Application of RNA-Seq in Ecotoxicogenomics: Exploring the Effects of Ibuprofen Exposure on Rainbow Trout and *C. elegans*

by

Lorraine Lok-Lun Yu Brown

B.Sc., University of British Columbia, 2007

Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of
Master of Science

in the
Department of Molecular Biology and Biochemistry
Faculty of Science

SIMON FRASER UNIVERSITY

Fall 2013

Approval

Name: **Lorraine Lok-Lun Yu Brown**

Degree: **Master of Science**

Title of Thesis: ***Application of RNA-Seq in Ecotoxicogenomics: Exploring the Effects of Ibuprofen Exposure on Rainbow Trout and* C. elegans**

Examining Committee: **Chair:** Nicholas Harden
Professor

Fiona Brinkman
Senior Supervisor
Professor

William Davidson
Supervisor
Professor

Steven Jones
Supervisor
Professor

Christopher Kennedy
Internal Examiner
Professor
Department of Biological Sciences

Date Defended/Approved: December 16, 2013

Partial Copyright Licence

Ethics Statement

The author, whose name appears on the title page of this work, has obtained, for the research described in this work, either:

a. human research ethics approval from the Simon Fraser University Office of Research Ethics,

or

b. advance approval of the animal care protocol from the University Animal Care Committee of Simon Fraser University;

or has conducted the research

c. as a co-investigator, collaborator or research assistant in a research project approved in advance,

or

d. as a member of a course approved in advance for minimal risk human research, by the Office of Research Ethics.

A copy of the approval letter has been filed at the Theses Office of the University Library at the time of submission of this thesis or project.

The original application for approval and letter of approval are filed with the relevant offices. Inquiries may be directed to those authorities.

Simon Fraser University Library
Burnaby, British Columbia, Canada

update Spring 2010

Abstract

RNA-Seq was applied in this ecotoxicogenomics study to investigate the effects of ibuprofen in two species, rainbow trout (*Oncorhynchus mykiss*), a fish routinely used in ecotoxicology tests, and *Caenorhabditis elegans*, a well-studied nematode with immense genomics information. Exposure to environmentally relevant levels of ibuprofen resulted in gene expression changes relating to stress, prostaglandin synthesis, reproduction and development in both species. In fish, we observed sex-dependent differences in vitellogenin and prostaglandin synthase gene expression, highlighting the importance of genetic sex determination of juvenile fish used in bioassays. In worms, we saw a decrease in progeny production count. Our results suggest that ibuprofen may have negative impacts on reproduction in both species but requires further investigation. This study demonstrated the benefits and challenges of using RNA-Seq in ecotoxicology and the value of studying diverse species, including traditional models and more microscopic organisms, to better understand toxicant impact on an entire ecosystem.

Keywords: RNA-Seq; rainbow trout; *C. elegans*; ibuprofen; toxicogenomics

To my family

Acknowledgements

First and foremost, I would like to thank my supervisor, Dr. Fiona Brinkman, for taking a chance on me and for being willing to branch out to a field (environmental toxicology) that was somewhat new to her when I first started. I am truly grateful for her support, guidance, and patience. Without her, all of this would not be possible.

Thank you to Graham van Aggelen for always encouraging me to pursue my goals. Thank you to my committee members, Dr. Willie Davidson and Dr. Steven Jones, for providing constructive feedback and guidance. Thank you to Dr. Heather Osachoff, for being a positive role model for me and for inspiring me to pursue a graduate degree.

Thank you to the Brinkman Lab staff members and students: Bhav Dhillon, Amir Foroushani, Erin Gill, Emma Griffiths, Matthew Laird, Raymond Lo, David Lynn, Mike Peabody, Thea Van Rossum, Geoff Winsor, and Matthew Whiteside. Thank you to the staff at Environment Canada: Norm Berke, Joy Bruno, Craig Buday, Michelle Linssen-Sauvé, Georgina Lush, Grant Schroeder, and Rachel Skirrow. Also thank you to all the co-op students and volunteers who have helped me with this project: Ishai Kones, Richard Chea, and Melanie Pylatuk.

I am grateful for the funding provided by SFU via Graduate Student Fellowships.

Finally, thank you to Aron, for his endless support and encouragement; my parents, Kin Kong and Wai Fan, and my sister, Rosa, for their unconditional love and support.

Table of Contents

List of Tables

List of Figures

List of Acronyms

BLAST	basic local alignment search tool
bp	base pairs
cDNA	complementary DNA
COX	cyclooxygenase
DEG	differentially expressed gene
DNA	deoxyribonucleic acid
ED	endocrine disruption
EDC	endocrine disrupting compound
FDR	false discovery rate
FPKM	Fragments Per Kilobase of transcript per Million mapped reads
gDNA	genomic DNA
GO	Gene Ontology
KEGG	Kyoto Encyclopedia of Genes and Genomes
IB	ibuprofen
mRNA	messenger RNA
NCBI	National Center for Biotechnology Information
NGM	nematode growth media
NGS	next-generation sequencing
NSAID	non-steroidal anti-inflammatory drug
PCR	polymerase chain reaction
PESC	Pacific Environmental Science Centre
PG	prostaglandin
QA/QC	quality assurance/quality control
qPCR	quantitative PCR
RNA	ribonucleic acid
SFU	Simon Fraser University
VTG	vitellogenin

Chapter 1.

Introduction

Introduction to the Research Problem and Objectives

RNA-Seq has recently emerged as a potent, effective, and reliable tool for transcriptomics research. Initially, it was mainly used in research with well-studied species such as human, mouse, and *Caenorhabditis elegans*; however, in recent years, RNA-Seq has rapidly gained popularity in its application on other organisms. Environmental toxicology, or more specifically, ecotoxicogenomics, is a field that uses a wide variety of species to study the environmental impact of chemicals and pollutants. Transcriptomics is commonly applied in ecotoxicogenomics to examine gene expression patterns caused by toxicant exposure and has, in the past, often relied on hybridization-based and probe-based methods (such as microarrays and quantitative polymerase chain reaction (qPCR)) to quantify gene expression. The potential use of RNA-Seq in ecotoxicogenomics is promising and a need to explore the application of Next Generation Sequencing (NGS) technology in this field has been recently reported (Mehinto *et al*. 2012, Jayapal 2012).

To demonstrate the application of RNA-Seq in ecotoxicogenomics, this thesis will focus on studying the molecular-level effects of ibuprofen (IB), a widely used drug that is readily detected in environmental aquatic systems and known to impact various aquatic species. Research to date has shown that IB can affect several biologically significant systems including immunity, reproduction, growth, heat shock, and stress responses. Whereas most of the previous research studies were focused on examining various specific gene markers indicative of endocrine disruption, this thesis will provide a wider view of the effects of IB exposure through the use of RNA-Seq to examine global gene expression responses.

The research aims for this project are: 1) To demonstrate the application of RNA-Seq in ecotoxicogenomics and 2) To investigate the gene-level effects of IB exposure in two species: rainbow trout (*Oncorhynchus mykiss*), a fish routinely used in ecotoxicology tests, and *Caenorhabditis elegans*, a well-studied worm with vast amounts of available genomic information.

Next Generation Sequencing: RNA-Seq

Overview

With the emergence of next generation sequencing (NGS) technology, mass sequencing of nucleic acids (DNA and RNA) is now feasible, affordable, and accessible by most standard laboratories. The first NGS machines were introduced in 2004 and 2006 by Roche (originally 454, Genome Sequencer FLX) and Illumina (originally Solexa) respectively and were followed by Applied Biosystems' SOLiD and Life Technologies' Ion Torrent™, both released in 2010 (Costa *et al*. 2010). Each platform's approach varies considerably, but there are some similarities in their underlying principles. 454-Pyrposequencing, Illumina sequencing, and Ion Torrent technology are based on *sequencing by synthesis*, while SOLiD is based on *oligonucleotide ligation*. The basis of *sequencing by synthesis* is to identify the nucleotides of target DNA or RNA strands during the synthesis of complementary strands by using reporter molecules such as fluorescent dyes (Illumina), hydrogen ions (Ion Torrent), or pyrophosphates (454). These sequencing reactions are conducted in parallel and can generate huge amounts of data in a relatively short time (as compared to the Sanger method). The process of high-throughput sequencing of messenger RNA (mRNA) is known as RNA-Seq and consists of three main stages: template preparation, sequencing and imaging, and data analysis (Costa *et al*. 2010, Metzker 2010).

Template preparation

In template preparation, ribosomal RNA molecules are removed from total RNA and the remaining mRNA population is reverse-transcribed into complementary DNA (cDNA). These cDNA strands are then fragmented and ligated to adapters at one or both

ends. Finally, each strand is amplified and sequenced in a massively parallel manner to generate short sequence reads from either one end (single-end sequencing) or both ends (paired-end sequencing) (Wang *et al*. 2009, Costa *et al*. 2010).

Sequencing and imaging (Illumina example)

The main differences amongst the various NGS platforms are found in the sequencing and imaging step. For Illumina, the platform used in this project, sequencing is performed in cycles where modified nucleotides are added to the sequencing reaction one base at a time. Each type of modified base (A,G,T,C) is fluorescently labelled with different dyes and has a reversible terminator, which prevents more than one nucleotide from being added at a time. At the beginning of a cycle, the template strands are flooded with modified bases. When a fluorescent base, complementary to the template strand, is incorporated by DNA polymerase, a fluorescent signal is emitted and captured by the sequencer. Thus, an image is generated at every cycle, capturing all the fluorescent signals across the entire library of template strands. This is known as parallel sequencing because the entire cDNA library is being sequenced simultaneously in parallel reactions during each cycle of the run. At the end of each cycle, cleavage and wash steps occur to remove the terminator molecules and any unincorporated bases. The cycle is then repeated and the target cDNA sequences are gradually revealed through many parallel synthesis reactions.

Finally, the sequencing instrument analyzes the set of images and, using base-calling algorithms, outputs millions of sequencing reads and their associated quality values (Costa *et al*. 2010, Metzker 2010). To increase the number of samples per run, and consequently decrease the cost per sample, sequencing reactions can be run in multiplex format, where multiple samples are run in the same lane with each sample library labelled with a unique index that acts as a barcode. This can save money and time but will reduce the number of reads generated for each sample.

Data analysis

Typical RNA-Seq runs can generate many gigabytes of data comprising of 10 to over 400 million reads of 30 to 400 base pairs (bp) in length, depending on the

technology used (Wang *et al*. 2009). For gene expression studies, the data analysis process can be summarized as five main steps: data preprocessing, read mapping, annotation, normalization and statistical analysis.

Data preprocessing usually involves adapter trimming and some filtering for quality assurance/quality control (QA/QC) measures. Recall that adapters are initially appended to target DNA or cDNA strands prior to sequencing reactions. These adapters are included in the sequence data output and are trimmed during preprocessing. If sequencing was conducted in multiplex format, the reads from each sample library are separated and binned by their index or barcode at this time. The sequencing instrument also outputs quality scores for each base of each read. Low quality reads can be filtered out and removed from subsequent analysis.

After initial quality checks and filtering, the next step is read mapping, which involves mapping high quality sequencing reads to an existing genome or transcriptome. For species without an available reference genome, *de novo* assembly is typically conducted first (Garber *et al*. 2011). In *de novo* assembly, sequencing reads are pieced together using overlapping regions to generate contiguous sequences (contigs) which represent transcripts or genes. Original sequencing reads are then mapped back to the newly assembled transcripts to determine read count values. A simplified overview of a genome-independent analysis process is shown in Figure 1.1, as represented by the dotted arrows. For species with a reference genome or a closely related species' genome, the output reads can either be aligned directly to the reference sequences or the reads can first undergo *de novo* assembly to create transcripts, which are then aligned to the reference. Then, similar to the genome-independent process, gene expression levels are derived from read count values. This genome-dependent process is represented by the solid arrows in Figure 1.1.

For organisms that have an available reference genome or a closely related reference species, annotation, and sometimes functional information, can be obtained from the reference sequences and transcripts. For those without a usable reference, depending on the application, transcripts can be annotated either following *de novo* assembly or after statistical gene expression analysis. If the goal is to characterize novel transcripts or the whole transcriptome, then annotation typically follows *de novo*

assembly. In contrast, for gene expression studies, annotation can be done after statistical analysis on the resulting gene list, to make more efficient use of computational resources. For transcripts derived from *de novo* assembly, annotation can be done by conducting a BLAST (Basic Local Alignment Search Tool) search against a database of known gene sequences, such as the National Center for Biotechnology Information (NCBI) non-redundant nucleotide and protein databases. Based on sequence similarity, the identity and function of the unknown transcripts can be drawn from homologs of related species. This approach works well for evolutionarily conserved genes but not for highly divergent or functionally unique genes for the organism of interest.

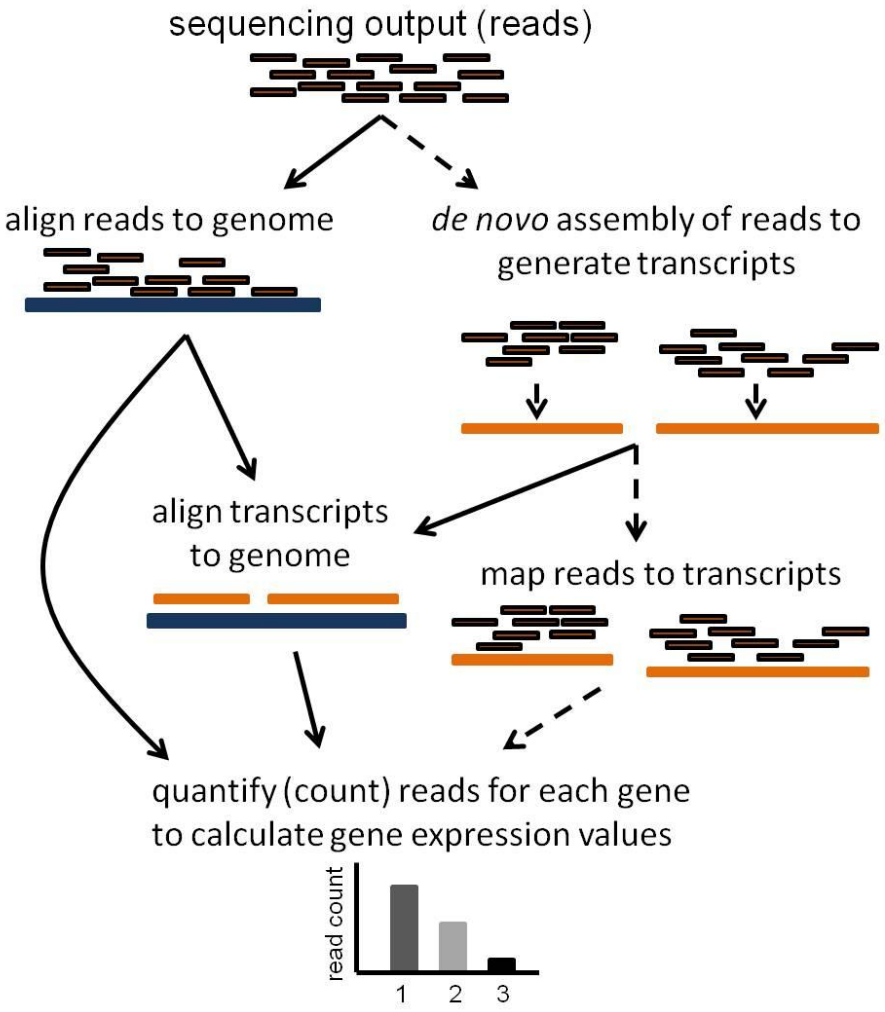

Figure 1.1. Simplified overview of RNA-Seq data analysis steps. Solid arrows represent genome-dependent processes while dotted arrows represent genome-independent processes.

There are many factors that can create biases when working with sequencing data. For example, the differences in the number of reads produced for each sample, the technical variations between samples generated during library preparation, and the differences in gene length and expression abundance can all contribute to analysis bias (Rapaport *et al*. 2013). As such, the raw read count values obtained for each gene during the read mapping step must be normalized before comparisons can be made between sample groups. A common approach used for paired end sequencing is to calculate the Fragments Per Kilobase of transcript per Million mapped reads (FPKM) value (Trapnell *et al*. 2010):

$$FPKM = \frac{\#\ of\ fragments}{length\ of\ transcript\ (kb)} \times \frac{10^6}{total\ number\ of\ mapped\ reads}$$

This calculation normalizes the raw reads data to account for differences in transcript length and the total number of mapped reads for each sample. In RNA-Seq, the relative expression level of a transcript is considered to be proportional to the number of mapped reads to that transcript. The representation of each transcript within the total number of reads is, therefore, dependent on the expression levels of all other transcripts. Highly expressed genes will yield large portions of the sequenced reads and can skew the apparent read counts of lowly expressed genes (Rapaport *et al*. 2013). To correct for this bias, some normalization methods compute a scaling factor on the assumption that most genes are not differentially expressed (DE). For example, DESeq, the method used in this project, expands on the FPKM metric by calculating a scaling factor for each sample (Anders and Huber 2010). It first computes, for each gene, the ratio of its read count in a given sample over its geometric mean read count in all samples and then takes the median of all the ratios for all genes to derive the scaling factor for that sample. There are many other methods for normalization but their discussion is beyond the scope of this work.

Lastly, statistical analysis is conducted on the normalized gene expression data to determine significant differences amongst the sample groups. Rapaport *et al.*(2013) evaluated an array of popular RNA-Seq analysis methods and software programs, including DESeq, edgeR, Cuffdiff, and limma, on key features such as accuracy and sensitivity of DE genes detection. Although no single method emerged as the best, they

found that DESeq was slightly superior in detecting differentially expressed genes with improved specificity, sensitivity and false positive errors. DESeq uses a variation of the Fisher exact test based on negative binomial distribution (Anders and Huber 2010). This method accounts for the commonly observed “overdispersion” problem where the variances of gene expression across many biological replicates are larger than their mean expression values. Because RNA-Seq is used to examine the significance of thousands of genes at once, multiple testing correction (i.e. Benjamini-Hochberg) should be conducted on statistical values to minimize false positive errors. At the end of this process, a list of genes that have statistically altered expression levels between the sample groups is obtained and other downstream analyses (i.e. pathway overrepresentation analysis) can be conducted.

Benefits and challenges of RNA-Seq

RNA-Seq has many advantages over existing methods such as microarray and qPCR. Unlike these hybridization-based and probe-based approaches, RNA-Seq does not require prior sequence knowledge. This allows for the discovery, characterization, and quantification of new genes or transcripts, which is helpful for organisms without much pre-existing genome and/or sequence information. Not only can RNA-Seq be used to generate gene expression data, it can also reveal other transcriptomic properties such as alternative splicing and differential isoform abundance. Using longer read lengths, RNA-Seq can reveal the connectivity across multiple exons, which is useful for studying complex transcriptomes (Wang *et al*. 2009). Moreover, with single-base resolution sensitivity and power, RNA-Seq can detect sequence variations such as single nucleotide polymorphisms (SNPs) within the transcriptome.

