a random nucleotide (Townsley et al., 2015). Finished libraries were quantified using the Quant-iTTM PicoGreen dsDNA high sensitivity kit (ThermoFisher), and normalized to 1ng/µl. We used 2ng per library and multiplexed up to 96 samples for sequencing. The libraries were sequenced in 4 lanes of an Illumina HiSeq X platform (Illumina, Inc.), generating 10.8 million paired reads per sample. Raw reads were quality checked with FastQC v.0.11.5 (Andrews 2018). Low quality reads (q<20), adapters, and reads less than 25-bp were removed using Trimmomatic v.0.36 (Bolger et al., 2014).

3.3.7 Read Mapping and Quantification

Trimmed reads were mapped to a reference transcriptome and mapped reads quantified with

Salmon v 1.3.0 (Patro et al., 2017) using the reference transcriptome (section 2.3.5) as an index in mapping-based mode. Average mapping rate for all paired-end read samples was 70.92%. The length-scaled TPM function in *tximport* was used to sum the transcript counts to the gene level (Soneson et al., 2015).

3.3.8 Differential Expression Analysis

We transformed read counts to stabilize the variance so that it was independent of the mean (where the variance neither increases systematically nor decreases systematically), according to Rocke et al. (Rocke et al., 2015). This was accomplished by regressing the gene-and-treatment variance on the gene-and-treatment mean for all samples. Once the mean variance relationship was defined, we estimated a transformation to the counts for which the slope was near zero (0.09e-8) which, practically speaking, stabilizes the variance, and applied this transformation across all gene counts. Low-abundance transcripts were filtered using the function filterByExpr where the minimum number of counts per transcript per day per treatment was set to 35 and the minimum sum of counts within a treatment was 500 (Robinson and Oshlack 2010). Counts were further normalized using CalcNormFactors, method=TMM. Normalized RNA-seq counts by sample were first graphed using multidimensional scaling (MDS) to visually identify primary sources of variance. The 96 well plate in which the sample library was prepared was found to be a significant source of variance across the dataset (batch effect). Therefore, a linear model (lm() function in RStudio Version 1.1.463) was used to evaluate the effect of population and oil treatment, with plate effect as a random effect in the model: lm(Count ~ Population * Oil + (1 | Plate)). For each developmental stage (day post fertilization), the linear model was set up with

population (AK-P, AK-S, WA-P) as a factor, oil exposure as a cofactor (expressed mean total PAHs per dry weight at 10 dpf per treatment), while plate effect was a random effect in the model. All p values were corrected with the False Discovery Rate (FDR) algorithm. Using this output we were able to select gene lists that have a significant population effect, and/or oil dose effect.

Differentially expressed genes in oil-exposed embryos, relative to unoiled control embryos, were defined as having higher or lower levels of expression (FDR <0.1). Individual genes were categorized across populations based on their dose-response at each time-point. This was achieved by calculating the slope of the regression at each in each population for genes with a significant oil-dose by population interaction effect (FDR<0.1), comparing a gene's absolute slope values across populations, then assigning a gene to a population based on the slope with the greatest value (i.e., largest dose response). Limma's function *removeBatchEffect* was used to remove the library plate effect for all visual representation of the data (Limma; Ritchie et al., 2015). Further visualizations utilized the *pheatmap* package (Kolde and Kolde, 2018) in R to generate heat maps, the *factoextra* package (Kassambara and Mundt, 2020) to generate PCA plots, and the *ggballoonplot* package to generate figure 2.3A and figure 2.4.

3.3.9 Functional Gene Enrichment Analysis

Gene functional enrichment analysis was performed with *ShinyGO V0.76* (Ge, Jung, and Yao 2020) using the Gene Ontology (GO) resource, and the Kyoto Encyclopedia Genes and Genomes

(KEGG) pathways database. Terms with FDR-corrected p values of < 0.1 were considered significantly enriched within pathways.

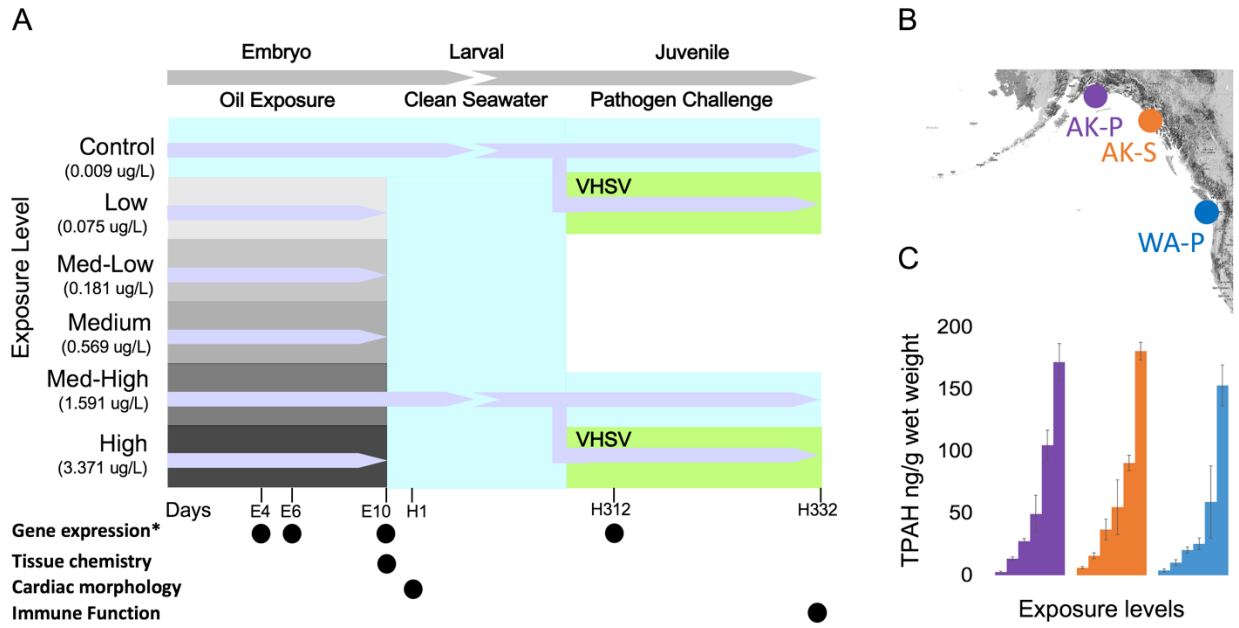

Figure 3.1: Map of sampling locations, experimental design schematic, and TPAH concentrations in embryo tissue. A) Schematic depicting experimental design showing length of exposure, time of hatching, and period of imaging/tissue collection/sample collecting for embryonic oil exposure and juvenile VHSV challenge experiments. Across all concentration gradients (Control-High dose), ANSCO was administered to the three populations as a continuous flow-through of dispersed oil droplets from 1 dpf - 10 dpf. Surviving fish from two oil treatments (Control and Medium-High) were hatched in clean seawater (~12 dpf) and grown-out through the juvenile stage, at which time they were exposed to either VHSV or no virus control. B) Map of sampling locations in Southern Alaska and Puget Sound of the three populations (AK-P Prince William Sound, AK-S Sitka Sound, WA-P Puget Sound). C) Embryo concentrations of TPAHs of ANSCO measured at the end of the exposure period (10 dpf) for each population (AK-P, AK-S, WA-P). Data is presented as mean ± SEM (n=4).

3.4 Results

3.4.1 Patterns of population variation

Individuals for this study were collected from Prince William Sound, AK, Sitka Sound, AK, and Puget Sound, WA, AK-P, AK-S, WA-P, respectively. Moderate population genetic structure exists between representative samples from the Alaska populations and geographically distant

Washington population. Pairwise F_{ST} was calculated using low-coverage (~1.5x) whole genome sequence data collected from ~50 individuals from each of the three focal populations (unpublished data). The F_{ST} between the two Alaska populations (AK-S vs AK-P= 0.0084) was lower than F_{ST} between Washington and each of the Alaska populations (WA-P vs AK-P = 0.0108, WA-P vs AK-S = 0.0096). This pattern of genetic differentiation is consistent with isolation by distance, where the more geographically isolated population (Washington) was more genetically distinct. This pattern of presumably neutral genetic variation sets the context for our null expectation of population variation in other traits (e.g., molecular and physiological responses to oil). If traits of the Alaska populations are more similar to each other than compared to the WA population, this would be consistent with neutral expectations. In contrast, if for example traits of the AK-P population are most different from the other two populations, then this would reject neutral expectations. The pattern for genes that vary in expression among populations (excluding genes that show an oil treatment effect) highlights WA-P as an outlier, consistent with our null expectation, which is to be expected since the majority of gene expression variation is predicted to evolve by random-neutral processes (Figure 3.2 and Figure 3.4, right panel) (Whitehead and Crawford 2006). In the absence of key external stressors of oil and pathogens, we establish a null hypothesis, that of neutral genetic drift, as the main driver of variation in gene expression among these populations. Under this hypothesis, we would expect the AK-P and AK-S populations to also be most similar for traits such as sensitivity to oil exposure and virus challenge, including the underlying molecular responses.

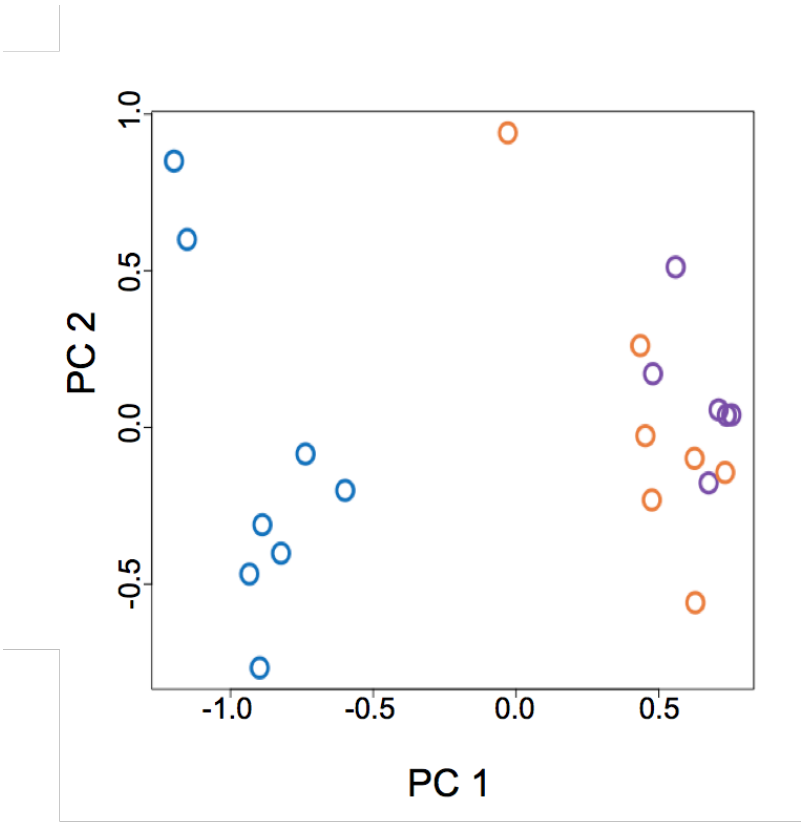

Figure 3.2: Population variation in gene expression. Principal Components Analysis represents the variation in population specific gene expression for genes significant for a main effect of population (FDR<0.001), excluding genes with a main effect of oil and a significant oil x population interaction effect (FDR>0.1). Circles represent treatment means for 10 dpf samples, and colors represent population assignment. PC1 discriminates transcriptional variation between the two Alaska populations (AK-P, AK-S; purple and orange, respectively) and the Washington population (WA-P; blue).