RNA-Seq is superior in some technical aspects as well. Like microarrays, RNA-Seq can produce massive amounts of data for many genes simultaneously, but unlike the former, RNA-Seq has been shown to generate much lower background noise and a greater dynamic range of detection (Costa *et al*. 2010). This increased sensitivity allows for the identification of low expressing genes, which is useful for transcriptome characterization. RNA-Seq experiments also require less RNA input as compared to microarrays. This is advantageous for working with small organisms or in cases where there is a limited amount of sample tissue available.

For gene expression studies, there are a few factors contributing to analysis bias that should be considered. Gene length has been shown to inherently bias the results of transcript quantification. Longer genes will produce more reads compared to shorter genes, even if they have similar expression levels. More reads means larger sampling size, which in turn leads to greater statistical power for detecting longer transcripts (Oshlack and Wakefield 2009). Despite normalization efforts, shorter transcripts will always be at a statistical disadvantage. This is an innate property of RNA-Seq and any method that involves the sequencing of short fragments generated from full length transcripts (Oshlack and Wakefield 2009).

Another source of bias comes from cDNA fragmentation. Due to current NGS technological limitations, it is not feasible to sequence the entire length of a gene as a whole and so cDNA (gene) must be fragmented into shorter segments before sequencing. In principle, fragmentation and sequencing processes are supposed to be uniformly distributed and random across all transcripts. However, in reality, there are both positional and sequence-specific biases, for which the likelihood of a fragment being selected for sequencing is affected by 1) its location in the original transcript (positional) and 2) the nucleotide sequence at the beginning or end of the fragment (sequence-specific) (Roberts *et al*. 2011). Because gene expression levels are inferred from fragment abundances (read counts), these biases should be taken into consideration during analysis.

The massive amount of data generated by RNA-Seq is both a benefit and a challenge. The gigabytes to terabytes of data from RNA-Seq experiments are in digital format, which means that the raw data can be re-analyzed indefinitely. As computational algorithms, software programs, and analysis tools continue to mature and advance, the old data can be re-examined in new ways to generate new insights. However, working with large datasets poses several bioinformatics challenges, including the need for efficient methods to store, retrieve and process large amounts of data (Wang *et al*. 2009). As sequencing costs continue to decrease, the amount of sequencing data generated can outgrow the speed at which the necessary computational infrastructures are developed. This trend has started to shift the bottleneck from genome/transcriptome data generation to challenges in data management, computational analysis, and biological interpretation (Costa *et al*. 2010).

There are also unique challenges that arise from working with organisms without a reference genome. Many existing software programs and standard data analysis pipelines are designed for use with available reference genomes and there are generally more resources and tools available for these well-studied species. Although there are methods and software packages developed for *de novo* transcript assembly, tools for downstream processes, including read mapping, normalization, and statistical analysis, are mainly optimized for use with a reference genome. Some of the specific challenges faced in this study include generating a unique set of transcripts from *de novo* assembly that is suitable for read mapping, annotating the assembled transcripts, and analyzing for functional pathway overrepresentation. These challenges will be discussed in Chapter 2.

Recent applications of RNA-Seq in environmental toxicogenomics

Environmental toxicogenomics, or ecotoxicogenomics, is a field that integrates principles in toxicology, molecular biology and bioinformatics to study molecular-level impacts of contaminants on environmental health (Nuwaysir *et al*. 1999, Ankley *et al*. 2006). Ecotoxicogenomics can be applied to extend traditional toxicology endpoints such as mortality, behaviour, reproduction, and growth, to provide genomic-level data, ranging from transcriptomics (i.e.: gene expression) to proteomics (i.e.: protein expression and function) and to metabolomics (i.e.: metabolite levels). The sensitivity of this approach allows scientists to study sub-lethal effects of both chronic (long-term) and acute (transient) exposure to pollutants, with aims to detect early indicators of toxicity. The data generated through toxicogenomics research has great potential to be used by the government and risk assessment agencies to set environmental regulatory standards and guidelines (Ankley *et al*. 2006). The knowledge gained from studying molecular level effects can improve our understanding of environmental pollutants' mechanism(s) of action on wildlife. Furthermore, these molecular endpoints (such as gene expression, protein levels, etc.) can potentially become biomarkers that can be used as early indicators of deleterious impact.

RNA-Seq is a relatively novel tool for environmental toxicogenomics research, as gene expression studies in the past often relied on hybridization-based and probe-based methods such as microarrays and qPCR. In the last two years, there has been an

increase in publications that use RNA-Seq to investigate gene expression responses to environmental toxicant exposure in organisms without a reference genome. The organisms studied range from larger, complex species such as fish (Garcia *et al*. 2012, Huang *et al*. 2012, Wiseman *et al*. 2013), to bivalves (Wang *et al*. 2012, Chen *et al*. 2013), to genetically simpler species such as algae (Domingos *et al*. 2013). With the exception of green algae, all other species used in these studies have no reference genome available and thus required genome-independent approaches for data analysis. Many of the studies were conducted on the Illumina sequencing platform, but various different *de novo* assemblers and data analysis programs were used downstream. A summary of the species, toxicant/stressor, platform, and software used is shown in Table 1.1.

Table 1.1. Summary of recent RNA-Seq studies using species lacking a reference genome.

Species	Toxicant or Stressor	Sequencing Platform	Software	Reference
Fundulus grandis **(salt marsh minnow)**	Oil	Illumina	Velvet, Oases, Bowtie, DESeq	Garcia *et al*. 2012
Pimephales promelas **(fathead minnow)**	Oil sands process water	Illumina	CLC Genomics	Wiseman *et al*. 2013
Oryzias melastigma	Perfluorooctane sulfonate	Illumina	Trinity	Huang *et al*. 2012
Chlamydomonas reinhardtii **(green algae)**	Metal nanoparticles	ABI SOLiD	MAQ, DESeq	Domingos *et al*. 2013
Villosa lienosa **(mussel)**	Heat stress	Illumina	Trans-ABySS	Wang *et al*. 2012
Corbicula fluminea **(clam)**	n/a	Illumina	Velvet, Oases	Chen *et al*. 2013
O. mykiss **(rainbow trout)**	n/a	454	SeqMan NGen	Salem *et al*. 2010
O. mykiss **(rainbow trout)**	Temp., salinity, DO, handling	454	Newbler, MIRA3	Sánchez *et al*. 2011
O. mykiss **(rainbow trout)**	Swimming	Illumina	CLC Genomics	Palstra *et al*. 2013

Although RNA-Seq has been used in ecotoxicogenomics studies for several fish species (i.e. minnow, Oryzias, zebrafish), there have been far fewer studies that use RNA-Seq on rainbow trout, a species routinely used in toxicology testing. Salem *et al.* 2010 was one of the first publications characterizing the transcriptome of rainbow trout using NGS technology. They used a combination of Sanger and 454-pyrosequencing datasets to produce high quality draft assemblies of the rainbow trout transcriptome. Although sequencing was performed using only one individual fish from a clonal line, they were able to generate over 1.2 million good quality expressed sequence tag (EST) sequences. The fish was not exposed to any toxicants or stressors in that study, so the draft transcriptome generated may not contain transcripts that are responsive to toxicant exposure. In 2011, Sánchez *et al.* published a rainbow trout transcriptome that included multiple tissues responding to multiple stressors. In that study, 454-pyrosequencing was used to sequence a pooled RNA sample from gill, brain, liver, spleen, kidney and muscle tissues of non-stressed (control) and stressed fish. Fish was stressed by perturbations in temperature (high and low), salinity, dissolved oxygen (DO) levels and handling/crowding during the experiment. *De novo* assembly was conducted and resulted in over 3.1 million ESTs, many of which were new. Both of these studies used NGS technology, but not in the context of studying gene expression changes in response to various exposures or conditions. In a recently published study (2013), Palstra *et al.* conducted RNA-Seq on female rainbow trout and reported gene expression differences between the red and white muscle of fish during various swimming exercises. That study was the first application of RNA-Seq on the skeletal muscle in rainbow trout and one of the first published studies to use RNA-Seq to investigate gene expression differences in trout.

The three aforementioned studies are amongst the first RNA-Seq studies conducted on rainbow trout. Although this species is commonly and routinely used in environmental toxicology testing, to date there is still a lack of studies that take advantage of the potency of RNA-Seq to investigate gene-level responses to toxicant exposure. The research conducted in this thesis aims to fill that gap and to demonstrate the efficacy of using RNA-Seq on rainbow trout in ecotoxicogenomics research.

Ibuprofen: A Non-Steroidal Anti-Inflammatory Drug

Overview of NSAIDs and their mode of action

Non-steroidal anti-inflammatory drugs (NSAIDs) are a structurally diverse group of analgesic, antipyretic, and anti-inflammatory agents that are commonly used to treat fever and pain. In North America, they are readily available over the counter and the popular forms of this class of drugs include ibuprofen (IB), aspirin, and naproxen. The activity of NSAIDs stems from inhibiting cyclooxygenase (COX) enzymes and subsequent prostaglandin (PG) synthesis (Vane and Botting 1998). The first, and still the most used, forms of NSAIDs are non-selective COX-inhibitors that inhibit both isoforms of the enzyme (COX-1 and COX-2). Prostaglandin biosynthesis begins with arachidonic acid (AA), which are unsaturated 20-carbon fatty acids that are embedded in cell membranes. Upon stimulation by a wide variety of triggers, AA are released into the cytoplasm and interact with COX enzymes, whereby AA are converted into prostaglandin G_2 (PGG_2). Then PPG_2 are reduced by peroxidase to become prostaglandin H_2, which are further converted into various biologically important, functionally active prostaglandins: PGD_2, PGE_2, and others (Rao and Knaus 2008). Because PGs, and accordingly COX enzymes, have important roles in maintaining gastric mucosa, recurrent use of NSAIDs has been shown to cause adverse gastrointestinal side-effects such as gastric bleeding and ulcers (Castellsague *et al*. 2012). In additional to gastric protection, PGs also act as essential physiological mediators in inflammation, pain, cardiovascular disease, and reproduction (Rao and Knaus 2008, Fent *et al*. 2006). Due to their impact on multiple physiological systems as potential endocrine disruptors and the high levels of global consumption and production, NSAIDs are an environmental concern, particularly since they are routinely found in sewage and other wastewater effluents all over the world (Stasinakis *et al*. 2013).

Presence of ibuprofen in the environment

NSAIDs are the most abundant class of pharmaceuticals found in the environment and IB is the most used and produced NSAID globally (Santos *et al*. 2010). It's listed as an *essential drug* by the World Health Organization and manufactured at estimated levels of several kilotons annually (WHO 2010, Han *et al*. 2010). Its presence

in waste water effluents has been well-documented and reported at levels of up to 6.7 ug/L in Canada, 8.2 ug/L in Europe and as high as 2.7 ug/L in natural surface water (Metcalfe *et al*. 2003, Heberer 2002). Approximately 70-80% of the therapeutic dose (600-1200 mg/day) is excreted from the human body as the parent compound or as metabolites (Buser *et al*. 1999). The three major metabolites of IB (hydroxy-IB, carboxy-IB, and carboxy hydratropic acid) are often found at even higher levels than the parent compound and have been detected in both natural aquatic environments and soils/sediments (Ferrando-Climent *et al*. 2012). Most research to date have focused on the activity and toxicity of the parent compound and not the metabolites. Although the reported half-life of IB varies considerably (ranging from 15 hours up to 32 days ,depending on water parameters and contents), it is continuously discharged into the environment and as such, aquatic organisms living near discharge sites are chronically exposed over their entire lifespan (Tixier *et al*. 2003, Lin and Reinhard 2005).

Several fish species have been shown to uptake water-borne IB readily. In a field study conducted by Brown *et al.*(2007), caged juvenile rainbow trout were exposed to nearby sewage discharge containing IB. At the test site, the IB level was detected to be 4.5 ng/L and IB was present in fish plasma at 4680 µg/L, which is highest amongst other NSAIDs examined including ketoprofen, diclofenac, and naproxen. In another study, under laboratory conditions, juvenile rainbow trout was exposed to a continuous flow of IB at 920 µg/L for 96 hours. After the initial 24 hours of exposure, the plasma concentration of IB increased from 7 µg/mL to 10.6 µg/mL (Huggett *et al.* 2004). Bioconcentration of IB was also observed in mosquito fish, where IB was detectable within the fish after only 5 hours of exposure (Wang and Gardinali 2013). The accumulation of IB in liver tissue was studied by (Nallani *et al.* 2011), where they found that IB was present in fathead minnow liver at as early as 24 hours of exposure and in catfish liver after 7 days. These findings show that IB is readily uptake by fish and may have negative impacts on continuously exposed organisms.

Known effects of ibuprofen on non-target organisms

Ibuprofen is a non-selective COX inhibitor, acting on both functionally important isoforms, COX-1 and COX-2, which are responsible for PG production (Cha *et al*. 2006). Due to its effects on PG synthesis, chronic exposure to ibuprofen has been shown to

cause endocrine disruption in vertebrates and invertebrates. In *Daphnia magna*, IB exposure at mg/L levels diminished mobility, reproduction and population growth (Han *et al*. 2010, Heckmann *et al*. 2007); while in bivalves, IB exposure at ug/L levels caused oxidative imbalance, hemocyte genotoxicity, and immunosuppression (Gonzalez-Rey and Bebianno 2012, Contardo-Jara *et al*. 2011, Parolini *et al*. 2011, Matozzo *et al*. 2012). Alterations of hematological parameters and hemo-genotoxicity were also reported in Indian carp and tilapia (Saravanan *et al*. 2012, Ragugnetti *et al*. 2011). Other effects of IB exposure at environmentally relevant levels (as low as 0.1 ug/L) were observed in Japanese medaka (decreased spawning and delayed egg hatching) and in rainbow trout (disruption of the heat shock response, corticosteroidogenesis, and seawater adaptation) (Han *et al*. 2010, Flippin *et al*. 2007, Gravel and Vijayan 2007, Gravel *et al*. 2009, van Anholt *et al*. 2003). Research to date has shown that IB can impact several biologically significant endocrine systems including immunity, reproduction, growth, heat shock, and stress response. Given that most previous research had focused on examining specific gene markers indicative of endocrine disruption, this thesis will provide a wider view of the effects of IB exposure through the use of RNA-Seq to examine global gene expression responses.

The endocrine system and endocrine disruption

The endocrine system consists of glands that produce and secrete various hormones which communicate with tissues, organs, and whole biological systems. Through the complex interactions of hormone cascades, the endocrine system influences and regulates essential biological functions such as metabolism, growth, development, immune system, and reproduction. Hormones typically work together in pathways, also known as an axis. Some well characterized endocrine pathways include: the hypothalamic-pituitary-adrenal (HPA) axis (stress response, immune system, digestion), the hypothalamic-pituitary-gonadal (HPG) axis (reproduction), the hypothalamic-pituitary-thyroid (HPT) axis (metabolism, growth), and the renin-angiotensin system (RAS) (blood pressure, fluid balance) (Hiller-Sturmhofel and Bartke1998).

Endocrine disruption (ED) occurs when the hormone signalling system is disturbed, resulting in homeostasis imbalance. Exogenous substances that interfere with

the synthesis, release, and function of hormones are known as endocrine disrupting compounds (EDCs). These agents can mimic or antagonize endogenous hormones thereby altering hormone expression, synthesis, elimination and receptor interaction. Mimics can fool target receptors by increasing the perceived amount of natural hormones thus activating a false response, while antagonists can block target receptors and prevent natural hormones from binding (Frye *et al*. 2012).

Not only are EDCs commonly present in pharmaceuticals and pesticides, but they are also found in many household products such as food and beverage containers (plastics, cans), detergents, perfumes, and cosmetics (Schug *et al*. 2011). Increased global consumption and production of EDCs, as well as increased industrial by-products and pollution, are causes of worldwide concern, particularly since ED has been linked to impacts on reproduction, sex differentiation, metabolic disorders, growth and developmental dysfunctions, and other harmful effects on both human and environmental health (Colborn *et al*. 1993, Tyler *et al*. 1998, WHO 2002, Jobling *et al*. 2009, Casals-Casas and Desvergne 2011). From an environmental perspective, the presence of EDCs in the natural aquatic environment is especially concerning. Because of the continuous discharge of sewage, industrial and agricultural waste laden with EDCs, receiving waters are chronically dosed with these chemicals, which, if persistent, can travel long distances and threaten aquatic ecosystems.

Endocrine disrupting compounds are often structurally similar to natural hormones and thus can have an impact even at low exposure concentrations. EDCs are typically categorized based on their activity and effects: estrogenic, anti-estrogenic, androgenic, anti-androgenic, or thyroidogenic (Schug *et al*. 2011). Much environmental toxicology research have been focused on estrogen, androgen, and thyroid hormone disruption because the metabolism and action of these hormones are highly conserved among vertebrates (Scholz *et al*. 2013). Moreover, these hormones regulate key aspects of sexual reproduction, growth, metamorphosis, survival, and population dynamics, all of which are crucial elements in ecosystem health and sustainability (Scholz and Mayer 2008).

In recent years, toxicogenomics methods have emerged as viable alternative assays for EDC testing, due to the ability to examine the sub-lethal impacts of low

concentration, chronic EDC exposure at the molecular (gene) level. Hormone-responsive genes are used as predictive biomarkers, whose altered expression levels are indicative of early signs of endocrine disruption. For example, genes (and proteins) such as vitellogenin and thyroid hormone receptor have been used extensively as biomarkers for estrogenic and thyroidogenic impact respectively, due to their reactivity to hormone mimics (Scholz *et al*. 2013).

Vitellogenesis in endocrine disruption

Vitellogenin (VTG) is the egg yolk precursor protein produced by the females of oviparous (egg-laying) organisms, both vertebrates and invertebrates. It is synthesized in the liver and transported via the bloodstream to the ovaries where it is taken up by growing oocytes. Vitellogenesis is the hormone-controlled process whereby VTG is produced and is part of the expanded hypothalamus-pituitary-gonadal-liver (HPGL) axis (Sumpter and Jobling 1995). In fish, the expression of VTG gene and subsequent protein synthesis is regulated mainly by estrogen (17β-estradiol) and its interaction with estrogen receptors within hepatocytes (liver cells). Estrogen or estrogen-mimics bind to estrogen receptors and trigger the transcription of VTG and vitelline envelop proteins (VEPs). In female salmonids, during peak reproductive season, plasma VTG concentrations can increase by up to a million-fold (Sumpter and Jobling 1995). Although minimal basal levels of VTG can be detected in mature male fish and all sexually immature fish, the expression of VTG can be induced by exogenous estrogen or estrogen-mimic exposure (Sumpter and Jobling 1995). The highly responsive and inducible nature of VTG is the basis of its effectiveness as an endocrine disruption biomarker.

Background Information of Test Species

Rainbow trout (Oncorhynchus mykiss)

Rainbow trout is the world's standard cool-water fish for freshwater pollution research in aquatic toxicology and is routinely used by global regulatory agencies including Environment Canada and the US Environmental Protection Agency. For more

than two decades, they have been used extensively in Canada for evaluating the acute lethal effects of exposure to environmental chemicals or effluents. Due to their sensitivity, underyearling rainbow trout are one of the key test organisms used to establish the "deleteriousness" of a discharged substance, as defined by the *Canadian Fisheries Act*, a violation of which is a federal offence (R.S.C. 1985). A substantial amount of toxicology data for rainbow trout has been published, including many recent molecular-level studies using toxicogenomics tools such as microarrays and qPCR. Underyearling rainbow trout have been shown to be vulnerable and responsive to endocrine disruption and thus make an effective and fitting model species for this project (Naderi *et al*. 2013, Osachoff *et al*. 2013).