3.4.2 Oil exposure effects on developing heart morphology

Cardiovascular morphology was assessed at hatch, several days after oil exposure had terminated. At this stage, the atrium is superior to the ventricle along the midline and the ventricle is beginning to balloon or extend posteriorly (see Figure 3.3 inset for depiction). Embryonic crude oil exposure caused concentration-dependent changes to cardiac morphology, most conspicuously on the size of the posterior ventricular outgrowth (PVO). These effects mirror those detected at a similar TPAH dose range (Incardona et al. 2021), but varied by source

population. WA-P larvae were the least affected by embryonic exposure with a lowest observed effect concentration (LOEC) on PVO area of 153.1 ng/g TPAHs (high dose) in comparison to AK-P and AK-S larvae (64.5 ng/g TPAH and 67.8 ng/g TPAH, both medium doses, respectively) (Figure 3.3). Development of ventricular ballooning was highly sensitive to crude oil exposure, particularly for AK-P and AK-S larvae which experienced 20 and 16% decreases in PVO areas, respectively, at the 0.53 µg/L TPAH aqueous concentration (medium dose). Overall, the Alaska populations (AK-P and AK-S) show similarly elevated sensitivity to oil-induced cardiotoxicity compared to the Washington population (WA-P).

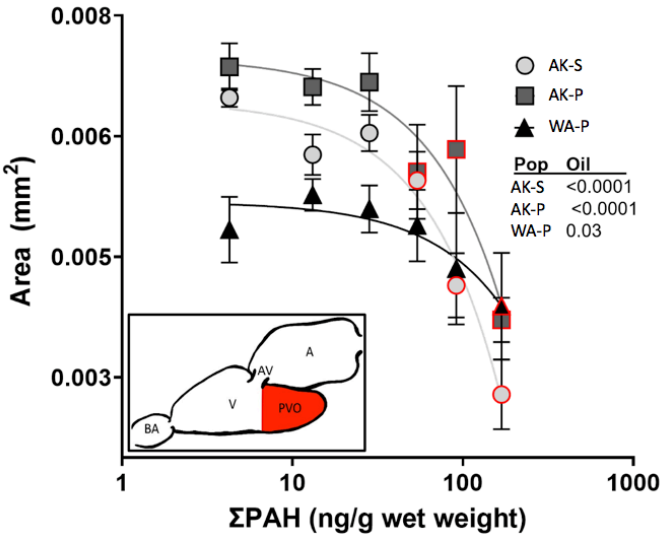

Figure 3.3: Posterior ventricular outgrowth (PVO) area at hatch, indicative of ventricular ballooning, in response to embryonic oil exposure for Sitka Sound (AK-S), Prince William Sound (AK-P), and Washington (WA-P) larvae. Oil effect by population (Pop) is depicted in a table within the figure (significant oil effect if $p<0.05$). PVO area varied by population ($p<0.0001$) and an interaction between population and oil ($p<0.0001$). Linear regression outcomes for dose response are depicted in a table within the figure. Points outlined in red indicate a significant difference ($p<0.05$) from the control (Tukey's post-hoc). Data are presented as mean ± SEM (n=20 larvae/4 tanks). The inset depicts the PVO within the context of a Pacific herring atrium (A), ventricle (V), atrioventricular valve (AV), and bulbous arteriosus (BA).

3.4.3 Gene expression: conserved population response to oil

Genome-wide whole animal transcriptome profiling is important for uncovering molecular mechanisms underlying the long-term impacts to herring health and fitness from embryonic exposure to crude oil. We first isolate the genes responsive to oil doses that are common across populations to provide insight into conserved core mechanisms. The number of genes involved in a conserved dose response to embryonic oil exposure across source populations increased with dose and time. The number of genes showing a transcriptional response in the herring embryo increased with dose, and increased from early to later developmental days. The number of genes showing a dose-response at 4 dpf, 6 dpf, and 10 dpf was 16, 19, and 1695, respectively (FDR<0.1) (Figure 3.4, left panel). At 4 dpf, genes that are differentially regulated during oil exposure were enriched for the KEGG pathway *Metabolism of Xenobiotics by Cytochrome P450* (KEGG Enrichment FDR: 0.003, Fold Enrichment: 6.2), which contained *cyp1a1, cyp1b1,* and *ugt1b5*. *cyp1a1* is a widely used biomarker for phase 1 xenobiotic metabolism, including PAHs. Like *cyp1a1, cyp1b1* also belongs to the phase 1 family of xenobiotic genes. *ugt1b5* encodes a phase 2 enzyme. As expected, these genes, as well as two additional xenobiotic metabolism genes (*cyp1c1* and *cyp1a2*), were induced for both oil exposure levels (Medium Low, Medium High). The conserved oil dose response at day 6 was largely characterized by an upregulated transcriptional response of a battery of xenobiotic response genes including, *cyp1a, cyp1a1, cyp1a2, cyp1c1, cyp1b1, ahrra, ugt1b5, and sult6b1* the latter of which is a newly recognized sulfotransferase involved in benz(a)anthracene metabolism. The dose response at 10 dpf is characterized by significantly greater transcriptional perturbation (1695 genes) than earlier time-points. The 1695 genes with a conserved oil dose

response were divided into up- and down-regulated genes (726 and 696, respectively) for functional enrichment analysis. There was no significant enrichment of KEGG pathways in 10 dpf induced genes. *Endocytosis*, however, was significantly enriched in genes with reduced expression relative to controls (KEGG Enrichment FDR: 0.004, Fold Enrichment: 1.3). The endocytosis pathway is responsible for the transport of exogenous nutrients, proteins and ligands into the cytoplasm. Additional key upregulate genes remain those involved in xenobiotic metabolism (*cyp1a1, cyp1a2, cyp1c1, cyp1b1, cyp39a1, ahrra, ugt1a1, ugt1b5, and sult6b1*).

A targeted analysis of oil-induced transcriptional response explored cardiac-related differential gene expression. A manually curated gene-set was selected for its involvement in heart development, function, and injury-repair response (*Cardiogenesis* panel). Enrichment analysis was conducted on those genes with a significant dose-response for each developmental time-point (conserved oil response and unique oil response). Among the genes with a conserved dose response across all populations at 4 and 6 dpf, none of the genes from the *Cardiogenesis* panel were present. Among genes significant for a conserved dose response at 10 dpf, 31 belonged the *Cardiogenesis* panel. However, *Cardiogenesis* was not statistically significantly enriched at 10 dpf. This gene panel, however, was significantly enriched at 6 dpf for genes with a population-dependent response to oil exposure (p= 0.07, fold enrichment= 1.75).

3.4.4 Gene Expression: Population-Specific Response to Oil

Differences in gene expression in response to oil exposure distinguish the Washington from

Alaska populations (Figure 3.4 middle panel). In contrast to the conserved response to oil exposure among the population, more genes showed a population-dependent response to oil exposure at earlier time-points (4 dpf) relative to later developmental time-points (6, 10 dpf). To find which genes are unique to each population, we correlated expression patterns across oil doses for each gene. WA-P contains the most genes with a unique dose response relative to the two Alaska populations (AK-P, AK-S) at every time-point (4 dpf: 91%, 6 dpf: 80%, 10 dpf: 74%). Functional enrichment analysis was used to identify molecular mechanisms that distinguish the molecular response of the Washington population from the Alaska populations. The *Intestinal Immune Network for IgA Production* pathway is significantly enriched for the WA-P specific dose-response at 4 dpf (KEGG Enrichment FDR: 0.03, Fold Enrichment: 11.19). Immunoglobulins are abundant in the fish mucosal system, acting as a first line of defense against disease (Mashoof and Criscitiello 2016). To gain further insights into the Washington population's response to oil we used the Enrichr database for functional enrichment analysis (Kuleshov et al. 2016) (FishEnrichr; Phenotype AutoRIF Predicted Z score). We found that the *Cardiac muscle cell action potential* pathway was enriched among population-dependent oil-responsive genes at 4 dpf (Phenotype.AutoRIF.Predicted FDR: 0.04, Fold Enrichment: 3.5; 13 genes). These early developmental cardiac genes were all down-regulated by oil exposure in the medium-high dose in the WA-P population. In contrast, only a minority of these genes were down-regulated at medium-high dose in both Alaskan populations. The KEGG *Cardiac muscle contraction* pathway was enriched among population-dependent oil-responsive genes at 6 dpf (Enrichment FDR: 0.07, Fold Enrichment: 7.36). Expression of genes in this pathway were all down-regulated at the high dose in the Washington population. Enrichr identified the same pathway, *Cardiac*

muscle cell action potential, as significantly enriched among population-dependent oil-responsive genes at day 4 and the same set of 14 genes within this pathway were all down-regulated at the high dose in WA-P. (Phenotype.AutoRIF.Predicted FDR: 0.04, Fold Enrichment: 3.6). No significant pathways at 10 dpf were enriched for the WA-P unique oil dose-response. Overall, these patterns cannot reject the null hypothesis that random-neutral drift is responsible for population variation in their response to oil.

Expression patterns of population-dependent oil responsive genes from in the manually curated *Cardiogenesis* panel were correlated among the two Alaska populations (AK-P, -S) with WA-P as an outlier. This agrees with the larger set of genes significant for a population-dependent oil response (Figure 3.4). Furthermore, all nine genes (*tpm4a, actc1c, actr3, ednraa, myl4, cxcl12b, tpm1, tpm3, cxcl12a)* enriching the *Cardiogenesis* panel at 6 dpf were down-regulated at every oil level relative to controls (with the exception of *cxcl12b*, which was down-regulated at the highest dose only).

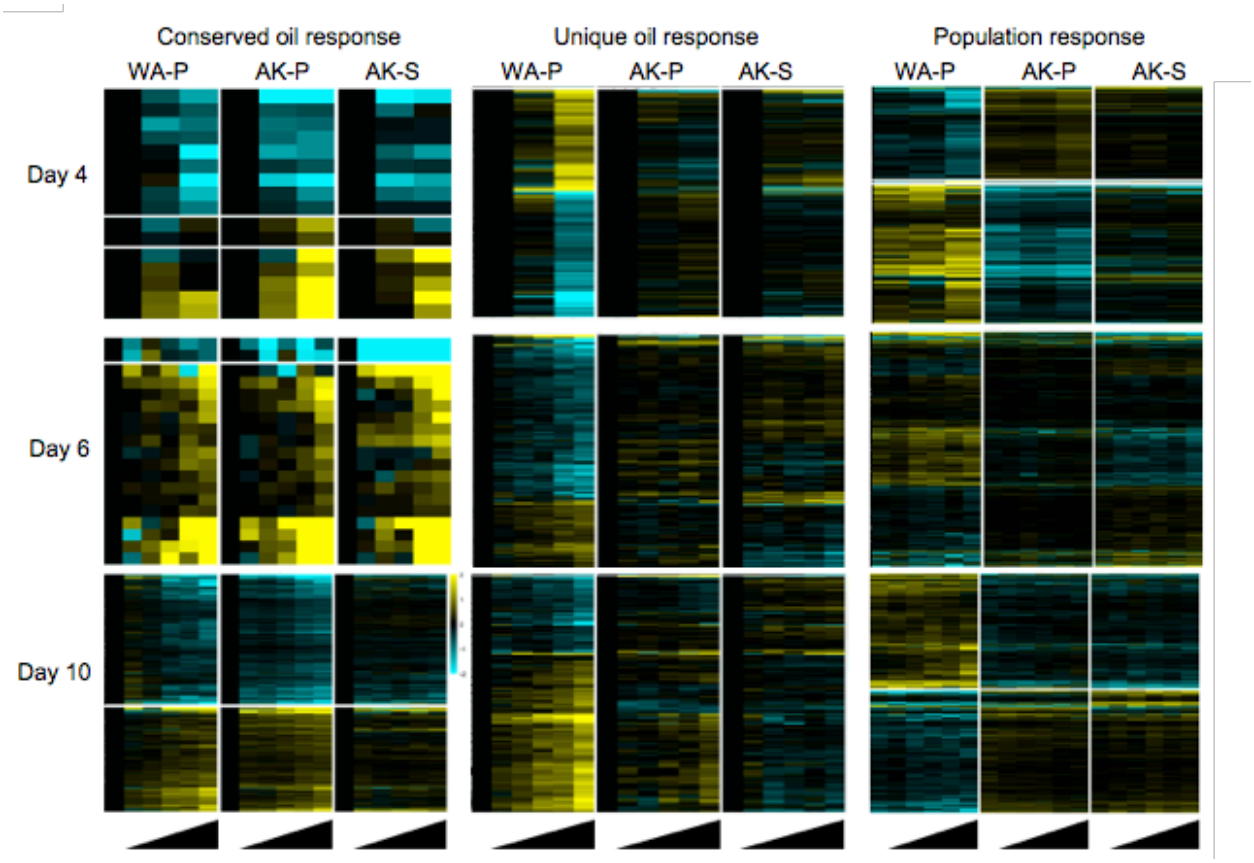

Figure 3.4: Fold change in gene expression for left panel) genes with a conserved dose response across populations, middle panel) genes with a significant dose by population interaction, and right panel) population variation (asterisk indicates this panel excludes genes with a dose response or interaction) at three developmental time-points (4, 6, 10 dpf) for three Pacific herring populations (WA-P, AK-P, AK-S). For each heatmap, the concentration of TPAHs increases from left to right, with only two doses (medium-low and medium-high) shown for 4 dpf. The mean trimmed mean of M values (TMM) was calculated from 6-8 biological replicates and was mean-centered normalized (yellow is upregulated, teal is downregulated). FDR was set to <0.1 for A and B and <0.001 for C.