Caenorhabditis elegans

C. elegans is a well-characterized transparent nematode that resides in both soil and aquatic environments. They are approximately 1.5 mm long (adults), grow rapidly (doubling time of half a day), have a short generation time (3-4 days) and can be cultured in Petri dishes with ease. There are two sexes in *C. elegans*, hermaphrodite and male, with the self-fertilizing hermaphrodites being the most common. They were first used as a model species to study animal development and behaviour in 1965 and have since then become one of the most studied and used multi-cellular species (Riddle *et al*. 1997). Because of its research value, popularity, and relatively small genome (97 Megabases), *C. elegans* became the first multi-cellular organism to have its genome sequenced in 1998 (*C. elegans* Sequencing Consortium 1998). Its success as a model organism in research areas such as neuroscience, development, genetics, and RNA interference has led to its recent increased usage in environmental toxicology studies (Williams and Dusenbery 1990, Freeman *et al*. 1999, Leung *et al*. 2008). Its rapid growth and short progeny generation time allows us to examine the potential growth and reproduction impacts of IB exposure which, when analyzed in conjunction with RNA-Seq data, can provide insight on both gene-level and organism-level effects of exposure in an effort to elucidate the link between molecular responses and ecological impact.

Research Objectives

The main focus of this research project is to investigate the effects of ibuprofen exposure on two laboratory species through the application of NGS technology, specifically RNA-Seq, in ecotoxicogenomics.

There are four main objectives for the research conducted in this thesis:

1. To evaluate the benefits and challenges of conducting RNA-Seq and subsequent data analysis on rainbow trout, a species without a fully sequenced genome commonly used in toxicology studies. (Chapter 2)

2. To investigate the acute transcriptomic response to ibuprofen exposure in underyearling, sexually immature, rainbow trout. (Chapter 2)

3. To determine the effects of ibuprofen exposure on *C. elegans* behaviour and progeny generation. (Chapter 3)

4. To investigate the transcriptomic response to ibuprofen exposure in *C. elegans*. (Chapter 3)

Chapter 2.

Effects of Ibuprofen Exposure in Rainbow Trout

Acknowledgements

Others who have contributed to the work presented in this chapter:

- Ishai Kones (SFU co-op student) assisted with the setup of rainbow trout bioassays.
- Chemical analysis of water samples was performed by staff at the Pacific Environmental Science Centre (Environment Canada).
- RNA-Seq library prep and sequencing runs were performed by Reza Falsafi (UBC).
- Sequence data pre-processing was done by Chris Fjell (UBC), Geoff Winsor (SFU), and Matthew Laird (SFU).
- Melanie Pylatuk (SFU co-op student) assisted with qPCR plate setups.

Introduction

RNA-Seq is a relatively new but promising tool for ecotoxicogenomics research. At the time of conducting this research project, we were not aware of any reported studies using RNA-Seq to investigate the global gene expression response of fish to environmental toxicants. The work in this chapter was originally conceived as a "proof-of-concept" exploratory study to investigate the feasibility of using RNA-Seq on rainbow trout in the context of ecotoxicology.

Rainbow trout are routinely used in toxicology bioassays to test the toxicity of environmental pollutants. In the recent decade, there has been a push to extend traditional toxicological endpoints such as mortality to include molecular-level endpoints such as gene and protein expression perturbations. This integration of toxicology and molecular biology is known as ecotoxicogenomics. Previous toxicogenomics studies using rainbow trout relied on the use of microarrays and qPCR for gene expression analysis. With the recent advancements in NGS technology, there is a need to explore the use of these new and promising tools, specifically RNA-Seq, on species without a sequenced genome such as rainbow trout.

Ibuprofen (IB) was chosen as the environmental pollutant of interest due to its widespread use, prominence in sewage effluent, and detectable levels in surface water. Waterborne ibuprofen exposure has been shown to have negative impacts on aquatic organisms such as small crustaceans, bivalves and various species of fish by affecting biologically significant functions including the immune response, heat shock, stress, egg hatching, and reproduction (van Anholt *et al*. 2003, Gravel and Vijayan 2007, Flippin *et al*. 2007, Heckmann *et al*. 2007, Han *et al*. 2010, Contardo-Jara *et al*. 2011, Parolini *et al*. 2011, Ragugnetti *et al*. 2011, Gonzalez-Rey and Bebianno 2012, Matozzo *et al*. 2012, Saravanan *et al*. 2012).

The objectives of the work presented in this chapter are 1) to investigate the use of RNA-Seq on rainbow trout in the context of ecotoxicogenomics, 2) to highlight the benefits and challenges of working with a species lacking a reference genome, and 3) to use RNA-Seq to explore the impact of ibuprofen exposure on global gene expression in underyearling rainbow trout liver.

Materials and Methods

Exposure bioassays and tissue collection

The care and treatment of animals used in this study were in compliance with the Canadian Council on Animal Care, with approval from the Simon Fraser University Animal Care Committee. Underyearling rainbow trout (*O. mykiss*) were obtained from Fraser Valley Hatchery (2009; Abbotsford, BC, Canada) and Sun Valley Trout Hatchery (2011; Mission, BC, Canada) and acclimated in the laboratory (15 ± 1 °C and 16:8-h light:dark photoperiod) for 14 weeks prior to use. Fish were fed 2.5% of their body weight of commercial feed pellet (Skretting, Vancouver, BC) daily, except during the test period.

Ibuprofen sodium salt (CAS = 31121-93-4) was purchased from Sigma-Aldrich Canada Ltd. (Oakville, ON, Canada) and dissolved in well water as a 1.5 g/L stock for the low dose exposures (2009) and as a 6 g/L stock for the high dose exposures (2011). The 96-hour exposures were conducted in 60 L glass aquaria at 15 ± 1 °C under a 16:8-h light:dark photoperiod and constant aeration at 6.5 ± 1 mL/min • L. Water quality parameters (dissolved oxygen, pH, conductivity and temperature) were measured at the beginning and end of the bioassays. Prior to the exposures, fish were randomly selected and placed into each test vessel containing up to 60 L of clean well water. The fish were allowed to acclimatize to the test vessels for 24 hours. Then at the start of the test, appropriate volumes of IB stock solutions were gradually added to each tank to obtain 60 L final volume. Six tanks and eight fish per tank were used for the low dose exposures; four tanks and ten fish per tank were used for the high dose exposures. For the low dose exposures, fish were exposed to environmentally relevant concentrations of IB: 0.5 ug/L and 1.0 ug/L and a higher concentration: 30 ug/L for 96 hours. For the high dose exposures, fish were exposed to a high concentration of IB: 1000 ug/L for 96 hours and then transferred into clean well water tanks for a recovery/depuration period of 24 hours.

For all exposure conditions, two fish were randomly removed from each tank at 4-h, 8-h, 24-h and 96-h time points (with an additional 120-h time point for the high dose experiment) and euthanized in 150 mg/L tricane methane sulfonate (MS-222; Syndel Labs, Vancouver, BC) dissolved in well water. At those time points, two fish were also

randomly removed from the control tanks. Because differential gene expression is calculated as relative to the control test samples, other factors (i.e. handling stress, lack of food, and loading density, etc.) are accounted for or are cancelled out because the control fish were treated in the same manner. Fish length and weight were measured post-mortem. The liver and brain were removed from each fish and stored in 1.5 mL tubes containing 1.0 mL of RNALater® (Ambion, Life Technologies, Inc., Burlington, ON) at -80 °C.

Water quality parameters and chemistry analysis

Basic water quality parameters, including temperature, dissolved oxygen levels, pH and conductivity, were measured at the beginning and end of the tests. One-litre water samples from each concentration were collected from the test vessels at the end of the test for chemical analysis of acidic drugs (IB) using liquid chromatography-mass spectrometry (LCMS). This analysis was performed by the chemistry section staff at the Pacific Environmental Science Centre (Environment Canada, North Vancouver, BC).

RNA and DNA extraction

Hepatic total RNA (n = 12 per treatment per time point) was extracted from 30 mg of liver tissue using RNeasy Mini Extraction kits (Qiagen, Mississauga, ON, Canada) following the manufacturer's protocol including the optional on-column DNase digestion step. RNA quality was evaluated using the Bio-Rad Experion Electrophoresis Station and the RNA StdSens Analysis Kit as per the manufacturer's protocol. RNA samples with RNA Quality Index (RQI) scores greater than 9.0 were considered acceptable for subsequent transcriptomic analyses.

Genomic DNA (gDNA) was extracted from 10 mg of brain tissue using DNeasy Blood & Tissue kits (Qiagen, Mississauga, ON) according to the manufacturer's protocol.

RNA-Seq

Library preparation and Illumina sequencing were performed at the R.E.W. Hancock Laboratory Sequencing Facility (University of British Columbia, Vancouver,

BC). Three RNA samples from the following conditions: 1.0 µg/L 96-h, 1000 µg/L 96-h, Recovery 120-h; plus three corresponding well water control samples were randomly selected for RNA-Seq (18 samples total). Sequencing libraries were generated using the Illumina TruSeq™ RNA Sample Prep Kit. Libraries were quantified using the KAPA Library Quant Kit for Illumina and normalized to 10 pM for cluster generation and sequencing, all following the manufacturers' protocols. Paired-end sequencing was done in multiplex across three lanes on an Illumina Genome Analyzer IIx sequencer as per manufacturer's protocols.

Adapter sequences and low quality bases were trimmed from the fastq reads and reads shorter than 35 bp were filtered out. *De Bruijn* graph-based *de novo* assembly was performed for each experiment dataset using ABySS version 1.3.5 and Trans-ABySS version 1.4.4. with k-mer values for assembly set at 36, 40, 44, 48, and 52 (Simpson *et al*. 2009, Robertson *et al*. 2010). Cufflinks 2.0.2 and reciprocal BLAST was used to merge redundant contigs to generate a final set of unique contigs (Roberts *et al*. 2010). Read mapping was performed using Bowtie2 2.0.2 and TopHat 2.0.6 with the unique contigs as the reference (Langmead and Salzberg 2012, Trapnell *et al*. 2009). Transcript sequencing depth was estimated using BEDTools 2.17.0 and subsequently used to generate read count values (Quinlan and Hall 2010). Finally, differential gene expression analysis and statistical testing were performed using R Bioconductor package, DESeq (Anders and Huber 2010) with FDR (false discovery rate) correction set at < 0.05 to compare IB-exposed samples to corresponding non-exposed controls. The gene list was annotated first by using BLASTx to search against the entire NCBI non-redundant protein database and then by manual review of the blast hits. The analysis workflow is summarized in Figure 2.1.

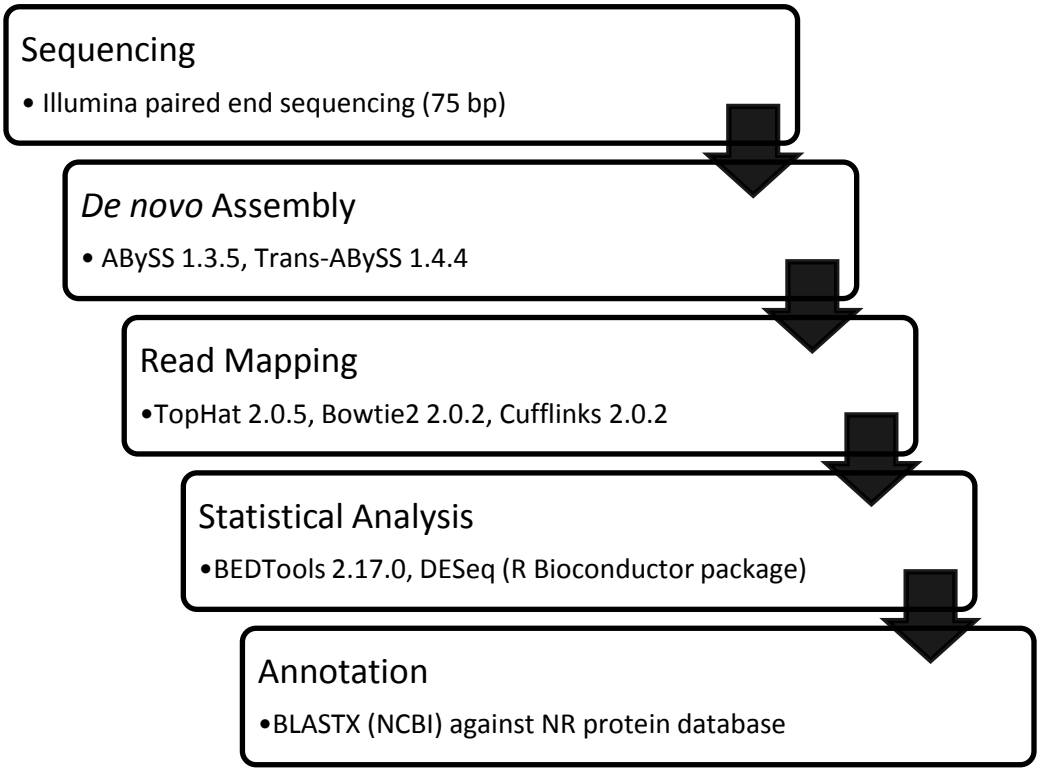

Figure 2.1. Summary of RNA-Seq analysis workflow outlining the software used in each step. All software used are open-source and free for public access.

QPCR

cDNA was prepared from 1 µg of total RNA using QuantiTect Reverse Transcription kits (Qiagen) according to the manufacturer's protocol. Nuclease-free water (Ambion) was used to generate a 10-fold working dilution for each sample. Samples were analyzed in triplicate 15 µL reactions consisting of 50% iTaq SYBR green Supermix (Bio-Rad, Mississauga, ON, Canada), nuclease-free water (Ambion), and 20 pmoles of each primer as per described in Osachoff et al. (2013). All primers produced a single amplicon as confirmed by dissociation melt-curve analysis; primer sequences are provided in Table 2.1. The reactions were run on Stratagene Mx3000P instruments (Agilent, Mississauga, ON, Canada) with an amplification profile consisting of 95°C for 1 min, followed by 40 cycles of denaturing for 20 s at 95°C, annealing for 30 s at 58°C and elongating for 40 s at 70°C.

Table 2.1. Rainbow trout genes used in qPCR analysis.

Gene	Gene Symbol	Primer Sequences	Amplicon Size (bp)
Vitellogenin	VTG	ATGAGAGCAGTAGTACTTG TCTTGCACACTCCCTGAGC	488
Prostaglandin D Synthase	PGDS	ATGCCTCAGAAAGACTTCAA AATGTTCCCAGAGAGGATTC	447
Ribosomal Protein L8	L8	CAGGGGACAGAGAAAAGGTG TGAGCTTTCTTGCCACAG	192
β-Actin	ACTB	GTCAGGCAGCTCGTAGCTCT CTGACCCTGAAGTACCCCATT	540
Ribosomal Protein S10	S10	CTGCACCTTCCACCTGAGAT TGTAGCTGAACCAGCACCTG	210
OmyY1 Y-specific	OmyY1	GTTCATATGCCAGGCTCAAC CGATTAGAAAGGCCTGCTTG	792

Gene expression data was normalized to the geometric mean expression of a suite of housekeeping genes: ribosomal protein L8, β-actin, and 40S ribosomal protein S10, using the method described in Vandesompele et al. (2002). Outliers were detected using Grubbs' test (Filby and Tyler 2007) and removed from analysis. Non-parametric statistical tests were performed using Prism 5.04 (GraphPad Software Inc., La Jolla, CA) to analyze differences between control and IB-exposed fish (Mann-Whitney test, $p < 0.05$).

Genetic sex identification

Genetic sex of experimental fish was determined by characterizing a sex-specific locus in the rainbow trout Y-chromosome following the protocol described by Brunelli *et al*. (2008). gDNA extracted from fish brain tissue were amplified in 35 uL polymerase chain reactions (PCR) containing 2 mmol/L MgCl2, 0.25 mmol/L dNTPs, 1 x Mg-free PCR buffer (Fisher Scientific, Ottawa, ON), 10 pmol of forward and reverse primers, and 1 U of Taq polymerase (Fisher Scientific, Ottawa, ON). OmyY1 Y-specific primers (Table 2.1), targeting a 792 bp fragment, were purchased from AlphaDNA (Montreal, QC). Two sets of primers were included in each reaction: OmyY1 and β-actin (positive control). The reactions were conducted in MJ Research PTC-200 thermocyclers (Bio-Rad

Laboratories, Hercules, CA) with the program consisting of initial denaturation at 94°C for 3 mins followed by 30 cycles of denaturation at 94°C for 50 s, annealing at 60°C for 50 s, and extension at 72°C for 60 s plus a final elongation at 72°C for 2 mins. The amplified PCR products were visualized using the Qiaxcel® electrophoresis system (Qiagen, Mississauga, ON).

Results

Water quality parameters and chemistry analysis

Water quality parameters were recorded for each test vessel at the beginning and end of the experiments. All measured parameters were determined to be acceptable for fish well-being and are summarized as follows: temperature: 15.0 ± 0.5 °C; dissolved oxygen: 9.1 ± 1.0 mg/L; pH: 7.5 ± 0.5; and conductivity: 465 ± 5 µS. There were no fish mortalities during the experiments.

At the end of the experiments, a 1-litre water sample from one tank from each concentration was collected and analyzed for acidic drug content. The measured concentrations of IB are shown in Table 2.2. The 1000 µg/L concentration was not analyzed. The measured concentrations were close to 70% of target for 0.5 µg/L and 30 µg/L but only 55% for 1.0 µg/L. Since these measurements were taken at the end of the test (96 hours of static non-renewal), some of the IB in the tanks may have been lost (degraded) or taken up by fish.

Table 2.2. Ibuprofen concentration in test vessels as determined by LCMS.

Target Ibuprofen Concentration	Measured Concentration	% of Target
30 mg/L (stock)	23 mg/L	76.7%
0.5 µg/L	0.35 µg/L	70%
1.0 µg/L	0.55 µg/L	55%
30 µg/L	20.7 µg/L	69%

Genetic sex of test fish

Fish sex was genetically determined by amplifying a male-specific 792 bp gene fragment through PCR analysis. The results were visualized using gel electrophoresis which separated the PCR products according to their sizes. Gel images were analyzed manually to detect the absence or presence of the 792 bp product band, which is indicative of male sex. To ensure that sample gDNA were appropriately loaded into the reactions, the PCRs were run in multiplex, with β-actin primers in addition to OmyY1 sex primers. The presence of a β-actin band act as a positive control and would confirm the correct inclusion of sample gDNA for each reaction.

An example of an analyzed gel image is shown in Figure 2.2. The Qiagen® gel electrophoresis software uses two alignment markers that flank the gel region and enable the alignment of multiple sample lanes for visual analysis. The example demonstrates the identification of male-positive bands and the presence of β-actin products in all reactions. In this example, samples one, five, and six are male and samples two, three, and four are female.