3.4.5 Gene Expression: analysis of immune system pathways

Similar to the cardiac gene panel, a manually curated gene-set was selected for involvement in innate immune development and function. There was no enrichment of genes belonging to the manually curated *Immunogenic* panel with a significant dose response conserved across all

populations during the first two timepoints (4 dpf and 6 dpf). There were 173 *Immunogenic* panel genes with a conserved dose response at 10 dpf, the final day of oil exposure and 3 days prior to hatch. However, similar to the *Cardiogenesis* gene panel the *Immunogenic* gene panel was not enriched in genes showing a conserved response to oil exposure. For the population-dependent response to oil exposure, every developmental time-point contained genes from the *Immunogenic* gene panel, with an increasing abundance of genes as embryogenesis progressed (4,6,10 dpf). However, only at 6 dpf were *Immunogenic* panel genes significantly enriched (p= 0.03, fold enrichment= +1.37, N = 37 genes). WA-P had a unique expression profile for 89% of the genes in this list relative to the two Alaska populations, two of which, *cxcl12a* and *arf2b*, were also unique to WA-P at 4 dpf, linking the two time-points.

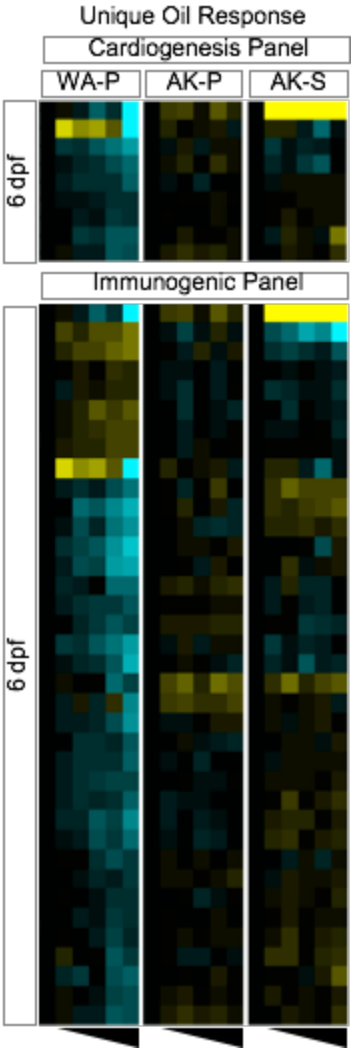

Figure 3.5: Gene expression of cardiac pathways and immune system panel. Heatmap visualizes genes enriching the manually curated cardiac and innate immune system gene panels at 6 dpf. Hypergeometric means testing was used to determine enrichment for manually curated gene panels. 6 dpf was the only time-point reaching statistical significance (p= 0.03, fold enrichment= +1.37). Days 4 and 10 were not significant and are not shown.

3.4.6 Effects of Embryonic Oil Exposure on Sensitivity to Virus Challenge

Pacific herring were similarly susceptible to VHSV across populations and oil doses measured by viral shedding and viral persistence (data not shown), and cumulative mortality (Figure 3.6). Cumulative mortalities were significantly greater in all populations and treatments that were VHSV-challenged relative to no virus controls (Figure 3.6 A, B, and C). Additionally, populations

previously exposed as embryos to oil did not differ in cumulative mortality as juveniles during virus challenge (Figure 3.6D). Among controls (no virus), VSHV was not isolated from spleen and viral shedding was not detected (data not shown).

Quantitative expression of *mx1*, an interferon-inducible viral response gene, significantly distinguishes the Washington population (WA-P) from the 2 Alaskan populations (AK-P, AK-S). This pattern of population varaition is consistent with random-neutral expectations (Figure 3.6). The induction of the *mx1* gene has previously been closely associated with VHSV-exposed Pacific herring (Hershberger et al. 2013), and is an biological indicator of virus infection. In our experiments, *mx1* expression was significantly upregulated in the WA-P population in both the oil and no-oil treatment groups during VHSV challenge relative to the pre-virus control treatment group (two-way ANOVA effect of virus only $p<0.0001$), demonstrating that embryonic oil exposure had no effect on *mx1* expression during juvenile VHSV exposure for the WA-P population (Figure 3.6E**)**. While *mx1* was upregulated in the oil-virus treatment group for the Alaskan populations (AK-P, AK-S), *mx1* was not induced in virus challenged fish in the unoiled exposure group (no oil-virus). This indicates that oil imposed a mediating effect on *mx1* expression during VHSV exposure in the Alaska populations only (two-way ANOVA effect of virus, oil, and virus-oil interaction $p<0.0001$). This suggests population variation in immune responses to VHSV exposure, but where those differences do not manifest as variable sensitivity to acute virus-induced mortality.

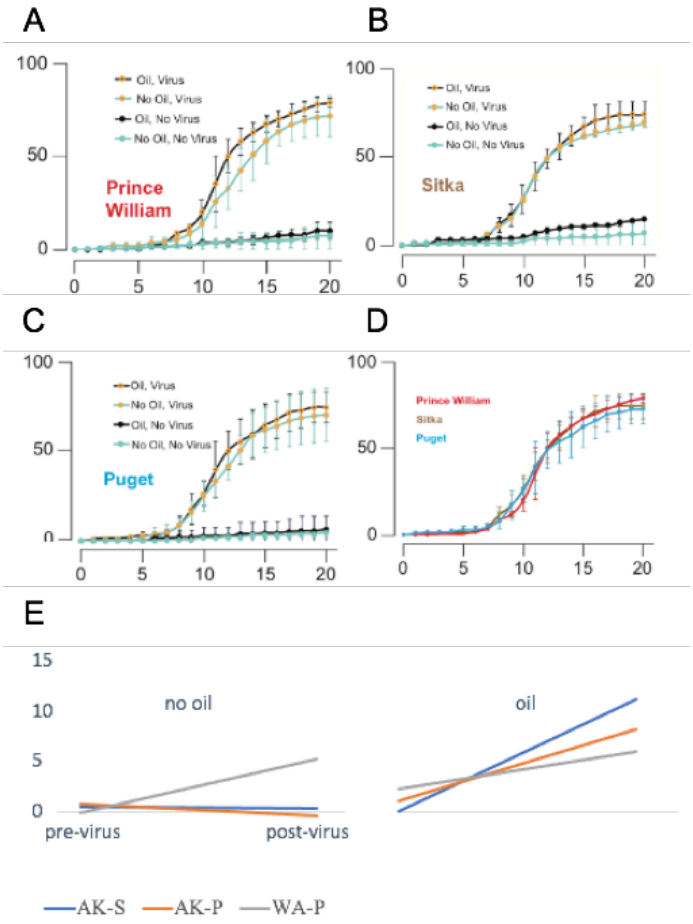

Figure 3.6: Survival of VHSV challenged juvenile Pacific herring, with and without embryonic exposure to oil.

A) AK-P, B) AK-S, and C) WA-P populations. D) Comparison of survival curves among populations. We only include the respective oil + virus treatments for each population. Bars indicate standard deviation. Data represent percentages corresponding to the means of arcsine-transformed proportions from triplicate treatment tanks. There were no significant differences between day 20 means between oil and no oil virus challenged fish (Tukey's HSD). E) Fold change in myxovirus (Influenza) Resistance Dynamin Like GTPase 1 (*mx1*) gene expression of herring juveniles for 3 populations, AK-P, AK-S, WA-P before and after VHSV exposure. Data represents mean fold-change relative to controls normalized to the geometric mean of three reference genes (See Methods). On the x-axes, pre-virus represents no virus control samples and post-virus represents virus exposed samples. Virus exposed samples were taken at experimental day 5 to measure a viral response before significant mortalities began. Left and right panels show no oil and oil exposure groups, respectively, before and during virus exposure. Significant differences ($p<0.05$) between populations after ANOVA and Tukey's post hoc analysis is indicated by an asterisk (n=4).

3.5 Discussion

The causes of the Prince William Sound Pacific herring fishery collapse in 1993-1994 remain controversial, and the reasons for their lack of recovery remain a mystery. The Exxon Valdez oil spill occurred three to four years prior to the collapse. While contaminating crude oil from the Exxon Valdez spill was largely absent from surface waters at the time of the collapse, the coincidence of events is conspicuous. Disease epizootics likely contributed to the collapse, and subsequent recurring outbreaks may be reasons for the lack of recovery (Marty 2008). Notably, closely related populations outside of AK-P were unaffected by the oil spill and did not experience a dramatic population decline at the time despite the presence of virus. We tested hypotheses about mechanistic connections between oil exposure and disease susceptibility, and whether these impacts may have resulted in evolutionary outcomes.

We designed experiments to test two hypotheses and their alternatives. Our first hypothesis was Adaptive selection. If adaptive selection happened then we would predict that oil exposure caused genotype-selective mortality, where surviving adults and their offspring had compromised immune function. We found no unique difference in the AK-P population's response to embryonic oil exposure and/or pathogen challenge relative to the two reference populations. We identified, however, variation among populations and in population response to oil that is consistent with random-neutral drift. Virus titers and mortality curves were similar across populations. However, the molecular immune response to virus infection varied among populations, again in a pattern that was consistent with random-neutral drift. We therefore conclude that the effect of divergent genetic background, as a function of random-neutral drift

caused by isolation-by-distance, accounts for population differences in sensitivity to oil, and the molecular responses that underlie them.

Our second hypothesis was Physiological impact. If oil exposure perturbed immune system development and function, then we would predict that oil exposure during early life impaired fish immune system development such that adults were more susceptible to disease. We found that exposure to crude oil during early embryogenesis does not compromise the ability of juveiles to survive VHSV exposure, and that populations were equally susceptible VHSV-induced mortality. However, since much of the immune system develops and matures after hatch, and oil exposures were confined to pre-hatch, it is plausible that oil exposure after hatch may perturb immune system development in ways that we could not detect. Additional studies, including additional windows of exposure during critical developmental periods, are necessary to fully test our oil-immune interaction hypothesis.

3.5.1 Patterns of population variation

The general patterns of population variation represented in Figure 3.2 is consistent with random neutral divergence caused by isolation-by-distance. Under a neutral drift model, the variation in a trait should be more similar among closely related populations than among more distantly related ones (Orr and Allen Orr 1998; Whitehead and Crawford 2006). The comparative population approach used here to identify oil-dose response genes depends on first accounting for the variation of neutral genetic drift apparent between geographically nearby and more distant populations. If population differences in traits are caused by random-

neutral drift, then we would expect that the geographically most distant population (WA-P in our experiment) would be most different. Alternatively, if environmental changes in PWS have driven evolved differences through natural selection, then we would expect traits of AK-P fish to differ relative to the two reference populations (AK-S and WA-P). We found that the two Alaska populations were most similar in their sensitivity to oil and virus exposure and the WA-P population was most different. We therefore conclude that EVOS oil-induced natural selection did not account for population variation in the traits that we measured.

3.5.2 Dose response heart morphometrics

To test the hypothesis of oil spill-induced natural selection in the AK-P population, we measured population variation in the developing heart, the canonical target of oil-induced injury in fish embryos. While embryos from AK-S, AK-P and WA-P all experienced cardiotoxicity in response to embryonic oil exposure, the severity of this toxicity differed. In accord with our null hypothesis (neutral drift) the Alaska populations (AK-P and AK-S) were most similar in their response to oil, and the Washington population (WA-P) most distinct. In addition to a similar response to oil exposure, the Alaska populations as a group were more sensitive to TPAHs, while the Washington population was relatively unaffected at similar doses (Figure 3.3). Within the Alaska populations, AK-P larvae exhibited a concentration-dependent oil-induced reduction in outgrowth of the posterior ventricle (Figure 3.3). Similar cardiotoxicity outcomes in response to similar oil concentration ranges has been observed in other studies (Incardona et al. 2021). The oil exposure levels used in our experiments were comparatively low relative to previous studies, and the heart injury phenotype observed here was mild for all populations, especially

WA-P. However, the concentration-dependent reduction in posterior ventricle outgrowth (particularly evident in the Alaska populations) has been shown to result in pathological remodeling of the ventricle and reduced cardiac performance in juveniles (Incardona et al. 2021). Our experimental design does not distinguish between neutral genetic drift and local adaptation in the Washington population as an explanation for its distinctive response.

3.5.3 Gene expression: conserved population response to oil

We extend the conventional indicators of abnormal heart development in early life stage herring following embryonic oil exposure to developmental patterns of genome-wide gene expression. Oil-perturbed gene expression and functional pathways can complement traditional injury assessments in the aftermath of oil spills, particularly in fish that appear grossly normal after exposure to sub-lethal doses of oil. Genes and pathways with a shared dose-dependent response to PAHs can even prove to be more sensitive diagnostic markers for embryonic oil-induced injury than morphometric or physiological endpoints that tend to be difficult to measure (Edmunds et al. 2015). Even at the trace levels of oil used in this study, a conserved response to oil exposure was measured at the earliest developmental time-point for 27 genes. Gene expression varied between developmental time points, with more genes eliciting a conserved dose response to increasing concentrations of TPAHs across populations in late (10 dpf) than early (4, 6 dpf) embryogenesis. Early developmental time-points (4, 6 dpf) were characterized by an upregulated set xenobiotic metabolism genes. However, at 10 dpf, genes enriched in the Endocytosis pathway were down-regulated. Endocytic subcellular organelles shuttle ligands and nutrients across the cell membrane and into intracellular compartments.

The enrichment of this pathway for downregulated genes could indicate that PAH metabolism is not as prevalent as in earlier time-points. At 4 dpf, the xenobiotic response is most prevalent, as PAHs accumulate and the development of metabolic capacity increases, leading to lower levels of PAH body burdens by 10 dpf. This is further supported by the 10 dpf specific induction of aryl-hydrocarbon receptor repressor (*ahrra*), a negative regulator of the AHR transcription factor in xenobiotic metabolism (Jenny et al. 2009).

3.5.4 Gene expression: population-specific response to oil

The reduced sensitivity to concentration-dependent ventricle outgrowth distinguishes the geographically distant and genetically distinct Washington population from the two Alaska populations. This distinction was reflected in genes significant for a population-by-dose response interaction effect (referred to here as population-specific dose response genes) where WA-P had more perturbed genes with higher fold change than both Alaska populations. In contrast to conserved dose-response genes, more genes elicited a dose-response earlier rather than later, at 4 dpf relative to 6 and 10 dpf. It appears that population transcriptional variation in response to oil converges over developmental time. This would also explain why there are relatively more conserved dose-response genes in late embryogenesis. It is clear from Figure 3.4, middle panel, that the WA-P population is remarkably different from the Alaska populations in their dose response across all developmental time-points. Correlated gene expression between populations confirms that WA-P is an outlier in its dose-response relative to the Alaska populations throughout embryogenesis. Intriguingly, the *Intestinal Immune*

Network for IgA Production pathway is significantly enriched for the WA-P specific dose-response at 4 dpf (KEGG Enrichment FDR: 0.03, Fold Enrichment: 11.19). At 4 dpf, herring are not feeding nor do they have a digestive system. However, immunoglobulin A plays a dominant role in mucosal immunity in fish, functioning as a neutralizing antibody on mucosal surfaces for toxins and viruses (Mashoof and Criscitiello 2016). WA-P is further distinguished from the two Alaska populations at 6 dpf by a unique set of oil-responsive genes significantly enriched for the KEGG Cardiac Muscle Contraction pathway, including three tropomyosin genes, *tpm1, tpm3*, and *tpm4a,* (Enrichment FDR: 0.07, Fold Enrichment: 5.59). Expression patterns varied across populations but all three genes were consistently downregulated across doses in WA-P. All three tropomyosins play a critical role in heart development. For example, *tpm1* knock-down in zebrafish causes ventricular trabeculae formation, increased apoptosis, and disrupted action potentials (Shih et al. 2015; England et al. 2017). Fewer genes had a significant population-dependent dose response at 10 dpf, and no pathways were significantly enriched for population-dependent oil-responsive genes at 10 dpf. These results suggest that the mechanism of reduced sensitivity to oil-induced cardiotoxicity in WA-P herring is related to increased resilience of their cardiac developmental pathways following exposures.

Overall, the outlier status of the WA-P population suggests neutral drift accounts for most of the population specific variation in response to embryonic oil exposure throughout embryogenesis. However, this does not rule out variation in gene expression resulting from natural selection in the Puget Sound WA region. While we cannot directly test for local adaptation in this population, given the nature of our population contrasts, it is intriguing to note that the WA-P samples (gametes) used in this study come from an historically distressed

stock located near Cherry Point in northern Puget Sound, WA. Cherry Point in Puget Sound is genetically distinct from other Washington herring, as well as from nearby British Columbia herring, and recruitment in this stock has fallen to extremely low levels since 1993, similar to AK-P (Kurt et al. 2012). Herring spawning success in these habitats is subject to environmental quality including hydrocarbons from vehicle combustion emissions and urban runoff. It is plausible that WA-P (Cherry Point) fish have a distinct xenobiotic metabolism physiology relative to the Alaska populations, so the opportunity to incur heart injury after PAH exposure is reduced and therefore protective.

3.5.5 Manually curated gene panels

We manually curated two genes panels to lend molecular support to physiological targets of oil injury. Genes included in the manually curated *Cardiogenesis* panel were selected for their involvement in heart development, function, or injury-repair response, likely underlying oil-induced heart injury in developing fish. We also curated a set of genes involved in innate immune system development and function. We hypothesized that early life exposure to oil impacts immune system development that would persist to affect later-life immune function, which may affect sensitivity to disease. This would provide a link between the EVOS oil spill and later VHSV-associated population collapse in PWS. These manually curated gene panel were not significantly enriched for genes with a population-conserved response to oil throughout embryogenesis. As is evident from figure 3.2, left panel, the few genes with a conserved dose response at 4 and 6 dpf were mostly those involved in the xenobiotic response. There was also

no enrichment of panel genes among those genes showing a conserved oil response at 10 dpf. However, the much larger subset of 10 dpf conserved dose response genes were enriched a single KEGG pathway, *Endocytosis*. In contrast, genes with a population-dependent response to oil exposure were significantly enriched in both gene panels at 6 dpf. As expected, for a majority of these genes, their expression patterns distinguished Washington fish from the two Alaska populations.

For the *Cardiogenesis* gene panel at 6 dpf, with the exception of *cxcl12b* which was downregulated at high dose only, every gene was down-regulated at every oil level relative to controls for WA-P. In contrast, gene expression was either slightly upregulated (AK-P) across most doses or slightly downregulated (AK-S). The long-term effect of oil induced cardiotoxicity visibly present in herring embryos results in an initial stage of reduced ventricle size (Incardona et al. 2021). During later larval stages pathological remodeling of the heart in these affected fish leads to ventricular hypertrophy, a relative increase in cardiomyocyte number and size (Incardona et al. 2021). If genes involved in cardiomyocyte proliferation are down-regulated in the WA-P population beginning in early embryogenesis, one may speculate that this may contribute to reduced oil-induced hypertrophy in those fish.

The immunogenic manually curated gene panel was enriched among population-dependent oil responsive genes at 6 dpf only. For the majority of these genes, WA-P showed a markedly

unique expression pattern relative to the two Alaska populations with a majority of genes for this population down-regulated across a majority of the doses.

3.5.6 Oil dose response virus challenge

The role of AhR in mediating immune function is well established. It has been demonstrated that AhR ligands, including environmental toxicants, bind AHR and alter both adaptive and innate immune development and function (Song et al. 2020; Quintana 2013). Impairments in AHR function could come at the cost of normal immune system function. Exposure to the 1989 Exxon Valdez oil spill could have been a strong selective force on the WA-P population. We tested responses to pathogen challenged juvenile herring for all three populations across two oil levels (medium-low and high) and no oil controls. All three populations were similarly susceptible to VHSV exposure across all measurements. Notably, unlike heart morphometrics and molecular indicators of cardiotoxicity, WA-P did not exhibit increased sensitivity to immune challenge from VHSV relative to the Alaska populations (Figure 3.6D). This suggests that population variation in historical exposure to oil or virus dose not affect sensitivity to VHSV mortality, nor does variation in genetic background. Additionally, oil exposure during early embryogenesis did not affect cumulative mortality to VHSV exposure as juveniles (Figure 3.6A, B, C). We are cautious to conclude that oil exposure during early life does not affect later-life sensitivity to pathogens. This is because our window of exposure was prior to much immune system development, such that exposures during other periods of immune system development could be consequential for later disease susceptibility. Furthermore, we tested for

only innate immune responses, where adaptive immune responses could have been modulated by oil exposure. And we tested only acute mortality in response to one pathogen, where other aspects of sensitivity, and in response to other pathogens, could be perturbed by oil. Further experiments are necessary to fully explore the hypothesis that early life exposure to oil may affect later-life sensitivity to disease.

Expression of the *mx1* gene, an interferon-inducible viral response gene (Hershberger et al. 2013), was different between the WA-P population compared to the two Alaskan populations. This pattern of population variation is consistent with the influence of random-neutral genetic drift caused by isolation-by-distance (Figure 3.6E), though we cannot rule out the influence of natural selection driving this divergence in the Washington population. Independent of previous oil exposure, WA-P, relative to both Alaska populations, is uniquely responsive to virus infection as measured by *mx1* expression. While this did not translate into decreased mortality for the WA-P population, we can speculate that a less virulent pathogen than VHSV may result in increased survivorship curves for WA-P. The mediating effect of embryonic oil exposure on *mx1* expression in the Alaska populations alone suggests that they have immune systems uniquely sensitive to contaminating PAHs. Additionally, why *mx1* was not induced by virus in these populations for fish that hadn't been exposed to oil in early life (control treatment group) is unclear. *mx1* induction is a reliable marker for virus exposure in Pacific herring (Hershberger et al. 2013). However, this is the first time to our knowledge that fish from Prince William Sound, AK, and Sitka Sound, AK, have been used in such a study. Additional studies that examine a detailed time-course of mx1 expression following virus challenge may provide

additional insights about population differences in immune response dynamics.