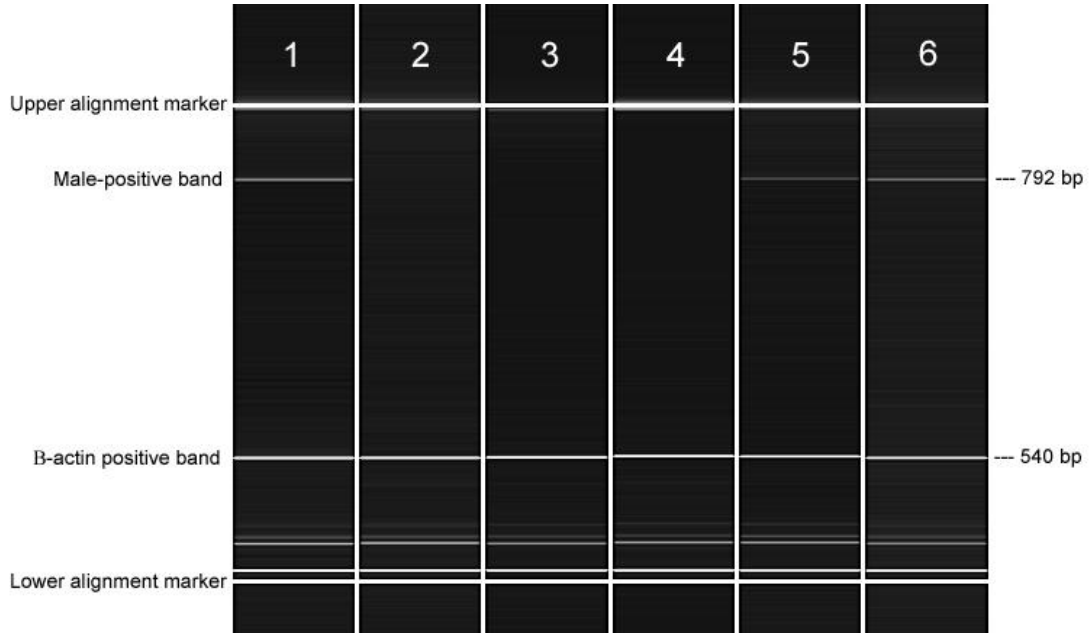

Figure 2.2. Gel electrophoresis image of sex identification PCR products. Upper and lower alignment markers flank the gel region. OmyY1 gene fragments are 792 bp (male-positive bands) and β-actin fragments are 540 bp (positive controls). In this example, lanes 1, 5, 6 are male fish and lanes 2, 3, 4 are female fish.

Experimental underyearling rainbow trout from the 96-hr time point of all exposure concentrations were genetically sexed. The results are shown in Table 2.3. The sex ratios are nearly equal (close to 50%) for all conditions reviewed, with the exception of the 1000 ug/L concentration. There were substantially more male fish in that exposure, an unfortunate coincidence which may bias the gene expression results as all three fish used in RNA-Seq analysis were male. At the time that fish were chosen for RNA-Seq, the genetic sexes of the fish were not yet determined.

Table 2.3. Genetic sex of rainbow trout used in experiments.

Ibuprofen Concentration (96-hr)	Total Num. of Fish	Num. Females (%)	Num. Males (%)	Ratio Females : Males Used in RNA-Seq
Control	12	5 (42%)	7 (58%)	2 : 1
0.5 µg/L	12	7 (58%)	5 (42%)	-
1.0 µg/L	12	5 (42%)	7 (58%)	1 : 2
30 µg/L	12	6 (50%)	6 (50%)	-
Control	8	4 (50%)	4 (50%)	2 : 1
1000 µg/L	8	1 (12.5%)	7 (87.5%)	0 : 3

RNA-Seq

Over 54 million 75-bp paired-end reads were obtained via high-throughput sequencing and were subsequently assembled into more than three million contigs via *de novo* assembly. Contigs less than 200 bp long were filtered out with 180,333 contigs remaining. Overlapping contigs were then merged to eliminate redundancy and resulted in 32,151 unique transcripts with at least five reads mapped. These unique transcripts were annotated using BLASTX to search against the NBCI non-redundant protein database. Of these, 16,831 contigs had BLASTX hits and were successfully annotated with alignment criteria set to e-value ≤ 0.01 or % identity ≥ 60%. Detailed information on sequencing and assembly statistics are shown in Table 2.4.

Table 2.4. Sequencing and assembly results.

Description	Value
Num. of libraries	18
Total num. of reads generated	54,464,467
Min. num. reads per sample	548,265
Max. num. reads per sample	7,324,955
Avg. num. reads per sample	3,025,804
Total num. contigs assembled	3,451,963
Num. contigs size > 200 bp	180,333
Unique transcripts	32,151
N50	536
Num. annotated (via BLASTX)	16,831

RNA-Seq gene expression

Statistical evaluation of contig read counts revealed 125 differentially expressed genes (DEG) that were affected in one or more of the conditions as follows: 47 DEGs (21 up, 26 down) for 1 µg/L IB, 60 DEGs (29 up, 31 down) for 1000 µg/L IB, and 43 DEGs (28 up, 15 down) for recovery (Table 2.5). Of these, 102 DEGs were successfully annotated. The overlap of genes differentially expressed in each condition is shown in Figure 2.3. Five genes were differentially expressed across all conditions (adjusted p-value < 0.05): prostaglandin D synthase (PTGDS or PGD synthase), vitellogenin (VTG), growth hormone 2, olfactory receptor 4C9, and MHC class I antigen. The observed impact on these five genes are supported by a larger sample size since all fish in all examined conditions were affected. Increased expression of PTGDS and decreased expression of VTG were observed during IB exposure and the effects persisted during depuration (recovery) phase (Figure 2.4).

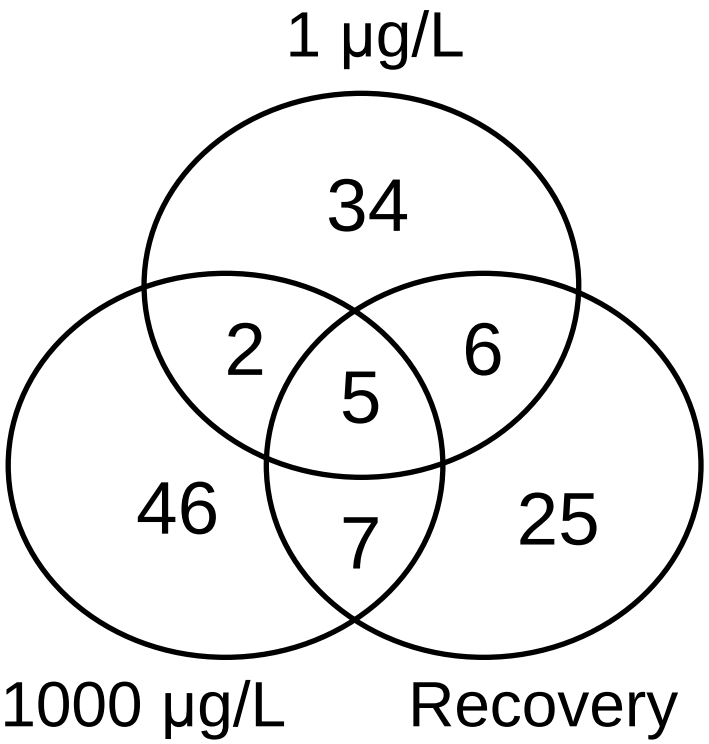

Figure 2.3. **Distribution of differentially expressed genes for each condition. The value within the intersections of the circles represent the number of overlapping genes amongst the conditions.**

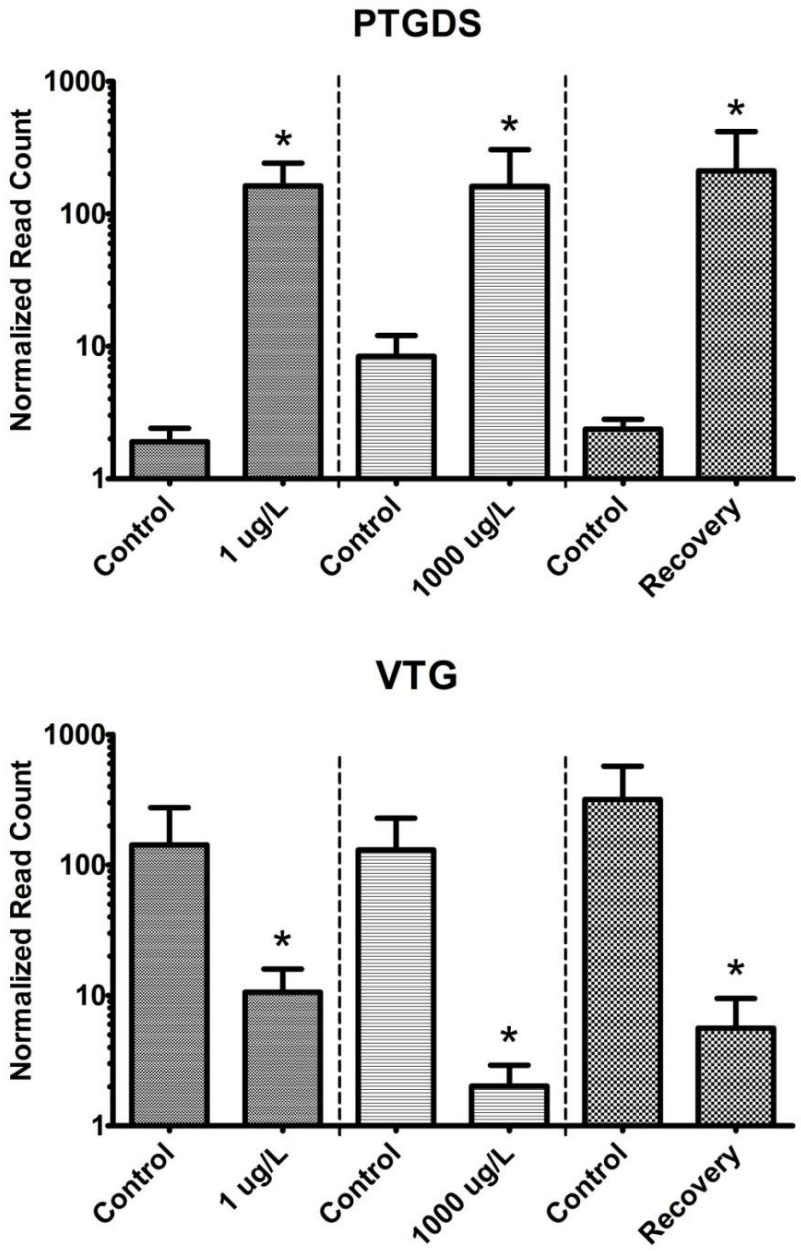

Figure 2.4. **Prostaglandin D Synthase and Vitellogenin gene expression in rainbow trout exposed to IB (1 µg/L, 1000 µg/L, Recovery). Asterisks denote statistically significant gene expression changes between fish exposed to IB and its corresponding control group as detected by DESeq method with adjusted P-value < 0.05 (FDR correction).**

Table 2.5. List of annotated genes (102) differentially expressed in at least one experimental condition. Fold change values are shown for genes with adjusted p-value < 0.05 (FDR correction).

Gene Class; Function	Gene	Abbrev.	Accession #	1.0 ug/L	1000 ug/L	Recovery
Blood						
oxygen transport	hemoglobin subunit alpha	HBA	NP_001118023	3.01	-	-6.63
oxygen transport	hemoglobin subunit beta-1	HBB	NP_001154027	2.98	-	<-10*
Cell Cycle						
apoptosis	serine/threonine protein kinase	PIM3	NP_001133339	-2.13	-	-
apoptosis	B-cell receptor-associated protein 29	BCAP29	NP_001019391	-	3.03	-
cell division	fibroblast growth factor 2	FGF2	ACJ02099	-	-3.39	-
cell division	centromere protein Q	CENPQ	NP_001140135	-	-4.35	-
cell growth	notch homolog 2	NOTCH2	NP_001108566	-	-	2.0
DNA						
repair	DNA repair protein homolog A	RAD51	NP_001134027	-	-3.56	-
repair	eosinophil cationic-type ribonuclease 5 precursor	EAR5	NP_062271	-	3.06	-
replication	transposon TX1 uncharacterized 149 kDa protein	ORF2	XP_003201132	>10*	-	-
transposition	olfactory receptor family C subfamily 4 member 9	-	AEB77803	-5.96	-4.79	2.83
Growth						
control	growth hormone 1	GH1	EU621900.1	-	-	<-10*
control	growth hormone 2	GH2	EU090916.1	-5.76	-6.26	-3.93
organ development	frizzled-related protein	FRZB	NP_571018	-	-	4.6
Immune System						
antiviral defense	fish virus induced TRIM protein	finTRIM	CAP08949	-	-	-5.91
chemotaxis	LECT2 neutrophil chemotactic factor	LECT2	NM_001124309.1	-	4.0	2.38
immune	b-cell receptor CD22	CD22	XP_002660657	-	-	<-10
immune	C1q-like adipose specific protein	C1Q	ABY55260	7.2	-	3.71
immune	cathelicidin-derived antimicrobial peptide 2 isoform A	CATH-2A	AAT44539	-	>10*	-
immune	CD209 molecule	CD209	NP_001118105	-	3.37	-
immune	complement C1q-like protein 2	C1QL2	NP_001134256	-	5.32	3.76
immune	complement component C8 gamma	C8G	NP_001117880	-	-	>10*
immune	immunoglobulin heavy chain IgH	IGH	AH014877.2	>10*	-	3.14
immune	inducible nitric oxide synthase	NOS2	NM_001124359.1	-4.56	-	-
immune	integrin beta-like protein 1	ITGBL1	NP_001019243	>10*	-	-
immune	interferon-inducible GTPase	IIGP1	EU221179.1	-	4.27	-
immune	interleukin-1 receptor type 2	IL1R2	NP_775465	-3.9	-	-

Gene Class; Function	Gene	Abbrev.	Accession #	1.0 ug/L	1000 ug/L	Recovery
immune	interleukin-1 receptor-associated kinase 3 IRAK3	IRAK3	XM_002823493.1	-	-	2.46
immune	MHC Class I A region	HLA-A	AB162342.1	>10*	-	-
immune	MHC Class I antigen	MHC	FJ969490.1	>10*	>10*	<-10*
immune	MHC Class I B region	HLA-B	AB162343.1	-	-2.33	-4.89
immune	MHC Class II alpha	DAA	CAB96451	-	4.33	-
immune	MHC Class II Antigen	MHC2	EU008541.1	4.56	-	-
immune	protein THEMIS isoform 3	THEMIS	NP_001158159	-	-	3.77
immune	complement protein component C7-1	C7-1	NP_001118090	-	2.92	-
inflammatory response	C-C motif chemokine 19	MIP3B	ACN09874	-6.8	-	-
inflammatory response	serum amyloid A	SAA	CAM12348	-	4.12	-
inflammatory response	toll-like receptor II	TLR2	NP_001117891	-	-	1.81
inflammatory response	uteroglobin, phospholipase A2 inhibitor	UGB	NP_035811	-4.19	-	-
innate immune response	cAMP-dependent transcription factor	ATF-1	ABA06591	-4.27	-	-
innate immune response	complement factor H1 protein	CFH	NP_001117899	<-10*	-	-
innate immune response	toll-like receptor 5	TLR5	BAC65467	<-10*	-	-
Iron						
storage	ferritin middle subunit	FRIM	BT073320.1	-	-2.81	-
binding	CYP19B1 promoter region	CYP19B1	JQ950348.1	<-10*	-	-
Metabolism						
amino acid	L-threonine 3-dehydrogenase	TDH	NP_998410	-3.25	-	-
amino acid	tyrosine aminotransferase	TAT	NP_001071022	3.14	-	-
drug	cytochrome P450 2K5	CYP2K5	NP_001118214	3.25	-	-
drug	cytochrome P450 2X12	CYP2X12	NP_001073322	3.37	-	-
drug	cytochrome P450 2X9	CYP2X9	NP_001070921	3.58	-	-
calcium ion	25-hydroxyvitamin D-1 alpha hydroxylase, responsive to estrogen stimulus	CYP27B1	XP_003199448	-	-2.05	-4.28
gluconeogenesis	ACN9 protein homolog	ACN9	ACM09009	-	-2.59	-
gluconeogenesis	phosphoenolpyruvate carboxykinase, cytosolic	PPCKC	NP_001133921	-2.52	-	-
lipid	prostaglandine D synthase	PTGDS	AF281353.1	7.08	4.97	6.82
proteolysis	matrix metalloproteinase 9	MMP-9	NP_001117842	-3.49	-	2.32
Reproduction						
egg protein	vitellogenin	VTG1	CAA63421	-4.09	-9.78	-6.07
Sensory						
neural	crestin	CRESTIN	NP_571068	-	-	3.44
olfactory	olfactory receptor family C subfamily 15 member 2	-	HM133611.1	-	-5.18	-
olfactory	pheromone receptor	V1R	NP_001119923	-	-	-2.09
Signal Transduction						
signal transduction	G-protein	P-ras	NM_001124705.1	-	3.27	-

Gene Class; Function	Gene	Abbrev.	Accession #	1.0 ug/L	1000 ug/L	Recovery
signal transduction	microfibrillar-associated protein 4	MFAP4	NP_001134717	-	4.16	-
signal transduction	SYPG1	SYPG1	DQ246664.1	-3.42	-	-
intracellular signaling	syndecan-4 precursor	SDC4	NP_001041614	-3.21	-	-
Stress Response						
DNA integration	retrotransposable element Tf2 155 kDa protein type 3	RT23	ACN60195	-3.91	-	-
DNA integration	retrotransposable element Tf2 155 kDa protein type 1-like	TF2-1	XP_003199928	-	3.56	-
stress response	DnaJ homolog subfamily C member 3, interacts with HSPA8	DNAJC3	NP_001134029	-	3.61	-
stress response	glucose-regulated protein 78kDa	HSPA5	BAD90025	-	3.34	-
stress response	heat shock 90kDa protein 1 beta	HSP90B	BAD90024	-	6.07	-
stress response	heat shock protein 90kDa alpha	HSP90A	NP_001167173	-	3.53	-
stress response	heat shock cognate 70 kDa protein, interacts with DNAJC3	HSPA8	NP_001117704	-3.83	3.87	-
oxidative damage	glutathione peroxidase 3	GPX3	BT072794.1	-	-	2.67
Structural						
skeletal	elastase-1	ELA1	Q7SIG3	-	-3.35	3.77
skeletal	myomesin-3	MYOM3	XP_001921065	-	-	2.3
skeletal	troponin I, slow skeletal muscle	TNNI1	ACI66721	4.11	-	-
skeletal	type II keratin E1	E1	NP_001118215	-	<-10*	-
muscle contraction	isotocin-neurophysin IT 1	IT	AB164430.1	-	-6.67	-
Transcription						
transcription factor	TATA box binding protein (Tbp)-associated factor	-	BC027813.1	-	4.49	-
regulation	aryl hydrocarbon receptor 2 alpha	AhR2a	FJ969489.1	-	<-10*	-
regulation	forkhead box protein I2	FOXI2	HM159472.1	-	-	4.75
Cellular Transport						
calcium	mucolipin-1	MCOLN1	XP_001336199	-	-6.93	7.03
calcium	nucleobindin 2a	NUCB2A	NP_958901	-	3.56	-
calcium	nucleobindin 2b precursor	NUCB2B	NP_958887	-	3.04	-
cellular transport	ferric-chelate reductase 1	FRRS1	NP_001167281	-	-	2.78
cellular transport	myosin 1	MYH1	EU221180.1	>10*	-	-
cellular transport	sodium bicarbonate cotransporter	NBC	NP_001117797	-	-	2.2
cellular transport	solute carrier family 28 member 3	SLC28A3	ACN58643	-	-	4.98
ion	P2X purinoceptor 4	P2RX4	NM_001140204.1	-	-1.92	-
lipid	apolipoprotein A-IV	APOA4	BT048822.1	-1.77	-	-
thyroid hormone	monocarboxylate transporter 8	MCT8	NP_006508	6.4	-	-
zinc	zinc transporter	ZIP7	HM208332.1	5.65	>10*	-

Gene Class; Function	Gene	Abbrev.	Accession #	1.0 ug/L	1000 ug/L	Recovery
zinc	membrane-associated ring finger 5b	march5b	FR749991.1	-	-3.92	-
Miscellaneous						
proteinase inhibitor	cystatin	CYT	AAA82049	-	-1.82	-
RNA-binding	transposase element	-	XP_003084866	-2.89	-	-
viral envelope	envelope glycoprotein syncytin-A	SYNA	NP_001013773	-	-1.45	-
Unknown	C-polysaccharide binding protein 1	TCBP1	AAG30020	-	-	-2.14
Unknown	C-polysaccharide binding protein 2	TCBP2	AAB35349	3.21	-	-
Unknown	CQ067 protein	C17orf67	NP_001139884	-	-3.94	-
Unknown	differentially regulated trout protein 1	-	NP_001118073	-	5.01	-
Unknown	hypothetical protein	-	XP_001919924	-	3.41	-
Unknown	hypothetical protein	-	NP_001129955	-	<-10*	-
Unknown	low quality protein	-	XP_685503	-	-	-2.81
Unknown	uncharacterized protein	-	NP_998672	-	-1.63	-
Unknown	uncharacterized protein	-	NP_001077318	-4.18	-	2.12

*When the expression value of the corresponding control sample = 0 and the IB-exposed sample expression value > 0, the fold change is denoted as > 10. Similarly, when the expression value of the corresponding control sample > 0 and the IB-exposed sample expression value = 0, the fold change is denoted as < -10.