In summary, these results demonstrate that random neutral genetic drift caused by isolation-by-distance contributes to population variation in sensitivity to oil, and the molecular response to oil exposure, in Pacific herring. Oil at very low doses causes cardiac impairment in fish, where Alaska populations are more sensitive than fish from Washington. The key mechanism distinguishing Alaska from the Washington populations is a greater sensitivity to oil-induced perturbations of ventricular ballooning during embryogenesis. However, brief embryo oil exposure does not modulate the lethal effects of acute VHSV exposure. Populations are equally sensitive to lethal effects of VHSV, but, virus-induced expression of *mx1*, a marker of innate immune response, does differ between populations, as does the modulating effect of early life oil exposure. This indicates that oil may have a subtle immunomodulatory effects in the Alaska populations. Shifting the window of oil exposure to later in development may perturb key immune developmental pathways that would in turn impact the immune response to disease in Pacific herring. To test this future studies could shift the oil exposure window to larvagenesis followed by VHSV challenge. Furthermore, a similar experimental design, instead focusing on the WA-P population, would be useful in determining whether their unique response to oil is the result of random genetic drift or natural selection.

4. References

Ameen, Mohamed, Laksshman Sundaram, Mengcheng Shen, Abhimanyu Banerjee, Soumya Kundu, Surag Nair, Anna Shcherbina, et al. 2022. “Integrative Single-Cell Analysis of Cardiogenesis Identifies Developmental Trajectories and Non-Coding Mutations in Congenital Heart Disease.” *Cell* 185 (26): 4937–53.e23.

Anderson, J. W., J. M. Neff, B. A. Cox, H. E. Tatem, and G. M. Hightower. 1974. “Characteristics of Dispersions and Water-Soluble Extracts of Crude and Refined Oils and Their Toxicity to Estuarine Crustaceans and Fish.” *Marine Biology*. https://doi.org/10.1007/bf00394763.

Andrews, S. (2010). FastQC: A Quality Control Tool for High Throughput Sequence Data [Online]. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Arkoosh, Mary R., Edmundo Casillas, Ethan Clemons, Anna N. Kagley, Robert Olson, Paul Reno, and John E. Stein. 1998. “Effect of Pollution on Fish Diseases: Potential Impacts on Salmonid Populations.” *Journal of Aquatic Animal Health*. https://doi.org/10.1577/1548-8667(1998)010<0182:eopofd>2.0.co;2.

Arkush, Kristen D., Holly L. Mendonca, Anne M. McBride, Susan Yun, Terry S. McDowell, and Ronald P. Hedrick. 2006. “Effects of Temperature on Infectivity and of Commercial Freezing on Survival of the North American Strain of Viral Hemorrhagic Septicemia Virus (VHSV).” *Diseases of Aquatic Organisms* 69 (2-3): 145–51.

Asai, Rieko, Yukiko Kurihara, Kou Fujisawa, Takahiro Sato, Yumiko Kawamura, Hiroki Kokubo, Kazuo Tonami, et al. 2010. “Endothelin Receptor Type A Expression Defines a Distinct Cardiac Subdomain within the Heart Field and Is Later Implicated in Chamber Myocardium Formation.” *Development* 137 (22): 3823–33.

Beffagna, Giorgia. 2019. “Zebrafish as a Smart Model to Understand Regeneration After Heart Injury: How Fish Could Help Humans.” Frontiers in Cardiovascular Medicine. https://doi.org/10.3389/fcvm.2019.00107.

Berchtold, Martin W., Triantafyllos Zacharias, Katarzyna Kulej, Kevin Wang, Raffaela Torggler, Thomas Jespersen, Jau-Nian Chen, Martin R. Larsen, and Jonas M. la Cour. 2016. “The Arrhythmogenic Calmodulin Mutation D129G Dysregulates Cell Growth, Calmodulin-Dependent Kinase II Activity, and Cardiac Function in Zebrafish.” *The Journal of Biological Chemistry* 291 (52): 26636–46.

Billiard, Sonya M., Mark E. Hahn, Diana G. Franks, Richard E. Peterson, Niels C. Bols, and Peter V. Hodson. 2002. “Binding of Polycyclic Aromatic Hydrocarbons (PAHs) to Teleost Aryl Hydrocarbon Receptors (AHRs).” *Comparative Biochemistry and Physiology. Part B, Biochemistry & Molecular Biology* 133 (1): 55–68.

Bolger, Anthony M., Marc Lohse, and Bjoern Usadel. 2014. “Trimmomatic: A Flexible Trimmer for Illumina Sequence Data.” *Bioinformatics* 30 (15): 2114–20.

Bonatesta, Fabrizio, Cameron Emadi, Edwin R. Price, Yadong Wang, Justin B. Greer, Elvis Genbo Xu, Daniel Schlenk, Martin Grosell, and Edward M. Mager. 2022. “The Developing Zebrafish Kidney Is Impaired by Deepwater Horizon Crude Oil Early-Life Stage Exposure: A Molecular to Whole-Organism Perspective.” *The Science of the Total Environment* 808 (February): 151988.

Brown, E. D., B. L. Norcross, and J. W. Short. 1996a. “Introduction to Studies on the Effects of the (Exxon Valdez) Oil Spill on Early Life History Stages of Pacific Herring, (Clupea Pallasi), in Prince William Sound, Alaska.” *Canadian Journal of Fisheries and Aquatic Sciences*. https://doi.org/10.1139/cjfas-53-10-2337.

Brown, E. D., B. L. Norcross, and J. W. Short.. 1996b. “Introduction to Studies on the Effects of the (Exxon Valdez) Oil Spill on Early Life History Stages of Pacific Herring, (Clupea Pallasi),

in Prince William Sound, Alaska." *Canadian Journal of Fisheries and Aquatic Sciences. Journal Canadien Des Sciences Halieutiques et Aquatiques* 53 (10): 2337–42.
Carls, M. G., M. M. Babcock, P. M. Harris, G. V. Irvine, J. A. Cusick, and S. D. Rice. 2001. "Persistence of Oiling in Mussel Beds after the Exxon Valdez Oil Spill." *Marine Environmental Research* 51 (2): 167–90.
Carls, M. G., G. D. Marty, and J. E. Hose. 2002. "Synthesis of the Toxicological Impacts of the*Exxon Valdez* oil Spill on Pacific Herring (*Clupea Pallasi*) in Prince William Sound, Alaska, U.S.A." *Canadian Journal of Fisheries and Aquatic Sciences*. https://doi.org/10.1139/f01-200.
Carreño Gutiérrez, Héctor, Sarah Colanesi, Ben Cooper, Florian Reichmann, Andrew M. J. Young, Robert N. Kelsh, and William H. J. Norton. 2019. "Endothelin Neurotransmitter Signalling Controls Zebrafish Social Behaviour." *Scientific Reports* 9 (1): 3040.
Cherr, Gary N., Elise Fairbairn, and Andrew Whitehead. 2017. "Impacts of Petroleum-Derived Pollutants on Fish Development." *Annual Review of Animal Biosciences* 5 (February): 185–203.
Chhangawala, Sagar, Gabe Rudy, Christopher E. Mason, and Jeffrey A. Rosenfeld. 2015. "The Impact of Read Length on Quantification of Differentially Expressed Genes and Splice Junction Detection." *Genome Biology* 16 (June): 131.
Dogra, Deepika, Suchit Ahuja, Hyun-Taek Kim, S. Javad Rasouli, Didier Y. R. Stainier, and Sven Reischauer. 2017. "Opposite Effects of Activin Type 2 Receptor Ligands on Cardiomyocyte Proliferation during Development and Repair." *Nature Communications* 8 (1): 1902.
Dorson, M., E. Quillet, M. G. Hollebecq, C. Torhy, and B. Chevassus. 1995. "Selection of Rainbow Trout Resistant to Viral Haemorrhagic Septicaemia Virus and Transmission of Resistance by Gynogenesis." *Veterinary Research* 26 (5-6): 361–68.
Dubansky, Benjamin, Andrew Whitehead, Jeffrey T. Miller, Charles D. Rice, and Fernando Galvez. 2013. "Multitissue Molecular, Genomic, and Developmental Effects of the Deepwater Horizon Oil Spill on Resident Gulf Killifish (Fundulus Grandis)." *Environmental Science & Technology* 47 (10): 5074–82.
Edmunds, Richard C., J. A. Gill, David H. Baldwin, Tiffany L. Linbo, Barbara L. French, Tanya L. Brown, Andrew J. Esbaugh, et al. 2015. "Corresponding Morphological and Molecular Indicators of Crude Oil Toxicity to the Developing Hearts of Mahi Mahi." *Scientific Reports*. https://doi.org/10.1038/srep17326.
England, Jennifer, Javier Granados-Riveron, Luis Polo-Parada, Diji Kuriakose, Christopher Moore, J. David Brook, Catrin S. Rutland, et al. 2017. "Tropomyosin 1: Multiple Roles in the Developing Heart and in the Formation of Congenital Heart Defects." *Journal of Molecular and Cellular Cardiology* 106 (May): 1–13.
England, Jennifer, and Siobhan Loughna. 2013. "Heavy and Light Roles: Myosin in the Morphogenesis of the Heart." *Cellular and Molecular Life Sciences: CMLS* 70 (7): 1221–39.
Galli, Lisa M., Roeben N. Munji, Susan C. Chapman, Ann Easton, Lydia Li, Ouma Onguka, Joseph S. Ramahi, et al. 2014. "Frizzled10 Mediates WNT1 and WNT3A Signaling in the Dorsal Spinal Cord of the Developing Chick Embryo." *Developmental Dynamics: An Official Publication of the American Association of Anatomists* 243 (6): 833–43.
Gao, Chen, Shuxun Ren, Jae-Hyung Lee, Jinsong Qiu, Douglas J. Chapski, Christoph D. Rau, Yu Zhou, et al. 2016. "RBFox1-Mediated RNA Splicing Regulates Cardiac Hypertrophy and Heart Failure." *The Journal of Clinical Investigation* 126 (1): 195–206.
Garcia-Puig, Anna, Jose Luis Mosquera, Senda Jiménez-Delgado, Cristina García-Pastor, Ignasi Jorba, Daniel Navajas, Francesc Canals, and Angel Raya. 2019. "Proteomics Analysis of Extracellular Matrix Remodeling During Zebrafish Heart Regeneration." *Molecular & Cellular Proteomics: MCP* 18 (9): 1745–55.
Gassmann, Aaron J., Yves Carrière, and Bruce E. Tabashnik. 2009. "Fitness Costs of Insect

Resistance toBacillus Thuringiensis." *Annual Review of Entomology* 54 (1): 147–63.
Gauvrit, Sébastien, Jaclyn Bossaer, Joyce Lee, and Michelle M. Collins. 2022. "Modeling Human Cardiac Arrhythmias: Insights from Zebrafish." *Journal of Cardiovascular Development and Disease* 9 (1). https://doi.org/10.3390/jcdd9010013.
Ge, Steven Xijin, Dongmin Jung, and Runan Yao. 2020. "ShinyGO: A Graphical Gene-Set Enrichment Tool for Animals and Plants." *Bioinformatics* 36 (8): 2628–29.
Griffin, F. J., M. C. Pillai, C. A. Vines, J. Kääriä, T. Hibbard-Robbins, R. Yanagimachi, and G. N. Cherr. 1998. "Effects of Salinity on Sperm Motility, Fertilization, and Development in the Pacific Herring, Clupea Pallasi." *The Biological Bulletin* 194 (1): 25–35.
Gunawan, Felix, Alessandra Gentile, Ryuichi Fukuda, Ayele Taddese Tsedeke, Vanesa Jiménez-Amilburu, Radhan Ramadass, Atsuo Iida, Atsuko Sehara-Fujisawa, and Didier Y. R. Stainier. 2019. "Focal Adhesions Are Essential to Drive Zebrafish Heart Valve Morphogenesis." *The Journal of Cell Biology* 218 (3): 1039–54.
Haas, BJ. https://github.com/TransDecoder/TransDecoder.
Hashimoto, Hisayuki, Zhaoning Wang, Glynnis A. Garry, Venkat S. Malladi, Giovanni A. Botten, Wenduo Ye, Huanyu Zhou, et al. 2019. "Cardiac Reprogramming Factors Synergistically Activate Genome-Wide Cardiogenic Stage-Specific Enhancers." *Cell Stem Cell* 25 (1): 69–86.e5.
Hay, D. E. 1985. "Reproductive Biology of Pacific Herring (*Clupea Harengus Pallasi*)." *Canadian Journal of Fisheries and Aquatic Sciences*. https://doi.org/10.1139/f85-267.
Heintz, Ron A. 2007. "Chronic Exposure to Polynuclear Aromatic Hydrocarbons in Natal Habitats Leads to Decreased Equilibrium Size, Growth, and Stability of Pink Salmon Populations." *Integrated Environmental Assessment and Management* 3 (3): 351–63.
Hendry, Andrew P., Thomas J. Farrugia, and Michael T. Kinnison. 2008. "Human Influences on Rates of Phenotypic Change in Wild Animal Populations." *Molecular Ecology* 17 (1): 20–29.
Henryon, Mark, Alfred Jokumsen, Peer Berg, Ivar Lund, Per B. Pedersen, Niels J. Olesen, and Wilhelmina J. Slierendrecht. 2002. "Genetic Variation for Growth Rate, Feed Conversion Efficiency, and Disease Resistance Exists within a Farmed Population of Rainbow Trout." *Aquaculture* 209 (1-4): 59–76.
Hershberger, P. K., M. K. Purcell, L. M. Hart, J. L. Gregg, R. L. Thompson, K. A. Garver, and J. R. Winton. 2013. "Influence of Temperature on Viral Hemorrhagic Septicemia (Genogroup IVa) in Pacific Herring, Clupea Pallasii Valenciennes." *Journal of Experimental Marine Biology and Ecology* 444 (June): 81–86.
Hicken, Corinne E., Tiffany L. Linbo, David H. Baldwin, Maryjean L. Willis, Mark S. Myers, Larry Holland, Marie Larsen, et al. 2011. "Sublethal Exposure to Crude Oil during Embryonic Development Alters Cardiac Morphology and Reduces Aerobic Capacity in Adult Fish." *Proceedings of the National Academy of Sciences of the United States of America* 108 (17): 7086–90.
Hirata, Hiromi, Takaki Watanabe, Jun Hatakeyama, Shawn M. Sprague, Louis Saint-Amant, Ayako Nagashima, Wilson W. Cui, Weibin Zhou, and John Y. Kuwada. 2007. "Zebrafish *relatively Relaxed* Mutants Have a Ryanodine Receptor Defect, Show Slow Swimming and Provide a Model of Multi-Minicore Disease." *Development*. https://doi.org/10.1242/dev.004531.
Hose, J. E., M. D. McGurk, G. D. Marty, D. E. Hinton, E. D. Brown, and T. T. Baker. 1996. "Sublethal Effects of the (E××on Valdez) Oil Spill on Herring Embryos and Larvae: Morphological, Cytogenetic, and Histopathological Assessments, 1989–1991." *Canadian Journal of Fisheries and Aquatic Sciences*. https://doi.org/10.1139/cjfas-53-10-2355.
Hsu, Po-Jui, Horng-Dar Wang, Yung-Che Tseng, Shao-Wei Pan, Bonifasius Putera Sampurna, Yuh-Jyh Jong, and Chiou-Hwa Yuh. 2021. "L-Carnitine Ameliorates Congenital Myopathy in a Tropomyosin 3 de Novo Mutation Transgenic Zebrafish." *Journal of Biomedical Science* 28 (1): 8.

Incardona, John P., Mark G. Carls, Heather L. Day, Catherine A. Sloan, Jennie L. Bolton, Tracy K. Collier, and Nathaniel L. Scholz. 2009. “Cardiac Arrhythmia Is the Primary Response of Embryonic Pacific Herring (Clupea Pallasi) Exposed to Crude Oil during Weathering.” *Environmental Science & Technology* 43 (1): 201–7.

Incardona, John P., Mark G. Carls, Larry Holland, Tiffany L. Linbo, David H. Baldwin, Mark S. Myers, Karen A. Peck, Mark Tagal, Stanley D. Rice, and Nathaniel L. Scholz. 2015. “Very Low Embryonic Crude Oil Exposures Cause Lasting Cardiac Defects in Salmon and Herring.” *Scientific Reports* 5 (September): 13499.

Incardona, John P., Tiffany L. Linbo, Barbara L. French, James Cameron, Karen A. Peck, Cathy A. Laetz, Mary Beth Hicks, et al. 2021. “Low-Level Embryonic Crude Oil Exposure Disrupts Ventricular Ballooning and Subsequent Trabeculation in Pacific Herring.” *Aquatic Toxicology* 235 (June): 105810.

Incardona, John P., Tanya L. Swarts, Richard C. Edmunds, Tiffany L. Linbo, Allisan Aquilina-Beck, Catherine A. Sloan, Luke D. Gardner, Barbara A. Block, and Nathaniel L. Scholz. 2013. “Exxon Valdez to Deepwater Horizon: Comparable Toxicity of Both Crude Oils to Fish Early Life Stages.” *Aquatic Toxicology* 142-143 (October): 303–16.

International Committee on Taxonomy of Viruses, M. H. V. Van Regenmortel, C. M. Fauquet, and D. H. L. Bishop. 2000. *Virus Taxonomy: Classification and Nomenclature of Viruses : Seventh Report of the International Committee on Taxonomy of Viruses*. Academic Press.

Jenny, Matthew J., Sibel I. Karchner, Diana G. Franks, Bruce R. Woodin, John J. Stegeman, and Mark E. Hahn. 2009. “Distinct Roles of Two Zebrafish AHR Repressors (AHRRa and AHRRb) in Embryonic Development and Regulating the Response to 2,3,7,8-Tetrachlorodibenzo-P-Dioxin.” *Toxicological Sciences: An Official Journal of the Society of Toxicology* 110 (2): 426–41.

Jiao, Shuang, Rui Xu, and Shaojun Du. 2021. “Smyd1 Is Essential for Myosin Expression and Sarcomere Organization in Craniofacial, Extraocular, and Cardiac Muscles.” *Journal of Genetics and Genomics = Yi Chuan Xue Bao* 48 (3): 208–18.

Julliard, Walker, John H. Fechner, and Joshua D. Mezrich. 2014. “The Aryl Hydrocarbon Receptor Meets Immunology: Friend or Foe? A Little of Both.” *Frontiers in Immunology* 5 (October): 458.

Kassambara, A., Mundt, F., 2020 Factoextra: Extract and Visualize the Results of Multivariate Data Analyses. R Package Version 1.0.7. https://CRAN.R-project.org/package=factoextra.

Khursigara, Alexis J., Jacob L. Johansen, and Andrew J. Esbaugh. 2018. “Social Competition in Red Drum (Sciaenops Ocellatus) Is Influenced by Crude Oil Exposure.” *Aquatic Toxicology* 203 (October): 194–201.

Kim, Daehwan, Ben Langmead, and Steven L. Salzberg. 2015. “HISAT: A Fast Spliced Aligner with Low Memory Requirements.” *Nature Methods* 12 (4): 357–60.

Kobayashi, Isao, Jingjing Kobayashi-Sun, Yuto Hirakawa, Madoka Ouchi, Koyuki Yasuda, Hiroyasu Kamei, Shigetomo Fukuhara, and Masaaki Yamaguchi. 2020. “Dual Role of Jam3b in Early Hematopoietic and Vascular Development.” *Development* 147 (1). https://doi.org/10.1242/dev.181040.

Kolde, Raivo, and Jaak Vilo. 2015. “GOsummaries: An R Package for Visual Functional Annotation of Experimental Data.” *F1000Research*. https://doi.org/10.12688/f1000research.6925.1.

Kramer, Frederike, Jens Dernedde, Artur Mezheyeuski, Rudolf Tauber, Patrick Micke, and Kai Kappert. 2020. “Platelet-Derived Growth Factor Receptor β Activation and Regulation in Murine Myelofibrosis.” *Haematologica* 105 (8): 2083–94.

Kujawinski, Elizabeth B., Christopher M. Reddy, Ryan P. Rodgers, J. Cameron Thrash, David L. Valentine, and Helen K. White. 2020. “The First Decade of Scientific Insights from the Deepwater Horizon Oil Release.” Nature Reviews Earth & Environment. https://doi.org/10.1038/s43017-020-0046-x.

Kuleshov, Maxim V., Matthew R. Jones, Andrew D. Rouillard, Nicolas F. Fernandez, Qiaonan Duan, Zichen Wang, Simon Koplev, et al. 2016. “Enrichr: A Comprehensive Gene Set Enrichment Analysis Web Server 2016 Update.” *Nucleic Acids Research* 44 (W1): W90–97.
Kurt C. Stick, Adam Lindquist, and Dayv Lowry. “2012 Washington State Herring Stock Status Report.” Washington Department of Fish & Wildlife. https://wdfw.wa.gov/publications/01628.
Langfelder, Peter, and Steve Horvath. 2008. “WGCNA: An R Package for Weighted Correlation Network Analysis.” *BMC Bioinformatics*. https://doi.org/10.1186/1471-2105-9-559.
Law, Charity W., Yunshun Chen, Wei Shi, and Gordon K. Smyth. 2014. “Voom: Precision Weights Unlock Linear Model Analysis Tools for RNA-Seq Read Counts.” *Genome Biology* 15 (2): R29.
Lazzerini, Pietro Enea, Pier Leopoldo Capecchi, Nabil El-Sherif, Franco Laghi-Pasini, and Mohamed Boutjdir. 2018. “Emerging Arrhythmic Risk of Autoimmune and Inflammatory Cardiac Channelopathies.” *Journal of the American Heart Association* 7 (22): e010595.
Liao, Yang, Gordon K. Smyth, and Wei Shi. 2013. “The Subread Aligner: Fast, Accurate and Scalable Read Mapping by Seed-and-Vote.” *Nucleic Acids Research* 41 (10): e108.
Li, Ting, Wei Wang, Shunyou Gong, Honghong Sun, Huqin Zhang, An-Gang Yang, Youhai H. Chen, and Xinyuan Li. 2018. “Genome-Wide Analysis Reveals TNFAIP8L2 as an Immune Checkpoint Regulator of Inflammation and Metabolism.” *Molecular Immunology* 99 (July): 154–62.
Lohman, Brian K., Jesse N. Weber, and Daniel I. Bolnick. 2016. “Evaluation of TagSeq, a Reliable Low-Cost Alternative for RNAseq.” *Molecular Ecology Resources* 16 (6): 1315–21.
Lubbers, Ellen R., and Peter J. Mohler. 2016. “Roles and Regulation of Protein Phosphatase 2A (PP2A) in the Heart.” *Journal of Molecular and Cellular Cardiology* 101 (December): 127–33.
Lu, Gang, Shuxun Ren, Paavo Korge, Jayoung Choi, Yuan Dong, James Weiss, Carla Koehler, Jau-Nian Chen, and Yibin Wang. 2007. “A Novel Mitochondrial Matrix Serine/threonine Protein Phosphatase Regulates the Mitochondria Permeability Transition Pore and Is Essential for Cellular Survival and Development.” *Genes & Development* 21 (7): 784–96.
Lynch, Michael, and Georgi K. Marinov. 2015. “The Bioenergetic Costs of a Gene.” *Proceedings of the National Academy of Sciences of the United States of America* 112 (51): 15690–95.
Lyon, Angeline M., and John J. G. Tesmer. 2013. “Structural Insights into Phospholipase C-β Function.” *Molecular Pharmacology* 84 (4): 488–500.
Ma, Feiyang, Brie K. Fuqua, Yehudit Hasin, Clara Yukhtman, Chris D. Vulpe, Aldons J. Lusis, and Matteo Pellegrini. 2019. “A Comparison between Whole Transcript and 3’ RNA Sequencing Methods Using Kapa and Lexogen Library Preparation Methods.” *BMC Genomics* 20 (1): 9.
Mager, Edward M., Andrew J. Esbaugh, John D. Stieglitz, Ronald Hoenig, Charlotte Bodinier, John P. Incardona, Nathaniel L. Scholz, Daniel D. Benetti, and Martin Grosell. 2014. “Acute Embryonic or Juvenile Exposure to*Deepwater Horizon*Crude Oil Impairs the Swimming Performance of Mahi-Mahi (*Coryphaena Hippurus*).” *Environmental Science & Technology*. https://doi.org/10.1021/es501628k.
Mandelboum, Shir, Zohar Manber, Orna Elroy-Stein, and Ran Elkon. 2019. “Recurrent Functional Misinterpretation of RNA-Seq Data Caused by Sample-Specific Gene Length Bias.” *PLoS Biology* 17 (11): e3000481.
Marty, Gary. 2008. “Effects of the Exxon Valdez Oil Spill on Pacific Herring in Prince William Sound, Alaska.” *The Toxicology of Fishes*. https://doi.org/10.1201/9780203647295.ch23.
Marty, Gary D., Terrance J. Quinn Ii, Greg Carpenter, Theodore R. Meyers, and Neil H. Willits. 2003. “Role of Disease in Abundance of a Pacific Herring (Clupea Pallasi) Population.” *Canadian Journal of Fisheries and Aquatic Sciences. Journal Canadien Des Sciences Halieutiques et Aquatiques* 60 (10): 1258–65.