Functional gene set enrichment analysis

Gene Set Enrichment Analysis (GSEA) was conducted on the 102 differentially expressed genes detected by RNA-Seq. Because rainbow trout lacks functional and pathway annotation in the Gene Ontology (GO) and KEGG databases, the differentially expressed genes were first matched for orthologs in zebrafish using BLASTX. Out of 102 genes, 96 had zebrafish orthologs and the corresponding zebrafish RefSeq accession numbers were used as input into WebGestalt software (Wang *et al*. 2013). Both GO and KEGG pathways were analyzed for enrichment using the zebrafish genome as the reference set and a hypergeometric statistical method with Benjamini-Hochberg multiple testing correction. Twenty-two biological process, seven molecular function and six cellular component GO categories were significantly affected by IB exposure, including many relating to immune function and stress response (Table 2.6). Four KEGG pathways were found to be significantly impacted by ibuprofen exposure, including a pathway related to oocyte maturation (Table 2.7).

Table 2.6. Gene Ontology overrepresentation analysis results.

GO	Description	Num. of Genes	Adj. P-value
Biological Process			
GO:0006950	Response to stress	17	1.05e-09
GO:0009617	Response to bacterium	7	2.17e-07
GO:0050896	Response to stimulus	31	2.50e-07
GO:0051707	Response to other organism	7	5.92e-07
GO:0051704	Multi-organism process	7	7.00e-07
GO:0009607	Response to biotic stimulus	7	7.00e-07
GO:0006952	Defense response	6	1.08e-05
GO:0009611	Response to wounding	6	0.0002
GO:0031347	Regulation of defense response	2	0.0056
GO:0006955	Immune response	5	0.0185
GO:0015671	Oxygen transport	2	0.0260
GO:0019882	Antigen processing and presentation	3	0.0310
GO:0015669	Gas transport	2	0.0310
GO:0070482	Response to oxygen levels	2	0.0420
GO:0002684	Positive regulation of immune system process	2	0.0420
GO:0001666	Response to hypoxia	2	0.0420
GO:0006954	Inflammatory response	2	0.0420
GO:0002376	Immune system process	6	0.0420
GO:0042742	Defense response to bacterium	2	0.0420
GO:0036293	Response to decreased oxygen levels	2	0.0464
GO:0045087	Innate immune response	2	0.0464
GO:0042060	Wound healing	3	0.0496
Molecular Function			
GO:0005506	Iron ion binding	7	0.0006
GO:0020037	Heme binding	5	0.0040
GO:0046906	Tetrapyrrole binding	5	0.0053
GO:0005344	Oxygen transporter activity	2	0.0240
GO:0019825	Oxygen binding	2	0.0240
GO:0005515	Protein binding	13	0.0411
GO:0051082	Unfolded protein binding	3	0.0411

GO	Description	Num. of Genes	Adj. P-value
Cellular Component			
GO:0005833	Hemoglobin complex	2	0.0246
GO:0042611	MHC protein complex	3	0.0267
GO:0044459	Plasma membrane part	6	0.0273
GO:0042612	MHC class I protein complex	2	0.0287
GO:0032991	Macromolecular complex	12	0.0394
GO:0043234	Protein complex	10	0.0465

Table 2.7. KEGG pathway overrepresentation analysis results.

Pathway Name	Num. of Genes	Adj. P-value
Protein processing in endoplasmic reticulum	5	0.0001
NOD-like receptor signaling pathway	2	0.0119
Cell adhesion molecules (CAMs)	2	0.0294
Progesterone-mediated oocyte maturation	2	0.0294

qPCR gene expression

qPCR was conducted on the remaining fish and corresponding controls from 0.5 µg/L, 1.0 µg/L, 30 µg/L, 1000 µg/L IB exposures. When female and male fish were analyzed together, both PGD synthase and VTG were not detected to be differentially expressed (data not shown). Incorporating genomic sex data in the analysis revealed sex-dependent gene expression differences. Female fish exposed to 0.5 µg/L and 30 µg/L of ibuprofen had lower PGD synthase expression compared to the female control fish, an effect not seen in male fish (Figure 2.5). The decrease in expression levels were calculated to be -3.22 and -2.29 respectively.

Similarly, sex-dependent gene expression levels were also seen in VTG data. Comparing the males and females within an exposure group revealed that sexually premature males and females have statistically different baseline VTG expression, regardless of ibuprofen exposure (Figure 2.6). A decrease in VTG transcript levels was

observed only in male fish exposed to 0.5 µg/L, 1.0 µg/L, 30 µg/L ibuprofen levels with fold changes calculated to be -1.46, -1.51, and -1.47 respectively.

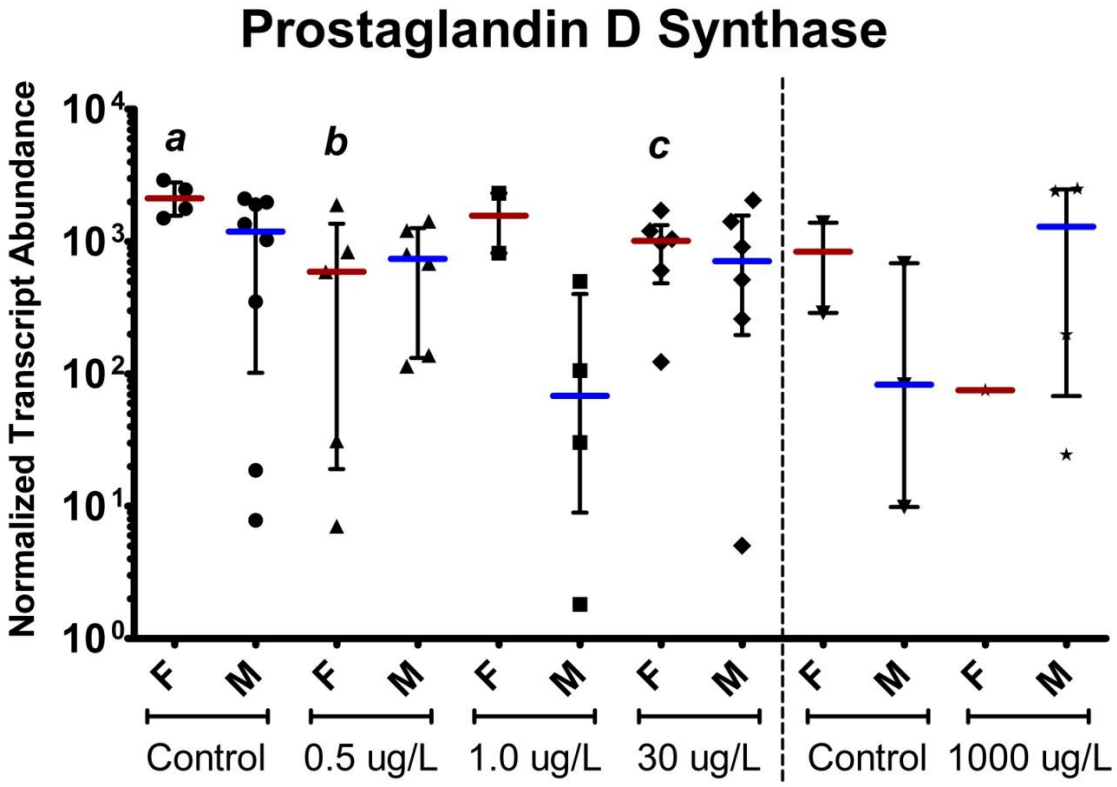

Figure 2.5. **qPCR detection of Prostaglandin D (PGD) Synthase gene expression in female (F) and male (M) underyearling rainbow trout. Statistical significance was determined using Mann-Whitney t-test with p-value ≤ 0.05. Data labels "a, b, c" denotes statistical significance, where a, b, and c are statistically different from each other. Ibuprofen exposure at 0.5 µg/L and 30 µg/L impacted PGD Synthase transcript abundance in female fish at statistically significant levels, with fold changes calculated to be -3.22 and -2.29 respectively.**

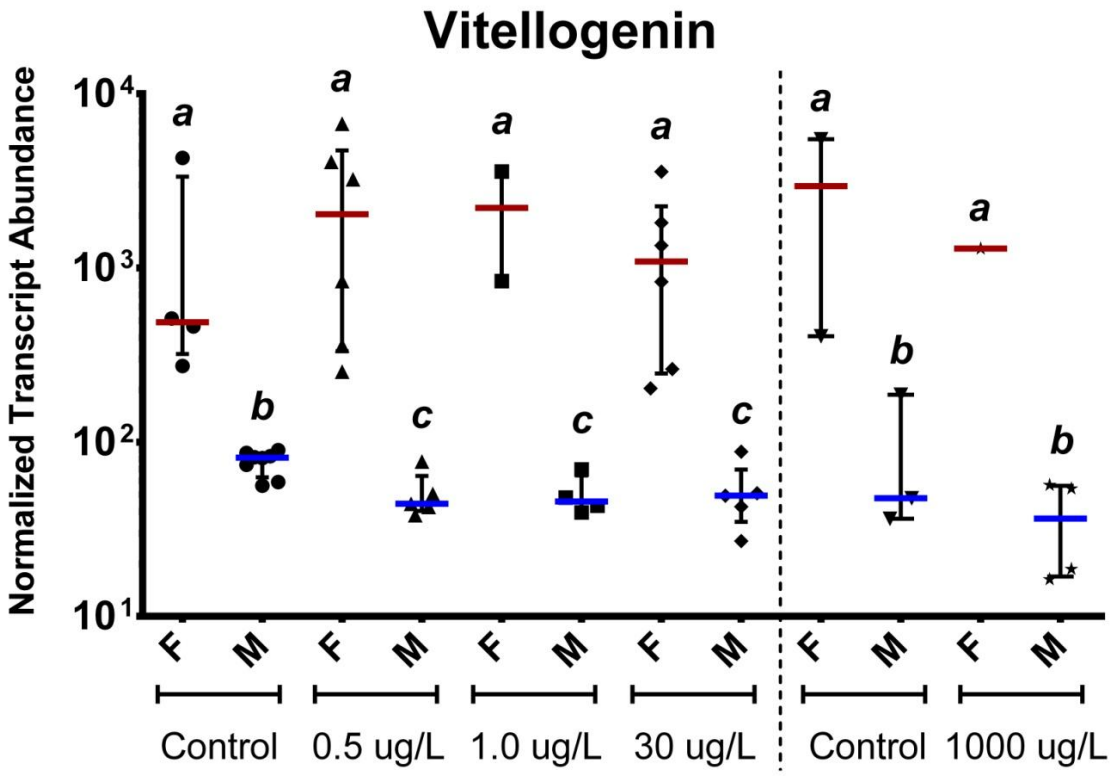

Figure 2.6. **qPCR detection of Vitellogenin (VTG) gene expression in female (F) and male (M) underyearling rainbow trout. Statistical significance was determined using Mann-Whitney t-test with p-value ≤ 0.05. Data labels "a, b, c" denotes statistical significance, where a, b, and c are statistically different from each other. Females and males have statistically different basal VTG expression levels. Ibuprofen exposure impacted VTG transcript abundance in male fish at statistically significant levels. Fold changes were calculated to be -1.46, -1.51, and -1.47 for ibuprofen levels 0.5 µg/L, 1.0 µg/L, 30 µg/L respectively.**

Discussion

General gene expression changes

Ibuprofen (IB) is a widely consumed and produced drug and is commonly found in sewage effluents at concentrations in the magnitude of micrograms per litre. Adverse effects of environmental IB exposure have been well-documented previously, including effects on osmoregulation, immunosuppression, and endocrine disruption in multiple

aquatic species (Gravel and Vijayan 2007, Matozzo *et al*. 2012, Bhandari and Venables 2011, Morthorst *et al*. 2013, Heckmann *et al*. 2007, Flippin *et al*. 2007, Han *et al*. 2010, Parolini *et al*. 2011). Several studies also reported on specific gene biomarkers using qPCR (Gravel and Vijayan 2007, Gravel *et al*. 2009) and general transcriptomic changes using microarrays (Heckmann *et al*. 2008, Milan *et al*. 2013). Amidst the next-generation sequencing era, this study is one of the first to use RNA-Seq in the context of environmental toxicology and one of the first studies to explore the effects of IB exposure on rainbow trout liver gene expression. The goal of this study is to examine liver gene expression changes in underyearling rainbow trout, a species routinely used in environmental toxicology, and to provide a framework for future toxicogenomics studies using RNA-Seq on other species currently without a reference genome but used widely in toxicology.

Using RNA-Seq, 125 transcripts were detected to have altered expression levels during and/or after IB exposure, 102 of which were successfully functionally annotated. GSEA on GO terms revealed that IB exposure affected genes in many biological processes associated with immune function and stress response (Table 2.6). An impact on these pathways have also been reported previously in other IB exposure studies involving rainbow trout, Indian major carp, and clams (Gravel and Vijayan 2007, Gravel *et al*. 2009, Matozzo *et al*. 2012, Saravanan *et al*. 2012). Furthermore, one of the KEGG pathways that were found to be significantly impacted by IB exposure is related to oocyte maturation (Table 2.7). Oocyte and reproduction-related effects were also reported in studies with *Daphnia magna* and other fish (Japanese medaka, zebrafish) (Flippin *et al*. 2007, Heckmann *et al*. 2007, Lister and Van Der Kraak 2008, Heckmann *et al*. 2008). The significance of these impacts and their biological implications are discussed later in this chapter.

qPCR validation

Five genes were detected by RNA-Seq to have been affected at all three examined IB conditions (1 µg/L, 1000 µg/L, and recovery). Because of time constraints, two genes were selected for qPCR validation: VTG and PGD synthase. Since it's known that IB acts on PG synthesis, PGD synthase was selected for qPCR validation to verify the impact of IB on this pathway. The observed decrease in VTG gene expression was

unexpected, so due to its wide usage as a biomarker for estrogenicity, VTG gene expression warranted a more detailed investigation. The other three genes (growth hormone 2, olfactory receptor 4C9, and MHC class I antigen) should be validated in future work. Using the remaining fish RNA samples from the 96-h time point for all exposure concentrations and their respective controls, qPCR was performed on VTG and PGD synthase to validate the RNA-Seq gene expression data. When female and male fish were analyzed together, we were unable to detect differential gene expression between control and IB-exposed fish. This is likely due to the substantial biological variations that we observed amongst the samples for both PGD synthase and VTG. However, when the genetic sex of the fish is taken into consideration, the analysis revealed sex-dependent gene expression differences for both genes. Female fish exposed to 0.5 µg/L and 30 µg/L of IB had lower PGD synthase expression compared to the female control fish, an effect not seen in male fish (Figure 2.5). Although PGD synthase was determined to be DE using qPCR (decreased expression), the direction of change is opposite to what was found via RNA-Seq analysis (increased expression). This discrepancy could be caused by the small sample size used for both analyses. The sample size of RNA-Seq analysis was small (n = 3) and the imbalance in sex distribution can likely confound gene expression analysis. Similarly, the sample size of qPCR data was also diminished, due to separating the sexes in the analysis. However, despite the challenges arising from small sample size, we have shown that it is important to take into account the sex of the test fish, even though the fish are sexually immature. This point is more apparent in VTG expression. Notably, we observed statistically significant differences between male and female basal vitellogenin gene expression levels regardless of IB exposure. Moreover, we were able to validate the RNA-Seq expression data, but statistically significant DE was only observed for male fish. (Figure 2.6). Further discussion on VTG is presented later in this chapter.

Stress response

Heat shock protein 70 (hsp70) is a highly conserved part of the heat shock protein family cascade that is involved in cellular response to stress as a defence mechanism to protect the cell (Kiang and Tsokos 1998). In addition to being responsive to heat stimulus, hsp70 is also known to be induced by heavy metals exposure

(Levinson *et al*. 1980), microbial infections, hormones, and antibiotics (Kaufmann 1990). In this study, hsp70 was found to have decreased expression during 1 µg/L IB exposure (fold change = -3.83) and increased expression at 1000 µg/L (fold change = 3.87); during depuration, hsp70 gene expression returned to normal levels (not significantly different compared to control non-exposed fish) (Table 2.5). Moreover, four other heat shock family and related genes had increased expression during 1000 µg/L IB exposure. Our results suggest that IB exposure perturbs the heat shock pathway, indicating a stress response to IB, and could ultimately have ramifications on fish cellular defense mechanisms. This effect was also observed by Gravel and Vijayan (2007), where hsp70 protein was initially induced in rainbow trout liver during four days of exposure to similar levels of IB. However, when fish were subjected to heat stress, IB-treated fish had significantly reduced hsp70 protein expression in both liver and gills, although transcript abundance was not altered. Their study showed that rainbow trout's ability to respond to heat shock stress was compromised due to IB exposure. Similar heat shock response disruption was observed in invertebrate species as well as mammalian systems where the highly conserved genes in the heat shock pathway were affected by IB exposure (Heckmann *et al*. 2008, Lagunas *et al*. 2004).

Prostaglandin biosynthesis pathway

The analgesic and anti-inflammatory activity of NSAIDs stems from inhibiting COX and subsequent PG production (Vane and Botting 1998, Rainsford 2009). Moreover, IB is a non-selective COX inhibitor, acting on both COX-1 and COX-2 (van Hecken *et al*. 2000). COX are functionally important enzymes in fatty acid metabolism, converting arachidonic acid into PGs, which are key regulators in reproduction, ion flux, and immunity in both vertebrates and invertebrates (Figure 2.7) (Rao and Knaus 2008, Rowley *et al*. 2005). Lister and Van Der Kraak (2008) showed that PGs played a significant role in zebrafish oocyte maturation and ovulation and suggested that inhibition of PG production (i.e. by NSAIDs) could disrupt and negatively impact zebrafish reproduction (Lister and Van Der Kraak 2008). In vertebrates, PG can be found in two classical forms: prostaglandin E_2 (PGE_2) and prostaglandin D_2 (PGD_2). Recent experiments of fish exposed to 20-500 µg/L IB concentrations showed significant reduction of PGE_2 (Bhandari and Venables 2011, Morthorst *et al*. 2013). Although we

were unable to directly detect effects of IB on COX gene expression using RNA-Seq, we found altered expression of two key enzymes that were upstream and downstream of COX action. Increased PGD synthase gene expression was seen at 1 µg/L and 1000 µg/L IB concentrations and the increase was sustained during depuration. Also at 1 µg/L, we detected a decrease in uteroglobin (UGB), which is a phospholipase A2 inhibitor (Levin *et al*. 1986). An increase in PGD synthase (positive control) and a decrease in UGB (negative control) could potentially be an adaptive response to an overall decrease in PG output due to IB action.

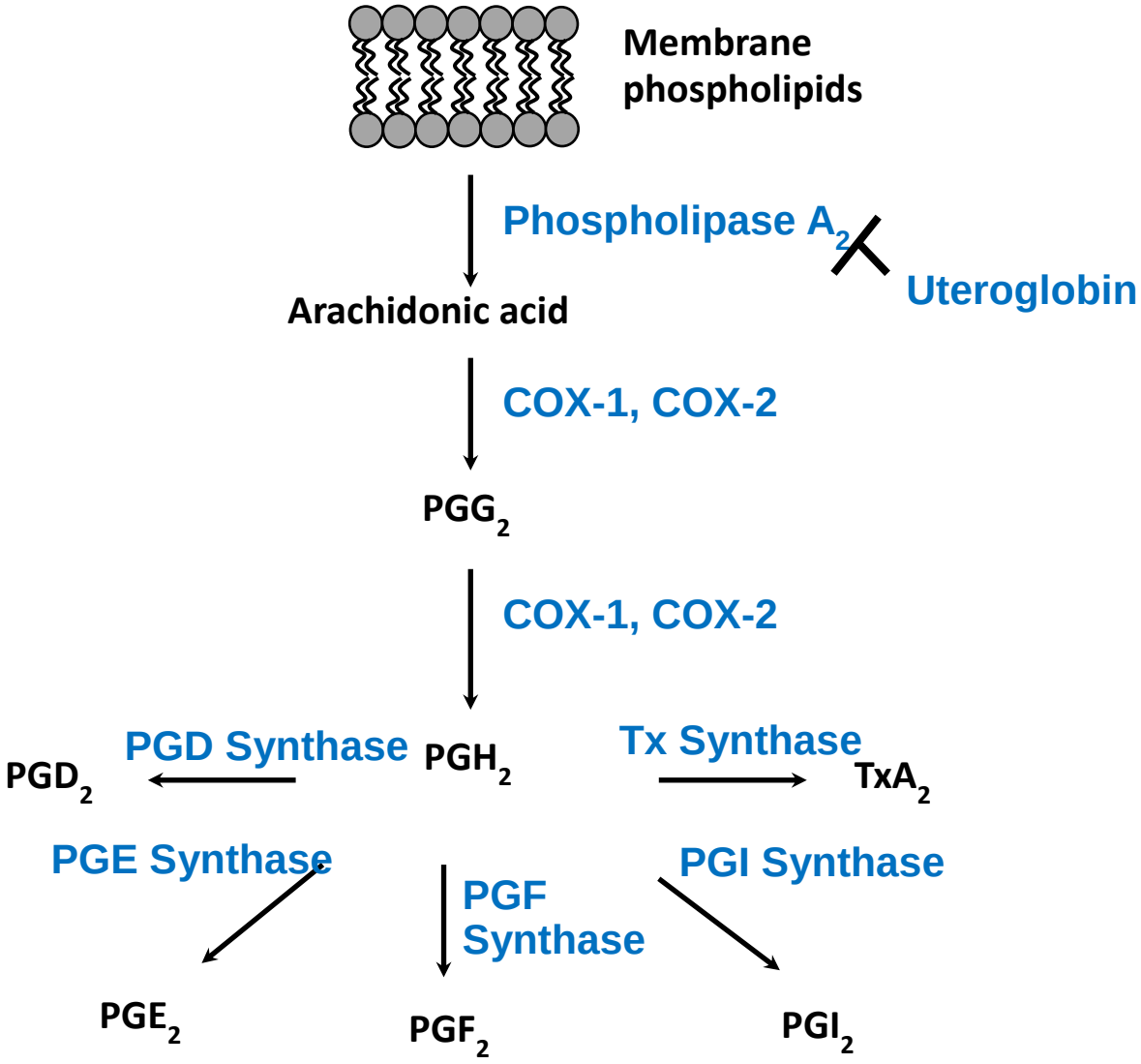

Figure 2.7. **Simplified representation of prostaglandin (PG) biosynthesis pathway (Rao and Knaus 2008, Rowley 2005). Enzymes that catalyze the reactions are shown in grey. Uteroglobin is an inhibitor of phospholipase A_2.**

Potential anti-estrogenic effects

The disruption of PG production could have negative repercussions on oogenesis and reproduction parameters. In both vertebrates and invertebrates, PGE2 regulates the expression of aromatase (CYP19), a key enzyme in endogenous estrogen synthesis from androgens (Brueggemeier *et al*. 2005, Richards and Brueggemeier 2003, Subbaramaiah *et al*. 2011). Therefore, a reduction in PGE2 (as seen in IB exposures)

could decrease aromatase activity, which subsequently could disrupt estrogen production resulting in potential anti-estrogenic effects. Negative impacts on reproduction parameters have been reported in numerous studies for multiple species exposed to IB. Heckmann et al. suspected IB disrupted eicosanoid metabolism and, consequently, oogenesis, which resulted in delayed reproduction, reduced fecundity, and overall reduced population growth rate observed in *Daphnia magna* (Heckmann *et al*. 2007, Heckmann *et al*. 2008). Similarly, in two independent studies of Japanese medaka exposed to IB, decreased number of spawning events and delayed hatching of fertilized eggs were observed (Flippin *et al*. 2007, Han *et al*. 2010).

Vitellogenin is the egg yolk protein precursor that is produced by females of almost all oviparous organisms (Wahli *et al*. 1981). In sexually mature female fish, it is synthesized in the liver and transported to maturing oocytes in the ovaries (Arukwe and Goksoyr 2003). Vitellogenesis can be induced in immature fish and male fish via exposure to xenoestrogens, compounds that mimic natural estrogens (Sumpter and Jobling 1995, Arukwe *et al.* 1997). Because of this property, VTG expression is commonly used as a biomarker for toxicant estrogenicity. In rainbow trout, using RNA-Seq and qPCR, we found that VTG gene expression was significantly decreased in sexually immature male fish at all exposure concentrations. It is unknown why this effect was not observed in females. One possibility is that females have much greater biological variation in VTG gene expression and this variability masked any significant effects of IB (Figure 2.6). The biological variation we observed is consistent with other studies that showed variable basal and induced VTG protein levels in same-sex fish of comparable age and sexual maturity (Tyler *et al*. 1991, Carlson and Williams 1999).

A possible mode of action for the observed reduction in VTG gene expression is based on aryl hydrocarbon receptor (AhR) activity. AhR plays a role in drug metabolism as a ligand-dependent transcription factor that activates the expression of several cytochrome P450 (CYP) genes involved xenobiotics metabolism pathway. Activated AhR have been shown to inhibit the expression of estrogen-responsive genes by disrupting estrogen receptor (ER) activity (Safe and Wormke 2003, Matthews and Gustafsson 2006, Mortensen and Arukwe 2007). Based on rainbow trout RNA-Seq data, we found that AhR and four CYP genes (CYP2K5, CYP2X12, CYP2X9, CYP19B1) were affected by IB exposure. Ibuprofen could have indirectly suppressed VTG gene

expression by activating AhR and disrupting the activity of ER. However, the actual mechanism of action might not be that simple and additional studies involving, for example, the measurement of AhR and ER activity and other key proteins in that pathway are needed to test that hypothesis. Although sexually immature fish have been documented to express basal levels of VTG, the biological implications of reduced VTG gene expression at this stage of their lives remain unknown (Tyler *et al*. 1991, Carlson and Williams 1999).

Sex-dependent VTG expression

Comparing the sexes of the non-exposed (control) fish, we also observed that sexually immature females and males expressed significantly differently basal levels of VTG gene (Figure 2.6). This could be due to differences in estrogen receptor abundance, affinity, and reactivity levels that have been observed previously in trout species (Campbell *et al*. 1994). Sex-specific levels of estrogen-induced, not basal, VTG production in rainbow trout was documented by Carlson and Williams (1999). Their results advocated the importance of considering the sex of immature fish when using vitellogenesis as a biomarker. At the time of their study, genetic sexing of juvenile rainbow trout was not available and fish were sexed by examining gonads post-mortem, which was difficult to do accurately. Given that a genetic sexing method for rainbow trout is now available (Brunelli *et al*. 2008) and our observations showed sex differences in basal VTG gene expression levels, we demonstrate and re-emphasize the importance of controlling for and/or considering the sex of fish, especially in studies using vitellogenesis as a biomarker for estrogenicity.

Benefits and challenges of RNA-Seq

As mentioned in Chapter 1, RNA-Seq has many advantages over existing hybridization and probe-based methods for conducting gene expression quantification experiments. The benefits include technical aspects such as increased sensitivity, greater dynamic range of detection, lower background levels, and a lower requirement of RNA sample input (Wang *et al*. 2009, Costa *et al*. 2010). Because RNA-Seq does not require existing sequence data, it can be applied on virtually any organism. This is

especially advantageous for a field (such as environmental toxicology) that uses a wide variety of organisms.

Given the massive amount of data that is generated from RNA-Seq experiments, there is a need for the development of computational infrastructure, including methods to store, retrieve, and process large amounts of data. In ecotoxicogenomics, a field that is still relatively young, obtaining and accessing the necessary bioinformatics support, both computational resources and data analysis expertise, can be quite challenging. Since the majority of the species commonly used in toxicological testing and research do not have fully sequenced genomes or sometimes no sequence data at all, there are more challenges that arise when using RNA-Seq in this field.

Some of the specific challenges faced in this study include generating a unique set of transcripts from *de novo* assembly that is suitable for read mapping and annotating the assembled transcripts. *De novo* assembly generated 180,333 contigs, which are not representative of unique transcripts. Before read mapping can be done, overlapping contigs and highly similar contigs must be merged so that reads that come from the same transcript are not mapped to multiple contigs that may represent the same gene. However, even after contig merging with Cufflinks software, there were still many similar contigs that were annotated to the same gene during BLASTX annotation. To resolve this issue, a reciprocal BLAST search was conducted where the set of contigs were aligned against each other and contigs that had sequence similarity score over a threshold value were grouped together. Within each group, the longest contig was kept and included in the final set of unique transcripts. The combination of Cufflinks merge and reciprocal BLAST steps filtered the 180,333 down to 32,151 unique transcripts.

The annotation efficacy of all de novo assembly transcripts (16,831 out of 32,151) and differentially expressed genes (102 out of 125) were 52.3% and 59.6% respectively; these values are comparable to those reported by other recent RNA-Seq studies (Coppe *et al*. 2010, Garcia *et al*. 2012, Palstra *et al*. 2013). Meaningful functional annotation is one of the key challenges arising from working with an organism without a sequenced reference genome (Vijay *et al*. 2013). For annotation methods based on BLAST and sequence similarity, success depends greatly on the stringency of alignment

match parameters and the filtering criteria if multiple "best hits" were found. Hornett and Wheat (2012) noted that there is a level of bias in annotation to be aware of, since well-annotated genes are typically those that are highly expressed and highly conserved amongst species while divergent genes, which are potentially more interesting, are more difficult to annotate and thus often excluded from the analyses. For this study, almost half of the transcripts generated through *de novo* assembly were unable to be annotated using BLASTX search against the entire NCBI non-redundant protein database. This suggests that a significant number of new sequences have been found and that as more annotation becomes available, these sequences can be re-analyzed.

Summary

This was one of the first studies to use RNA-Seq on rainbow trout in the context of environmental toxicogenomics. We found that IB exposure altered the expression of 125 genes in rainbow trout liver, including those involved in stress, immune response, heat shock, vitellogenesis, and prostaglandin biosynthesis. Notably, an effect on VTG gene expression was observed (only in male fish). We also found that genetically female and male fish expressed statistically different basal levels of VTG gene, despite being sexually immature, an observation that demonstrate the importance of controlling for and/or considering the sex of fish when using VTG as a marker for endocrine disruption.

Chapter 3.

Effects of Ibuprofen Exposure in *Caenorhabditis elegans*

Acknowledgements

Others who have contributed to the work presented in this chapter:

- Raymond Lo and Rylan Fernandez (SFU) cultured the worms used in the experiments and conducted all *C. elegans* exposure bioassays (including progeny counting).
- Emma Griffiths (SFU) performed RNA-Seq library prep and sequencing.
- Thea Van Rossum performed RNA-Seq adapter trimming, QA/QC, and assisted with gene expression analysis.

Introduction

The *C. elegans* experiments described in this chapter were conducted after observing the gene expression changes in underyearling rainbow trout exposed to ibuprofen. In previous studies, ibuprofen (IB) has been shown to have effects on reproduction and egg hatching in *Daphnia magna* (small crustacean) and also in Japanese medaka (fish) (Heckmann *et al*. 2007, Flippin *et al*. 2007, Heckmann *et al*. 2008, Han *et al*. 2010). Because *C. elegans* have a well-studied reproductive system, a short progeny generation time, and a fully sequenced genome, they are ideal model organism to use in this project for examining the effects of IB exposure on multiple endpoints. Using *C. elegans* allows us to look at the effects of IB not only at the organism level, by monitoring progeny production and behaviour, but also at the molecular level, by examining gene expression changes using RNA-Seq. It also allows us to explore the benefits of using organisms from multiple trophic levels, including traditional toxicology test species (rainbow trout) and more microscopic organisms (*C. elegans*), to better understand the impact of an environmental toxicant on an ecosystem.

The objectives of the work described in this chapter are 1) to determine if ibuprofen exposure has an impact on progeny generation in *C. elegans* and 2) to examine the gene expression response of *C. elegans* during ibuprofen exposure.

Materials and Methods

Exposure bioassays

Synchronized worms at late L3 or early L4 were incubated at 24°C in the absence of IB or with the indicated concentrations of ibuprofen in NGM plates covered with *E. coli* OP50. One individual worm was transferred onto a fresh plate, using six plates for each treatment. At 24 hour intervals, the worm was transferred onto a fresh plate with the same concentration of ibuprofen. Worms were monitored for behaviour and mobility differences.

For the RNA-Seq experiment, worms were exposed to 0.2 µg/L IB for 24 hours, using the same set up as mentioned above with three biological replicates on separate

E. coli OP50-covered NGM plates spiked with IB. Worms were submersed in RNA-Later and stored at -80°C.

Progeny counting

Every 24 hours, the number of eggs released and the number of hatched progeny from individual worms were counted right after transferring the original worms to a fresh plate. The counting and transferring of worms were repeated for up to five consecutive days. The number of eggs laid and hatched progeny per 24 hours was calculated by adding the total number of eggs and progeny counted for each plate before transferring to fresh plate. The cumulative egg count was also calculated by adding the total number of eggs released and hatched progeny for the indicated period of time.

RNA extraction

Total RNA was extracted from each worm using RNeasy Mini Extraction kits (Qiagen, Mississauga, ON, Canada) following the manufacturer's protocol, including the optional on-column DNase digestion step. RNA quality was evaluated using the Agilent 2100 Bioanalyzer and the Agilent RNA 6000 Nano Kit, as per the manufacturer's protocol.

RNA-Seq and analysis

Library preparation and Illumina sequencing were performed at the Simon Fraser University (Burnaby, BC). Ribosomally depleted RNA from three non-exposed worms (control) and three IB-exposed worms were used for sequencing, for a total of six samples. Sequencing libraries were generated using the Illumina TruSeq™ RNA Sample Prep Kit; libraries were quantified using NanoDrop™ spectrophotometer and normalized to 10 pM for cluster generation and sequencing, all following the manufacturers' protocols. Paired-end sequencing (150 bp) was done in multiplex in a single lane on an Illumina MiSeq sequencer, as per manufacturer's protocols.

Adapter sequences were removed from the raw fastq reads and low quality reads and reads shorter than 100 bp were filtered out. Reads were mapped to the *C. elegans* reference genome WS210 using TopHat 2.0.6 and Bowtie 2.0.0.7 with default settings (Langmead and Salzberg 2012, Trapnell *et al*. 2009). Cufflinks 2.0.2 was used to determine transcript abundance and read count values, which represent gene expression levels (Roberts *et al*. 2010). Genes with low and inconsistent read counts, either due to technical artefacts or biological variations, can decrease the sensitivity of statistical tests by introducing noise in the data which can confound multiple testing correction. As such, Bioconductor package HTSFilter was used to remove these genes and to normalize the dataset prior to statistical analysis (Rau *et al*. 2013). To compare the gene expression of IB-exposed worms to corresponding non-exposed controls, Bioconductor package DESeq with FDR (false discovery rate) correction set at adjusted p-value ≤ 0.1 was used (Anders and Huber 2010).

Results

Behaviour and survival

Due to subtle timing differences in the start of progeny production in different worms, multiple experiments were carried out to verify the reproducibility of our observations. A total of nine exposure experiments were carried out over a period of nine months in which *C. elegans* were exposed to various concentrations of IB ranging from 0.002 µg/L up to 5000 µg/L. In all experiments, we did not observe any noticeable differences in the behaviour and mobility of control and exposed worms. The worms did not exhibit any typical outward signs of stress (i.e. they did not try to escape or climb out of the Petri dishes). For all IB concentrations, the mortality rates of exposed worms were not significantly different from unexposed control worms (Table 3.1).

Table 3.1. Mortality rates of *C. elegans* exposure experiments.

Ibuprofen Concentration	Num. of Experiments*	Total Num. Mortalities	Average % Mortality
Control	9	6	11%
0.002 µg/L	3	1	6%
0.02 µg/L	3	4	22%
0.2 µg/L	9	9	17%
2.0 µg/L	5	7	23%
20 µg/L	1	1	17%
5000 µg/L	5	4	13%

*N = 6 for each condition in each experiment conducted.