Marty, Gary D., Terrance J. Quinn II, Greg Carpenter, Theodore R. Meyers, and Neil H. Willits. 2003. “Role of Disease in Abundance of a Pacific Herring (*Clupea Pallasi*) Population.” *Canadian Journal of Fisheries and Aquatic Sciences*. https://doi.org/10.1139/f03-109.

Marty, G. D., E. F. Freiberg, T. R. Meyers, J. Wilcock, T. B. Farver, and D. E. Hinton. 1998a. “Viral Hemorrhagic Septicemia Virus, Ichthyophonus Hoferi, and Other Causes of Morbidity in Pacific Herring Clupea Pallasi Spawning in Prince William Sound, Alaska, USA.” *Diseases of Aquatic Organisms*. https://doi.org/10.3354/dao032015.

Marty, G. D., E. F. Freiberg, T. R. Meyers, J. Wilcock, T. B. Farver, and D. E. Hinton. 1998b. “Viral Hemorrhagic Septicemia Virus, Ichthyophonus Hoferi, and Other Causes of Morbidity in Pacific Herring Clupea Pallasi Spawning in Prince William Sound, Alaska, USA.” *Diseases of Aquatic Organisms* 32 (1): 15–40.

Mashoof, Sara, and Michael F. Criscitiello. 2016. “Fish Immunoglobulins.” *Biology* 5 (4). https://doi.org/10.3390/biology5040045.

McFadden, David G., Ana C. Barbosa, James A. Richardson, Michael D. Schneider, Deepak Srivastava, and Eric N. Olson. 2005. “The Hand1 and Hand2 Transcription Factors Regulate Expansion of the Embryonic Cardiac Ventricles in a Gene Dosage-Dependent Manner.” Development. https://doi.org/10.1242/dev.01562.

Meador, J. P., J. E. Stein, W. L. Reichert, and U. Varanasi. 1995. “Bioaccumulation of Polycyclic Aromatic Hydrocarbons by Marine Organisms.” *Reviews of Environmental Contamination and Toxicology*. https://doi.org/10.1007/978-1-4612-2542-3_4.

Merks, Anne Margarete, Marie Swinarski, Alexander Matthias Meyer, Nicola Victoria Müller, Ismail Özcan, Stefan Donat, Alexa Burger, et al. 2018. “Planar Cell Polarity Signalling Coordinates Heart Tube Remodelling through Tissue-Scale Polarisation of Actomyosin Activity.” *Nature Communications* 9 (1): 2161.

Meyers, T. R., S. Short, K. Upson, W. N. Batts, J. R. Winton, J. Wilcock, and E. Brown. 1994. “Association of Viral Hemorrhagic Septicemia Virus with Epizootic Hemorrhages of the Skin in Pacific Herring Clupea Harengus Pallasi from Prince William Sound and Kodiak Island, Alaska, USA.” *Diseases of Aquatic Organisms* 19: 27–37.

Mi, Gu, Yanming Di, Sarah Emerson, Jason S. Cumbie, and Jeff H. Chang. 2012. “Length Bias Correction in Gene Ontology Enrichment Analysis Using Logistic Regression.” *PloS One* 7 (10): e46128.

Monosson, Emily. 2012. *Evolution in a Toxic World*. Springer My Copy UK.

Münch, Juliane, Dimitrios Grivas, Álvaro González-Rajal, Rebeca Torregrosa-Carrión, and José Luis de la Pompa. 2017. “Notch Signalling Restricts Inflammation and Expression in the Dynamic Endocardium of the Regenerating Zebrafish Heart.” *Development* 144 (8): 1425–40.

Muradian, Melissa L., Trevor A. Branch, Steven D. Moffitt, and Peter-John F. Hulson. 2017. “Bayesian Stock Assessment of Pacific Herring in Prince William Sound, Alaska.” *PloS One* 12 (2): e0172153.

National Academies of Sciences, Engineering, and Medicine 2022. Oil in the Sea IV: Inputs, Fates, and Effects. Washington, DC: The National Academies Press. https://doi.org/10.17226/26410.

Nebert, Daniel W., Timothy P. Dalton, Allan B. Okey, and Frank J. Gonzalez. 2004. “Role of Aryl Hydrocarbon Receptor-Mediated Induction of the CYP1 Enzymes in Environmental Toxicity and Cancer.” *The Journal of Biological Chemistry* 279 (23): 23847–50.

Nordtug, Trond, Anders Johny Olsen, Dag Altin, Sonnich Meier, Ingrid Overrein, Bjørn Henrik Hansen, and Øistein Johansen. 2011. “Method for Generating Parameterized Ecotoxicity Data of Dispersed Oil for Use in Environmental Modelling.” *Marine Pollution Bulletin* 62 (10): 2106–13.

O’Leary, Nuala A., Mathew W. Wright, J. Rodney Brister, Stacy Ciufo, Diana Haddad, Rich McVeigh, Bhanu Rajput, et al. 2016. “Reference Sequence (RefSeq) Database at NCBI:

Current Status, Taxonomic Expansion, and Functional Annotation." *Nucleic Acids Research* 44 (D1): D733–45.
Orr, H. Allen, and H. Allen Orr. 1998. "Testing Natural Selection vs. Genetic Drift in Phenotypic Evolution Using Quantitative Trait Locus Data." *Genetics*. https://doi.org/10.1093/genetics/149.4.2099.
Oshlack, Alicia, and Matthew J. Wakefield. 2009. "Transcript Length Bias in RNA-Seq Data Confounds Systems Biology." *Biology Direct* 4 (April): 14.
Padang, Ratnasari, Richard D. Bagnall, Tatiana Tsoutsman, Paul G. Bannon, and Christopher Semsarian. 2015. "Comparative Transcriptome Profiling in Human Bicuspid Aortic Valve Disease Using RNA Sequencing." *Physiological Genomics* 47 (3): 75–87.
Panáková, Daniela, Andreas A. Werdich, and Calum A. Macrae. 2010. "Wnt11 Patterns a Myocardial Electrical Gradient through Regulation of the L-Type Ca(2+) Channel." *Nature* 466 (7308): 874–78.
Pater, Emma de, Metamia Ciampricotti, Florian Priller, Justus Veerkamp, Ina Strate, Kelly Smith, Anne Karine Lagendijk, et al. 2012. "Bmp Signaling Exerts Opposite Effects on Cardiac Differentiation." *Circulation Research* 110 (4): 578–87.
Patro, Rob, Geet Duggal, Michael I. Love, Rafael A. Irizarry, and Carl Kingsford. 2017. "Salmon Provides Fast and Bias-Aware Quantification of Transcript Expression." *Nature Methods* 14 (4): 417–19.
Pearson, Walter H., Richard B. Deriso, Ralph A. Elston, Sharon E. Hook, Keith R. Parker, and Jack W. Anderson. 2012. "Hypotheses Concerning the Decline and Poor Recovery of Pacific Herring in Prince William Sound, Alaska." *Reviews in Fish Biology and Fisheries*. https://doi.org/10.1007/s11160-011-9225-7.
Pearson, W. H., R. A. Elston, R. W. Bienert, A. S. Drum, and L. D. Antrim. 1999. "Why Did the Prince William Sound, Alaska, Pacific Herring (*Clupea Pallasi*) Fisheries Collapse in 1993 and 1994? Review of Hypotheses." *Canadian Journal of Fisheries and Aquatic Sciences*. https://doi.org/10.1139/f98-207.
Phillips, D. H. 1983. "Fifty Years of Benzo(a)pyrene." *Nature* 303 (5917): 468–72.
Pogoda, Hans-Martin, and Dirk Meyer. 2002. "Zebrafish Smad7 Is Regulated by Smad3 and BMP Signals." *Developmental Dynamics: An Official Publication of the American Association of Anatomists* 224 (3): 334–49.
Qi, Jialing, Annegret Rittershaus, Rashmi Priya, Shivani Mansingh, Didier Y. R. Stainier, and Christian S. M. Helker. 2022. "Apelin Signaling Dependent Endocardial Protrusions Promote Cardiac Trabeculation in Zebrafish." *eLife* 11 (February). https://doi.org/10.7554/eLife.73231.
Quillet, E., M. Dorson, G. Aubard, and C. Torhy. 2001. "In Vitro Viral Haemorrhagic Septicaemia Virus Replication in Excised Fins of Rainbow Trout: Correlation with Resistance to Waterborne Challenge and Genetic Variation." *Diseases of Aquatic Organisms* 45 (3): 171–82.
Quintana, Francisco J. 2013. "The Aryl Hydrocarbon Receptor: A Molecular Pathway for the Environmental Control of the Immune Response." *Immunology*. https://doi.org/10.1111/imm.12046.
Reid, Noah M., Dina A. Proestou, Bryan W. Clark, Wesley C. Warren, John K. Colbourne, Joseph R. Shaw, Sibel I. Karchner, et al. 2016. "The Genomic Landscape of Rapid Repeated Evolutionary Adaptation to Toxic Pollution in Wild Fish." *Science* 354 (6317): 1305–8.
Reifers, F., E. C. Walsh, S. Léger, D. Y. Stainier, and M. Brand. 2000. "Induction and Differentiation of the Zebrafish Heart Requires Fibroblast Growth Factor 8 (fgf8/acerebellar)." *Development* 127 (2): 225–35.
Reynaud, S., C. Duchiron, and P. Deschaux. 2003. "3-Methylcholanthrene Inhibits Lymphocyte Proliferation and Increases Intracellular Calcium Levels in Common Carp (Cyprinus Carpio