Progeny generation

Ibuprofen exposure impacted *C. elegans* progeny production in all nine experiments. No significant effects were observed at the lowest concentration, 0.002 µg/L. Statistically significant reductions of cumulative progeny were observed at exposure levels as low as 0.02 µg/L and were seen consistently at 0.2 µg/L, noticeably starting as early as Day 3 of exposure (Figure 3.1). These IB concentrations, where a reduction in progeny generation was observed, are below reported environmental levels. In the highest experimental concentration (5000 µg/L), a decrease in cumulative progeny was evident in all experiments. Although this level is much higher than environmental levels, IB was not acutely lethal to *C. elegans* even at this high dose.

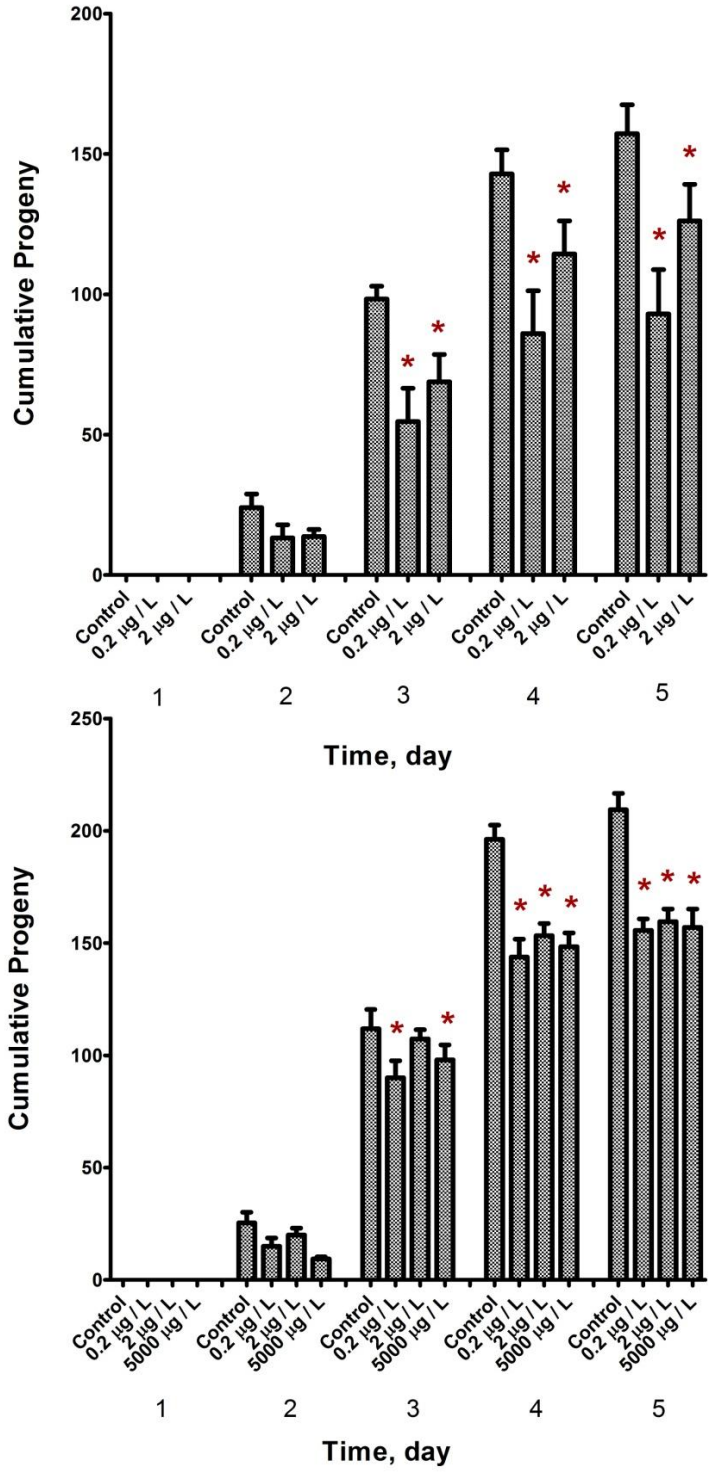

Figure 3.1. ***C. elegans*** **cumulative progeny production during IB exposure at various concentrations (two experiments shown out of a total of nine experiments performed). Asterisks denote statistical significant differences as compared to control worms (T-test p-value ≤ 0.05).**

Gene expression

After quality trimming, adapter trimming and minimum length filtering, an average of 1.1 million paired end reads (140-150 bp) per sample, with a range of 563k to 1.83 million, were used in subsequent analysis steps. Using Bioconductor packages HTSFilter and DESeq, 26 genes were found to be differentially expressed after 24 hours of exposure to 0.2 µg/L IB (FDR-corrected p-value ≤ 0.1) (Table 3.2). We found that 24 genes had higher expression in control (non-exposed) worms while only 2 genes had higher expression in IB-exposed worms. Out of 26 differentially expressed genes, 14 had functional annotation in WormBase and NCBI. Out of the 14 annotated genes that are affected by IB exposure, 8 had functions related to growth and development, while 2 of 14 were related to reproduction. Other functional groups affected by IB exposure include metabolism and cell structure. WebGestalt functional gene set enrichment analysis (GSEA) also revealed three statistically significant functional groups (Table 3.3).

Table 3.2. ***C. elegans*** **genes affected by IB exposure as determined by statistical analysis (DESeq) and FDR correction (adjusted p-value < 0.1).**

Gene Class; Function*	Gene	Abbrev.	WormBaseID	Refseq	0.2 ug/L
Development					
growth	stearoyl-CoA desaturase	FAT7	WBGene00001399	NP_504814.1	-0.91
growth	osmotic avoidance	OSM11	WBGene00003891	NP_510823.1	-0.97
growth	eukaryotic translation initiation factor 5A-2	EIF-5A-2	WBGene00002065	NP_495807.1	-0.80
growth	chaperonin homolog	HSP60	WBGene00002025	NP_497429.1	-0.61
growth	osteonectin	OST1	WBGene00003893	NP_500039.1	-0.64
embryo	papilin	MIG6	WBGene00003242	NP_505017.1	-0.90
embryogenesis	gut-specific cysteine proteinase	CPR1	WBGene00000781	NP_506002.2	0.51
cuticle	cuticlin-2	CUT2	WBGene00009983	NP_506325.1	-1.18
Metabolism					
fatty acid	palmitoyl-CoA fatty acid desaturase	FAT5	WBGene00001397	NP_507482.1	-0.67
metabolism	S-Adenosyl methionine synthase	SAMS1	WBGene00008205	NP_510002.1	-0.69
Reproduction					
egg laying	myosin light chain	MLC3	WBGene00003371	NP_741145.1	-0.68
fertility	elongation factor 1-alpha	EFT4	WBGene00001169	NP_509323.1	-0.53
Structure					
cell membrane	basement membrane proteoglycan	UNC52	WBGene00006787	NP_001022488.1	-0.94
locomotion	actin	ACT4	WBGene00000066	NP_508842.1	-0.69
Miscellaneous					

Gene Class; Function*	Gene	Abbrev.	WormBaseID	Refseq	0.2 ug/L
unknown	protein K08D12.6	Q8MXU8	WBGene00019540	NP_741325.1	-0.91
unknown	protein SSQ-4	Q23062	WBGene00006053	NP_500704.2	-0.89
unknown	protein FAAH-2	Q17447	WBGene00015048	NP_501366.1	1.37
unknown	protein SSQ-1	Q21294	WBGene00006050	NP_501768.1	-0.93
unknown	protein F41F3.3	Q20281	WBGene00018297	NP_504251.1	-0.84
unknown	protein COL-140	Q19813	WBGene00000713	NP_504525.1	-0.52
unknown	-	Q6A723	-	-	-0.57
unknown	protein COL-144	Q17458	WBGene00000717	NP_505374.3	-0.74
unknown	protein COL-159	Q20922	WBGene00000732	NP_506283.1	-0.75
unknown	protein K02E11.10	Q5FC49	WBGene00044109	NP_001024024.1	-1.56
unknown	protein SWSN-1	PSA1	WBGene00004203	NP_001256906.1	-0.76
unknown	protein C45B2.1	Q95YC6	WBGene00016658	NP_509017.1	-1.13

*Gene Class and Function were determined from gene descriptions obtained from WormBase annotations.

Table 3.3. WebGestalt functional enrichment analysis results. Functional analysis of differentially expression genes (determined by RNA-Seq) revealed that genes associated with growth, reproduction and fatty acid biosynthesis (prostaglandin precursors) were affected by IB exposure.

Function	# Genes	Overrepresentation
Positive regulation of growth rate (GO)	9	q = 0.0004
Reproduction (GO)	11	q = 0.001
Biosynthesis of unsaturated fatty acids (KEGG)	2 (FAT-5, FAT-7)	q = 0.0001

Discussion

Reduced progeny generation

In multiple repeated experiments, we observed that *C. elegans* progeny generation was reduced in IB-exposed worms at concentrations as low as 0.2 µg/L. To our knowledge, this has not been reported in literature previously. There have been very few, if any, toxicological studies of IB using *C. elegans* as the model organism. Recent studies involving IB and *C. elegans* have been focused on the potential of NSAIDs as protectors against Alzheimer's disease (Morita *et al*. 2009). In these studies, there were no reports of reproductive and/or growth anomalies in IB-exposed worms. A possible link between IB exposure and our observed impact on reproduction might be established

through prostaglandins' importance in *C. elegans* oocyte and sperm communication, which regulates the fertilization process.

The worms used in this study are hermaphrodites, with each worm producing both egg (oocyte) and sperm during self-fertilization. Fertilization occurs internally within the reproductive tract in the spermatheca (Han *et al*. 2010). Oocytes secrete signals derived from polyunsaturated fatty acids (PUFAs) that guide motile sperm to fertilization site. Examples of PUFAs include eicosapentaenoic acids and arachidonic acids (AA), which, recall from Chapter 2 (Figure 2.3), are precursors for prostaglandins (PGs) (Funk 2001). In *C. elegans*, the fatty acids used to synthesize PUFAs are obtained through their *E. coli* diet (Watts 2009). Unlike the mammalian PG synthesis pathway, *C. elegans'* do not involve cyclooxygenase (COX) enzymes, as evident by the lack of COX genes in the worm genome. There is recent evidence that *C. elegans* generate a novel series of F-series prostaglandins that regulate sperm guidance through a cyclooxygenase-independent PG synthesis mechanism (Kubagawa *et al*. 2006, Hoang *et al*. 2013). Our results show that IB-exposure is linked to reduced progeny generation. However, it is unclear whether the observed reduction is due to interference with sperm guidance through PG disruption or some other mechanism. Ibuprofen is a COX- inhibiting agent and since *C. elegans* are currently known to synthesize PGs independent of COX action, it is difficult to speculate how IB may interfere with PG production in this species. Further investigation into this issue is needed.

Gene expression changes

In a separate experiment, worms were exposed to 0.2 ug/L IB for 24 hours for subsequent RNA-Seq analysis. Although the impact on progeny count did not manifest until Day 3 of exposure in the previous experiments, the 24-hour time point was chosen for RNA-Seq analysis in order to examine the early response of *C. elegans* to IB exposure. Moreover, gene-level changes are expected to occur before protein- or higher-level impacts and the manifestation of reproductive interference. We found 26 genes to be differentially expressed in *C. elegans* after 24 hours of IB exposure, 14 of which had functional annotation available from WormBase (Table 3.2). Ten of these genes have functions related to development, growth, embryogenesis, and reproduction, with the majority of the affected genes having decreased expression levels in IB-

exposed worms. Although the gene expression fold change values are small (max = 1.18 fold change), their effects can be linked to the observed reduction in progeny generation. For example, as reported in other studies, reduced expression levels of OSM11, EIF-5A-2, OST1, and MIG6 have been linked to abnormalities in gonad structure and function, resulting in fertility defects (Komatsu *et al*. 2008, Hanazawa *et al*. 2004, Fitzgerald and Schwarzbauer 1998, Kawano *et al*. 2009). The impact of IB (or other NSAIDs) exposure on the expression of these genes has not been previously reported and this is the first time down-regulation of these genes have been linked to IB exposure. However, the mechanism through which the expression levels of these genes are altered by IB is not known and would require further investigation.

Other genes that may be associated with fertility and reproduction are Δ9 desaturases, FAT5 and FAT7, which are both observed in this study to have decreased expression in IB-exposed worms. Δ9 desaturases are lipogenic enzymes that convert saturated fatty acids into unsaturated fatty acids (i.e. PUFAs) (Brock *et al*. 2007). Recall that PUFAs are precursors to PGs, which are key signalling molecules secreted by *C. elegans* oocytes to guide sperm to the site of fertilization. Kubagawa *et al.* 2006 showed that mutant *C. elegans* lacking FAT2 or FAT3 had deficient PUFA levels and they observed defects in sperm motility and subsequent reduced progeny generation. They also reported that FAT1 and FAT4 mutants did not impact sperm motility, however they did not examine FAT5 and FAT7 mutants. FAT5 and FAT7 are involved in PUFA production upstream of FAT1, FAT2, FAT3, and FAT4 and have been reported to affect PUFA levels (Brock *et al*. 2007). The impact of IB exposure on FAT5 or FAT7 gene expression has not been previously reported. The mechanism of how IB might influence Δ9 desaturase gene expression is unknown.

At the organism-level, we observed that IB-exposed worms had lower progeny production than non-exposed worms, an impact that was consistently seen at IB concentrations as low as 0.2 µg/L. RNA-Seq data revealed that genes related to growth, gonad development and fatty acid synthesis were differentially expressed in IB-exposed worms. Considering both organism-level data and gene expression-level data together, our data suggests that IB exposure could have impeded *C. elegans* progeny generation through multiple mechanisms. There is gene-level evidence suggesting that 1) IB affects gonad structure and gonad development (decreased levels of OSM11, EIF-5A-2, OST1,

and MIG6 genes) and 2) IB affects oocyte-to-sperm communication through reducing PUFAs and PGs production (decreased FAT5 and FAT7 gene expression). Although we are unable to confirm ibuprofen's mechanism of action in *C. elegans*, we have generated new questions and hypotheses that warrant further investigation. qPCR assays should be conducted to verify the expression levels of the aforementioned genes in combination with protein assays to evaluate protein expression levels. Because *C. elegans* have transparent bodies, gonad structure and sperm motility could be accessed in real-time in future studies.

In summary, we consistently observed reduced progeny production and cumulative progeny production in *C. elegans* exposed to IB concentrations as low as 0.02 µg/L. Although nematodes are not traditionally used in environmental toxicology testing, *C. elegans* has recently emerged as a promising toxicological model (Freeman *et al*. 1999). The progeny reduction seen at environmentally relevant levels of IB reveals the impact of low dose IB on a common microscopic organism found in many ecosystems. As we move towards examining toxicological impacts on whole ecosystems, these results highlight the need to not only examine potential toxicant effects on larger organisms such as fish, but also smaller organisms such as microbes and nematodes. This study was undertaken as an exploratory investigation into the potential impacts of IB exposure on *C. elegans* and was successful in providing new hypotheses for future research and showing the benefits of examining both gene- and organism-level effects in order to link molecular responses to population or ecological impact.

Chapter 4.

Conclusion

Summary of Research Results

The aim of this research project was to explore the application of RNA-Seq in ecotoxicogenomics through investigating the effects of ibuprofen exposure on rainbow trout and *C. elegans* and linking gene-level and organism-level observations to better understand toxicant impact. Although ibuprofen is not known to be a highly acutely toxic substance (fish LC_{50} > 100 mg/L), we observed an impact on rainbow trout and *C. elegans* exposed to low environmentally relevant concentrations.

In underyearling rainbow trout, we found that ibuprofen exposure altered the expression of 125 genes in rainbow trout liver, including those involved in stress response, immune system, inflammatory response, vitellogenesis and prostaglandin biosynthesis. Using qPCR and a genetic sex determination method, we observed some sex-dependent gene expression changes. Notably, we observed a decrease in VTG gene expression in male fish but not females; however, the biological significance or implications of this observation is not yet clear. We also found that genetically female and male fish expressed statistically different basal levels of VTG gene, despite being sexually immature, an observation that demonstrates the importance of controlling for and/or considering the sex of fish when using VTG as a marker for endocrine disruption.

One of the goals of this thesis project was to explore the usage of RNA-Seq on rainbow trout in the context of ecotoxicogenomics. We demonstrated that, through the use of various open-source software programs, we were able to generate a set of rainbow trout transcripts via *de novo* assembly, conduct statistical analysis and discover a list of differentially expressed genes that were altered in response to ibuprofen exposure. When working with a species lacking a sequenced genome, it is often

necessary to explore and test a variety of software programs to find one that is suitable. Many of the software programs used in RNA-Seq requires bioinformatics support, including both computational infrastructure and some scripting/programming knowledge. Thus, a major challenge in using RNA-Seq in ecotoxicology is obtaining the necessary bioinformatics support. Although the process of learning and conducting RNA-Seq analysis is not trivial, the analysis pipeline used in this thesis is not a complex one, but it is effective. As such, we anticipate that the data analysis process used here can be adapted for use with other organisms lacking a reference genome for future studies in ecotoxicology.

In *C. elegans*, we observed that ibuprofen-exposed worms had lower progeny production than non-exposed worms, an impact that was consistently seen at ibuprofen concentrations as low as 0.2 µg/L, a dose that is below some reported environmentally detected levels. RNA-Seq data revealed that 26 genes were altered by ibuprofen exposure, of which 14 genes related to growth, gonad development and fatty acid synthesis had decreased expression levels. Considering both organism-level data and gene expression-level data together, our data suggests that ibuprofen exposure could have impeded *C. elegans* progeny generation through multiple mechanisms. Although we are unable to confirm ibuprofen's mechanism of action in *C. elegans*, we have generated new questions and hypotheses that warrant further investigation.

Future Studies

For rainbow trout experiments, due to the challenges that arose from small sample sizes coupled with the need to examine female and male fish separately, future experiments should be conducted with enough replicates so that a sufficient sample size can be obtained for both sexes. Also, the genetic sex of the test fish should be determined before subsequent downstream analysis and taken into consideration during data analysis. Since this work was conducted, a sex-determining gene (sdY) in rainbow trout has been identified (Yano *et al.* 2012, Phillips *et al.* 2013). This gene is considered to be the master sex-determining gene in rainbow trout and can be used in future work to determine the genetic sex of the test population. Based on the observed differences in basal levels of VTG gene expression between male and female juvenile rainbow trout, it

would be interesting to conduct a study to determine the age at which rainbow trout begin to express sex-dependent basal levels of VTG. Characterizing the sex-dependent changes in basal VTG levels in sexually immature rainbow trout has not been done previously and could potentially generate interesting insights. Other potential future work could include challenge studies in which fish are first exposed to IB alone and then another substance is later added to the mixture to observe whether IB exposure affects the fish's response to a secondary substance. Because we observed a decrease in VTG expression in IB-exposed male fish, an interesting challenge substance could be an estrogenic compound. A potential hypothesis is that IB may have protective activity against estrogenic effects.

Although we are unable to confirm IB's mechanism of action in *C. elegans*, we have generated new questions and hypotheses that warrant further investigation. Our observations suggest that IB exposure could have impeded *C. elegans* progeny generation through multiple mechanisms: 1) IB affects gonad structure and gonad development (decreased levels of OSM11, EIF-5A-2, OST1, and MIG6 genes) and 2) IB affects oocyte-to-sperm communication through reducing PUFAs and PGs production (decreased FAT5 and FAT7 gene expression). Before further experiments are performed, qPCR assays should be conducted to verify the expression levels of the aforementioned genes. Then new experiments can include protein assays to evaluate the corresponding protein expression levels of these genes. Because *C. elegans* have transparent bodies, gonad structure and sperm motility could be accessed in real-time in future studies.