L).” *Aquatic Toxicology* 63 (3): 319–31.
Ritchie, Matthew E., Belinda Phipson, Di Wu, Yifang Hu, Charity W. Law, Wei Shi, and Gordon K. Smyth. 2015. “Limma Powers Differential Expression Analyses for RNA-Sequencing and Microarray Studies.” *Nucleic Acids Research* 43 (7): e47.
Roberts, Steven B., Lorenz Hauser, Lisa W. Seeb, and James E. Seeb. 2012. “Development of Genomic Resources for Pacific Herring through Targeted Transcriptome Pyrosequencing.” *PloS One* 7 (2): e30908.
Robinson, Mark D., Davis J. McCarthy, and Gordon K. Smyth. 2010. “edgeR: A Bioconductor Package for Differential Expression Analysis of Digital Gene Expression Data.” *Bioinformatics* 26 (1): 139–40.
Robinson, Mark D., and Alicia Oshlack. 2010. “A Scaling Normalization Method for Differential Expression Analysis of RNA-Seq Data.” *Genome Biology*. https://doi.org/10.1186/gb-2010-11-3-r25.
Rocke, David M., Luyao Ruan, J. Jared Gossett, Blythe Durbin-Johnson, and Sharon Aviran. 2015. “Controlling False Positive Rates in Methods for Differential Gene Expression Analysis Using RNA-Seq Data.” BioRxiv. https://doi.org/10.1101/018739.
Rottbauer, W., K. Baker, Z. G. Wo, M. A. Mohideen, H. F. Cantiello, and M. C. Fishman. 2001. “Growth and Function of the Embryonic Heart Depend upon the Cardiac-Specific L-Type Calcium Channel alpha1 Subunit.” *Developmental Cell* 1 (2): 265–75.
Rouillard, Andrew D., Gregory W. Gundersen, Nicolas F. Fernandez, Zichen Wang, Caroline D. Monteiro, Michael G. McDermott, and Avi Ma’ayan. 2016. “The Harmonizome: A Collection of Processed Datasets Gathered to Serve and Mine Knowledge about Genes and Proteins.” *Database: The Journal of Biological Databases and Curation* 2016 (July). https://doi.org/10.1093/database/baw100.
Samuel, C. E. 1991. “Antiviral Actions of Interferon. Interferon-Regulated Cellular Proteins and Their Surprisingly Selective Antiviral Activities.” *Virology* 183 (1): 1–11.
Schmidt, J. V., and C. A. Bradfield. 1996. “Ah Receptor Signaling Pathways.” *Annual Review of Cell and Developmental Biology* 12: 55–89.
Schoenebeck, Jeffrey J., and Deborah Yelon. 2007. “Illuminating Cardiac Development: Advances in Imaging Add New Dimensions to the Utility of Zebrafish Genetics.” Seminars in Cell & Developmental Biology. https://doi.org/10.1016/j.semcdb.2006.12.010.
Scott, Camille, Tessa Pierce, Lisa K. Johnson, C. Titus Brown, Chaz Reid, The Gitter Badger, and Luiz Irber. 2019. *Dib-Lab/dammit: Dammit v1.2*. Zenodo. https://doi.org/10.5281/ZENODO.3569831.
Segner, Helmut, Christyn Bailey, Carolina Tafalla, and Jun Bo. 2021. “Immunotoxicity of Xenobiotics in Fish: A Role for the Aryl Hydrocarbon Receptor (AhR)?” *International Journal of Molecular Sciences*. https://doi.org/10.3390/ijms22179460.
Shankar, Thirupura S., Dinesh K. A. Ramadurai, Kira Steinhorst, Salah Sommakia, Rachit Badolia, Aspasia Thodou Krokidi, Dallen Calder, et al. 2021. “Cardiac-Specific Deletion of Voltage Dependent Anion Channel 2 Leads to Dilated Cardiomyopathy by Altering Calcium Homeostasis.” *Nature Communications* 12 (1): 4583.
Shih, Yu-Huan, Yuji Zhang, Yonghe Ding, Christian A. Ross, Hu Li, Timothy M. Olson, and Xiaolei Xu. 2015. “Cardiac Transcriptome and Dilated Cardiomyopathy Genes in Zebrafish.” *Circulation. Cardiovascular Genetics* 8 (2): 261–69.
Siepel, Adam, Gill Bejerano, Jakob S. Pedersen, Angie S. Hinrichs, Minmei Hou, Kate Rosenbloom, Hiram Clawson, et al. 2005. “Evolutionarily Conserved Elements in Vertebrate, Insect, Worm, and Yeast Genomes.” *Genome Research* 15 (8): 1034–50.
Simão, Felipe A., Robert M. Waterhouse, Panagiotis Ioannidis, Evgenia V. Kriventseva, and Evgeny M. Zdobnov. 2015. “BUSCO: Assessing Genome Assembly and Annotation Completeness with Single-Copy Orthologs.” *Bioinformatics*. https://doi.org/10.1093/bioinformatics/btv351.

Sloan, C., D. Brown, R. Pearce, R. Boyer, J. Bolton, D. Burrows, D. Herman, and M. Krahn. 2005. “Determining Aromatic Hydrocarbons and Chlorinated Hydrocarbons in Sediments and Tissues Using Accelerated Solvent Extraction and Gas Chromatography/mass Spectrometry.” *Techniques in Aquatic Toxicology, Volume 2*. https://doi.org/10.1201/9780203501597.ch35.

Soneson, Charlotte, Michael I. Love, and Mark D. Robinson. 2015. “Differential Analyses for RNA-Seq: Transcript-Level Estimates Improve Gene-Level Inferences.” *F1000Research* 4 (December): 1521.

Song, Jun-Young, Ayako Casanova-Nakayama, Anja-Maria Möller, Shin-Ichi Kitamura, Kei Nakayama, and Helmut Segner. 2020. “Aryl Hydrocarbon Receptor Signaling Is Functional in Immune Cells of Rainbow Trout ().” *International Journal of Molecular Sciences* 21 (17). https://doi.org/10.3390/ijms21176323.

Sørensen, Lisbet, Marta S. Silva, Andy M. Booth, and Sonnich Meier. 2016. “Optimization and Comparison of Miniaturized Extraction Techniques for PAHs from Crude Oil Exposed Atlantic Cod and Haddock Eggs.” *Analytical and Bioanalytical Chemistry* 408 (4): 1023–32.

Sørensen, Lisbet, Elin Sørhus, Trond Nordtug, John P. Incardona, Tiffany L. Linbo, Laura Giovanetti, Ørjan Karlsen, and Sonnich Meier. 2017. “Oil Droplet Fouling and Differential Toxicokinetics of Polycyclic Aromatic Hydrocarbons in Embryos of Atlantic Haddock and Cod.” *PloS One* 12 (7): e0180048.

Tandonnet, Sophie, and Tatiana Teixeira Torres. 2017. “Traditional 3’ RNA-Seq in a Non-Model Species.” *Genomics Data* 11 (March): 9–16.

Thai, W., C. Mylrea, I. Meredith, and R. Peverill. 2010. “Percutaneous Transluminal Septal Myocardial Ablation Improves Outflow Tract Obstruction and Symptoms Without Changing Long Axis Function in Patients with Hypertrophic Obstructive Cardiomyopathy.” *Heart, Lung and Circulation*. https://doi.org/10.1016/j.hlc.2010.06.467.

Thorne, Richard E., and Gary L. Thomas. 2008. “Herring and the ‘Exxon Valdez’ Oil Spill: An Investigation into Historical Data Conflicts.” *ICES Journal of Marine Science: Journal Du Conseil* 65 (1): 44–50.

Tiitu, V., and M. Vornanen. 2003. “Ryanodine and Dihydropyridine Receptor Binding in Ventricular Cardiac Muscle of Fish with Different Temperature Preferences.” *Journal of Comparative Physiology. B, Biochemical, Systemic, and Environmental Physiology* 173 (4): 285–91.

Townsley, Brad T., Michael F. Covington, Yasunori Ichihashi, Kristina Zumstein, and Neelima R. Sinha. 2015. “BrAD-Seq: Breath Adapter Directional Sequencing: A Streamlined, Ultra-Simple and Fast Library Preparation Protocol for Strand Specific mRNA Library Construction.” *Frontiers in Plant Science* 6 (May): 366.

Tsedeke, Ayele Taddese, Srinivas Allanki, Alessandra Gentile, Vanesa Jimenez-Amilburu, Seyed Javad Rasouli, Stefan Guenther, Shih-Lei Lai, Didier Y. R. Stainier, and Rubén Marín-Juez. 2021. “Cardiomyocyte Heterogeneity during Zebrafish Development and Regeneration.” Developmental Biology 476 (August): 259–71.

Tu, Shu, and Neil C. Chi. 2012. “Zebrafish Models in Cardiac Development and Congenital Heart Birth Defects.” *Differentiation; Research in Biological Diversity* 84 (1): 4–16.

Veerkamp, Justus, Franziska Rudolph, Zoltan Cseresnyes, Florian Priller, Cécile Otten, Marc Renz, Liliana Schaefer, and Salim Abdelilah-Seyfried. 2013. “Unilateral Dampening of Bmp Activity by Nodal Generates Cardiac Left-Right Asymmetry.” *Developmental Cell* 24 (6): 660–67.

Whitehead, Andrew. 2013. “Interactions between Oil-Spill Pollutants and Natural Stressors Can Compound Ecotoxicological Effects.” *Integrative and Comparative Biology* 53 (4): 635–47.

Whitehead, Andrew, Bryan W. Clark, Noah M. Reid, Mark E. Hahn, and Diane Nacci. 2017. “When Evolution Is the Solution to Pollution: Key Principles, and Lessons from Rapid Repeated Adaptation of Killifish () Populations.” *Evolutionary Applications* 10 (8): 762–83.

Whitehead, Andrew, and Douglas L. Crawford. 2006. “Neutral and Adaptive Variation in Gene Expression.” *Proceedings of the National Academy of Sciences*. https://doi.org/10.1073/pnas.0507648103.

Xiong, Yuguang, Magali Soumillon, Jie Wu, Jens Hansen, Bin Hu, Johan G. C. van Hasselt, Gomathi Jayaraman, et al. 2017. “A Comparison of mRNA Sequencing with Random Primed and 3’-Directed Libraries.” *Scientific Reports* 7 (1): 14626.

Xu, Elvis Genbo, Edward M. Mager, Martin Grosell, Christina Pasparakis, Lela S. Schlenker, John D. Stieglitz, Daniel Benetti, et al. 2016. “Time- and Oil-Dependent Transcriptomic and Physiological Responses to *Deepwater Horizon* Oil in Mahi-Mahi (*Coryphaena Hippurus*) Embryos and Larvae.” *Environmental Science & Technology*. https://doi.org/10.1021/acs.est.6b02205.

Yamada, Taisho, Hiromasa Horimoto, Takeshi Kameyama, Sumio Hayakawa, Hiroaki Yamato, Masayoshi Dazai, Ayato Takada, et al. 2016. “Constitutive Aryl Hydrocarbon Receptor Signaling Constrains Type I Interferon–mediated Antiviral Innate Defense.” *Nature Immunology* 17 (6): 687–94.

Ye, Ding, Huaping Xie, Bo Hu, and Fang Lin. 2015. “Endoderm Convergence Controls Subduction of the Myocardial Precursors during Heart-Tube Formation.” *Development* 142 (17): 2928–40.

Zhou, Zuoqiong, Lan Zheng, Changfa Tang, Zhanglin Chen, Runkang Zhu, Xiyang Peng, Xiushan Wu, and Ping Zhu. 2020. “Identification of Potentially Relevant Genes for Excessive Exercise-Induced Pathological Cardiac Hypertrophy in Zebrafish.” *Frontiers in Physiology* 11 (November): 565307.

Zuberi, Zia, Lutz Birnbaumer, and Andrew Tinker. 2008. “The Role of Inhibitory Heterotrimeric G Proteins in the Control of in Vivo Heart Rate Dynamics.” *American Journal of Physiology. Regulatory, Integrative and Comparative Physiology* 295 (6): R1822–30.