Conclusion

The objectives of this thesis project were 1) to evaluate the feasibility of conducting RNA-Seq analysis on rainbow trout in the context of ecotoxicogenomics, 2) to explore the gene expression response of juvenile rainbow trout and *C. elegans* to ibuprofen exposure using RNA-Seq and 3) to examine the effects of ibuprofen exposure on *C. elegans* behaviour and progeny generation.

Although IB is not known to be highly toxic or lethal, we observed an impact on both juvenile rainbow trout and *C. elegans* exposed to low environmentally relevant concentrations of IB. For both species, we found that IB exposure affected the expression of genes relating to stress, prostaglandin synthesis, reproduction and development. Notably in fish, we observed sex-dependent differences in VTG and PGD synthase gene expression, emphasizing the importance of genetic sex determination of juvenile fish used in bioassays. In worms, we saw a decrease in progeny production count, which can be linked to gene expression changes relating to gonad development and oocyte-to-sperm communication. Our results suggest that ibuprofen may have negative impacts on reproduction in both species but requires further investigation.

Additionally, we have successfully demonstrated the use of RNA-Seq on rainbow trout and described a data analysis pipeline that could be adapted for other species used in ecotoxicogenomics research. Our observations highlight the benefits of studying the effects of environmental toxicants not only on traditionally used species such as rainbow trout, but also microscopic organisms such as *C. elegans*, to better access toxicological impact on whole ecosystems.

References

Anders, S., & Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biology, 11*(10), R106-2010-11-10-r106. Epub 2010 Oct 27. doi: 10.1186/gb-2010-11-10-r106

Ankley, G. T., Daston, G. P., Degitz, S. J., Denslow, N. D., Hoke, R. A., Kennedy, S. W., . . . Versteeg, D. (2006). Toxicogenomics in regulatory ecotoxicology. *Environmental Science & Technology, 40*(13), 4055-4065.

Arukwe, A., & Goksoyr, A. (2003). Eggshell and egg yolk proteins in fish: Hepatic proteins for the next generation: Oogenetic, population, and evolutionary implications of endocrine disruption. *Comparative Hepatology, 2*(1), 4.

Arukwe, A., Knudsen, F. R., & Goksoyr, A. (1997). Fish zona radiata (eggshell) protein: A sensitive biomarker for environmental estrogens. *Environmental Health Perspectives, 105*(4), 418-422.

Ashton, D., Hilton, M., & Thomas, K. V. (2004). Investigating the environmental transport of human pharmaceuticals to streams in the united kingdom. *The Science of the Total Environment, 333*(1-3), 167-184. doi: 10.1016/j.scitotenv.2004.04.062

Bendz, D., Paxeus, N. A., Ginn, T. R., & Loge, F. J. (2005). Occurrence and fate of pharmaceutically active compounds in the environment, a case study: Hoje river in sweden. *Journal of Hazardous Materials, 122*(3), 195-204. doi: 10.1016/j.jhazmat.2005.03.012

Bhandari, K., & Venables, B. (2011). Ibuprofen bioconcentration and prostaglandin E2 levels in the bluntnose minnow pimephales notatus. *Comparative Biochemistry and Physiology, Toxicology & Pharmacology : CBP, 153*(2), 251-257. doi: 10.1016/j.cbpc.2010.11.004

Brock, T. J., Browse, J., & Watts, J. L. (2007). Fatty acid desaturation and the regulation of adiposity in Caenorhabditis elegans. *Genetics, 176*(2), 865-875. doi: 10.1534/genetics.107.071860

Brown, J. N., Paxeus, N., Forlin, L., Larsson, D. G. J. (2007) Variations in bioconcentration of human pharmaceuticals from sewage effluents into fish blood plasma. *Environmental Toxicology and Pharmacology, 24*(3), 267-274.

Brueggemeier, R. W., Diaz-Cruz, E. S., Li, P. K., Sugimoto, Y., Lin, Y. C., & Shapiro, C. L. (2005). Translational studies on aromatase, cyclooxygenases, and enzyme inhibitors in breast cancer. *The Journal of Steroid Biochemistry and Molecular Biology, 95*(1-5), 129-136. doi: 10.1016/j.jsbmb.2005.04.013

Brunelli, J. P., Wertzler, K. J., Sundin, K., & Thorgaard, G. H. (2008). Y-specific sequences and polymorphisms in rainbow trout and Chinook salmon. *Genome / National Research Council Canada = Genome / Conseil National De Recherches Canada, 51*(9), 739-748. doi: 10.1139/G08-060

Buser, H., Poiger, T., & Müller, M. D. (1999). Occurrence and environmental behavior of the chiral pharmaceutical drug ibuprofen in surface waters and in wastewater. *Environmental Science & Technology, 33*(15), 2529-2535.

C. elegans Sequencing Consortium. (1998). Genome sequence of the nematode C. elegans: A platform for investigating biology. *Science (New York, N.Y.), 282*(5396), 2012-2018.

Campbell, P. M., Pottinger, T. G., & Sumpter, J. P. (1994). Changes in the affinity of estrogen and androgen receptors accompany changes in receptor abundance in brown and rainbow trout. *General and Comparative Endocrinology, 94*, 329-340.

Carlson, D. B., & Williams, D. E. (1999). Sex-specific vitellogenin production in immature rainbow trout. *Environmental Toxicology and Chemistry, 18*(10), 2361-2363.

Casals-Casas, C., & Desvergne, B. (2011). Endocrine disruptors: From endocrine to metabolic disruption. *Annual Review of Physiology, 73*, 135-162. doi: 10.1146/annurev-physiol-012110-142200

Castellsague, J., Riera-Guardia, N., Calingaert, B., Varas-Lorenzo, C., Fourrier-Reglat, A., Nicotra, F., . . . Safety of Non-Steroidal Anti-Inflammatory Drugs (SOS) Project. (2012). Individual NSAIDs and upper gastrointestinal complications: A systematic review and meta-analysis of observational studies (the SOS project). *Drug Safety : An International Journal of Medical Toxicology and Drug Experience, 35*(12), 1127-1146. doi: 10.2165/11633470-000000000-00000

Cha, Y. I., Solnica-Krezel, L., & DuBois, R. N. (2006). Fishing for prostanoids: Deciphering the developmental functions of cyclooxygenase-derived prostaglandins. *Developmental Biology, 289*(2), 263-272. doi: 10.1016/j.ydbio.2005.10.013

Colborn, T., vom Saal, F. S., & Soto, A. M. (1993). Developmental effects of endocrine-disrupting chemicals in wildlife and humans. *Environmental Health Perspectives, 101*(5), 378-384.

Contardo-Jara, V., Lorenz, C., Pflugmacher, S., Nutzmann, G., Kloas, W., & Wiegand, C. (2011). Exposure to human pharmaceuticals carbamazepine, ibuprofen and bezafibrate causes molecular effects in dreissena polymorpha. *Aquatic Toxicology (Amsterdam, Netherlands), 105*(3-4), 428-437. doi: 10.1016/j.aquatox.2011.07.017

Coppe, A., Pujolar, J. M., Maes, G. E., Larsen, P. F., Hansen, M. M., Bernatchez, L., . . . Bortoluzzi, S. (2010). Sequencing, de novo annotation and analysis of the first anguilla anguilla transcriptome: EeelBase opens new perspectives for the study of the critically endangered European eel. *BMC Genomics, 11*, 635-2164-11-635. doi: 10.1186/1471-2164-11-635

Costa, V., Angelini, C., De Feis, I., & Ciccodicola, A. (2010). Uncovering the complexity of transcriptomes with RNA-seq. *Journal of Biomedicine & Biotechnology, 2010*, 853916. doi: 10.1155/2010/853916

Farre, M. L., Ferrer, I., Ginebreda, A., Figueras, M., Olivella, L., Tirapu, L., . . . Barcelo, D. (2001). Determination of drugs in surface water and wastewater samples by liquid chromatography-mass spectrometry: Methods and preliminary results including toxicity studies with vibrio fischeri. *Journal of Chromatography.A, 938*(1-2), 187-197.

Fent, K., Weston, A. A., & Caminada, D. (2006). Ecotoxicology of human pharmaceuticals. *Aquatic Toxicology (Amsterdam, Netherlands), 76*(2), 122-159. doi: 10.1016/j.aquatox.2005.09.009

Ferrando-Climent, L., Collado, N., Buttiglieri, G., Gros, M., Rodriguez-Roda, I., Rodriguez-Mozaz, S., & Barcelo, D. (2012). Comprehensive study of ibuprofen and its metabolites in activated sludge batch experiments and aquatic environment. *The Science of the Total Environment, 438*, 404-413. doi: 10.1016/j.scitotenv.2012.08.073

Filby, A. L., & Tyler, C. R. (2007). Appropriate 'housekeeping' genes for use in expression profiling the effects of environmental estrogens in fish. *BMC Molecular Biology, 8*, 10. doi: 10.1186/1471-2199-8-10

Fitzgerald, M. C., & Schwarzbauer, J. E. (1998). Importance of the basement membrane protein SPARC for viability and fertility in Caenorhabditis elegans. *Current Biology : CB, 8*(23), 1285-1288.

Flippin, J. L., Huggett, D., & Foran, C. M. (2007). Changes in the timing of reproduction following chronic exposure to ibuprofen in Japanese medaka, Oryzias latipes. *Aquatic Toxicology, 81*(1), 73-78. doi: 10.1016/j.aquatox.2006.11.002

Freeman, M. N., Peredney, C. L., & Williams, P. L. (1999). A soil bioassay using the Nematode *caenorhabditis elegans*. In D. S. Henshel, M. C. Black & M. C. Harrass (Eds.), *Environmental toxicology and risk assessment: Standardization of biomarkers for endocrine disruption and environmental assessment: Eighth volume, ASTM STP 1364* (pp. 305-318). West Conshohocken, PA: American Society for Testing and Materials.

Frye, C. A., Bo, E., Calamandrei, G., Calza, L., Dessi-Fulgheri, F., Fernandez, M., . . . Panzica, G. C. (2012). Endocrine disrupters: A review of some sources, effects, and mechanisms of actions on behaviour and neuroendocrine systems. *Journal of Neuroendocrinology, 24*(1), 144-159. doi: 10.1111/j.1365-2826.2011.02229.x

Funk, C. D. (2001). Prostaglandins and leukotrienes: Advances in eicosanoid biology. *Science (New York, N.Y.), 294*(5548), 1871-1875. doi: 10.1126/science.294.5548.1871

Gagne, F., Blaise, C., & Andre, C. (2006). Occurrence of pharmaceutical products in a municipal effluent and toxicity to rainbow trout (Oncorhynchus mykiss) hepatocytes. *Ecotoxicology and Environmental Safety, 64*(3), 329-336. doi: 10.1016/j.ecoenv.2005.04.004

Garber, M., Grabherr, M.G., Guttman, M., Trapnell, C. (2011) Computational methods for transcriptome annotation and quantification using RNA-Seq. *Nature Methods, 8*(6), 469-477. doi: 10.1038/NMETH.1613

Garcia, T. I., Shen, Y., Crawford, D., Oleksiak, M. F., Whitehead, A., & Walter, R. B. (2012). RNA-Seq reveals complex genetic response to deepwater horizon oil release in Fundulus grandis. *BMC Genomics, 13*, 474-2164-13-474. doi:10.1186/1471-2164-13-474

Gonzalez-Rey, M., & Bebianno, M. J. (2012). Does non-steroidal anti-inflammatory (NSAID) ibuprofen induce antioxidant stress and endocrine disruption in mussel metals galloprovincialis? *Environmental Toxicology and Pharmacology, 33*(2), 361-371. doi: 10.1016/j.etap.2011.12.017

Gravel, A., & Vijayan, M. M. (2007). Non-steroidal anti-inflammatory drugs disrupt the heat shock response in rainbow trout. *Aquatic Toxicology (Amsterdam, Netherlands), 81*(2), 197-206. doi: 10.1016/j.aquatox.2006.12.001

Gravel, A., Wilson, J. M., Pedro, D. F., & Vijayan, M. M. (2009). Non-steroidal anti-inflammatory drugs disturb the osmoregulatory, metabolic and cortisol responses associated with seawater exposure in rainbow trout. *Comparative Biochemistry and Physiology, Toxicology & Pharmacology : CBP, 149*(4), 481-490. doi: 10.1016/j.cbpc.2008.11.002

Han, S., Choi, K., Kim, J., Ji, K., Kim, S., Ahn, B., . . . Giesy, J. P. (2010). Endocrine disruption and consequences of chronic exposure to ibuprofen in Japanese medaka (Oryzias latipes) and freshwater cladocerans Daphnia magna and Moina macrocopa. *Aquatic Toxicology (Amsterdam, Netherlands), 98*(3), 256-264. doi: 10.1016/j.aquatox.2010.02.013

Han, S. M., Cottee, P. A., & Miller, M. A. (2010). Sperm and oocyte communication mechanisms controlling C. elegans fertility. *Developmental Dynamics : An Official Publication of the American Association of Anatomists, 239*(5), 1265-1281. doi: 10.1002/dvdy.22202

Hanazawa, M., Kawasaki, I., Kunitomo, H., Gengyo-Ando, K., Bennett, K. L., Mitani, S., & Iino, Y. (2004). The Caenorhabditis elegans eukaryotic initiation factor 5A homologue, IFF-1, is required for germ cell proliferation, gametogenesis and localization of the P-granule component PGL-1. *Mechanisms of Development, 121*(3), 213-224. doi: 10.1016/j.mod.2004.02.001

Huang, Q., Dong, S., Fang, C., Wu, X., Ye, T., & Lin, Y. (2012). Deep sequencing-based transcriptome profiling analysis of *Oryzias melastigma* exposed to PFOS. *Aquatic Toxicology, 120-121*, 54-58. doi:10.1016/j.aquatox.2012.04.013

Heberer, T. (2002). Occurrence, fate, and removal of pharmaceutical residues in the aquatic environment: A review of recent research data. *Toxicology Letters, 131*(1-2), 5-17.

Heckmann, L. H., Callaghan, A., Hooper, H. L., Connon, R., Hutchinson, T. H., Maund, S. J., & Sibly, R. M. (2007). Chronic toxicity of ibuprofen to daphnia magna: Effects on life history traits and population dynamics. *Toxicology Letters, 172*(3), 137-145. doi: 10.1016/j.toxlet.2007.06.001

Heckmann, L. H., Sibly, R. M., Connon, R., Hooper, H. L., Hutchinson, T. H., Maund, S. J., . . . Callaghan, A. (2008). Systems biology meets stress ecology: Linking molecular and organismal stress responses in daphnia magna. *Genome Biology, 9*(2), R40-2008-9-2-r40. Epub 2008 Feb 21. doi: 10.1186/gb-2008-9-2-r40

Hiller-Sturmhofel, S., & Bartke, A. (1998). The endocrine system: An overview. *Alcohol Health and Research World, 22*(3), 153-164.

Hoang, H. D., Prasain, J. K., Dorand, D., & Miller, M. A. (2013). A heterogeneous mixture of F-series prostaglandins promotes sperm guidance in the Caenorhabditis elegans reproductive tract. *PLoS Genetics, 9*(1), e1003271. doi: 10.1371/journal.pgen.1003271

Hornett, E. A., & Wheat, C. W. (2012). Quantitative RNA-Seq analysis in non-model species: Assessing transcriptome assemblies as a scaffold and the utility of evolutionary divergent genomic reference species. *BMC Genomics, 13*, 361-2164-13-361. doi: 10.1186/1471-2164-13-361

Huggett, D. B., Ericson, J. F., Cook, J. C. , and R. T. Williams, R. T. (2004). Plasma concentrations of human pharmaceuticals as predictors of pharmacological responses in fish. *Pharmaceuticals in the environment - Sources, fate, effects and risks*, edited by K. Kümmerer. Berlin, Germany: Springer.

Jayapal, M. (2012). Integration of next-generation sequencing based multi-omics approaches in toxicogenomics. *Frontiers in Genetics, 3*, 88. doi: 10.3389/fgene.2012.00088

Jobling, S., Burn, R. W., Thorpe, K., Williams, R., & Tyler, C. (2009). Statistical modeling suggests that antiandrogens in effluents from wastewater treatment works contribute to widespread sexual disruption in fish living in English rivers. *Environmental Health Perspectives, 117*(5), 797-802. doi: 10.1289/ehp.0800197

Kaufmann, S. H. E. (1990). Heat shock proteins and the immune response. *Immunology Today, 11*(0), 129-136. doi: 10.1016/0167-5699(90)90050-J

Kawano, T., Zheng, H., Merz, D. C., Kohara, Y., Tamai, K. K., Nishiwaki, K., & Culotti, J. G. (2009). C. elegans mig-6 encodes papilin isoforms that affect distinct aspects of DTC migration, and interacts genetically with mig-17 and collagen IV. *Development (Cambridge, England), 136*(9), 1433-1442. doi: 10.1242/dev.028472

Kiang, J. G., & Tsokos, G. C. (1998). Heat shock protein 70 kDa: Molecular biology, biochemistry, and physiology. *Pharmacology & Therapeutics, 80*(2), 183-201.

Komatsu, H., Chao, M. Y., Larkins-Ford, J., Corkins, M. E., Somers, G. A., Tucey, T., . . . Hart, A. C. (2008). OSM-11 facilitates LIN-12 notch signaling during Caenorhabditis elegans vulval development. *PLoS Biology, 6*(8), e196. doi: 10.1371/journal.pbio.0060196

Kubagawa, H. M., Watts, J. L., Corrigan, C., Edmonds, J. W., Sztul, E., Browse, J., & Miller, M. A. (2006). Oocyte signals derived from polyunsaturated fatty acids control sperm recruitment in vivo. *Nature Cell Biology, 8*(10), 1143-1148. doi: 10.1038/ncb1476

Lagunas, L., Bradbury, C. M., Laszlo, A., Hunt, C. R., & Gius, D. (2004). Indomethacin and ibuprofen induce Hsc70 nuclear localization and activation of the heat shock response in HeLa cells. *Biochemical and Biophysical Research Communications, 313*(4), 863-870.

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with bowtie 2. *Nature Methods, 9*(4), 357-359. doi: 10.1038/nmeth.1923

Leung, M. C., Williams, P. L., Benedetto, A., Au, C., Helmcke, K. J., Aschner, M., & Meyer, J. N. (2008). Caenorhabditis elegans: An emerging model in biomedical and environmental toxicology. *Toxicological Sciences : An Official Journal of the Society of Toxicology, 106*(1), 5-28. doi: 10.1093/toxsci/kfn121

Levin, S. W., Butler, J. D., Schumacher, U. K., Wightman, P. D., & Mukherjee, A. B. (1986). Uteroglobin inhibits phospholipase A2 activity. *Life Sciences, 38*(20), 1813-1819.

Levinson, W., Oppermann, H., & Jackson, J. (1980). Transition series metals and sulfhydryl reagents induce the synthesis of four proteins in eukaryotic cells. *Biochimica Et Biophysica Acta, 606*(1), 170-180.

Lin, A. Y., & Reinhard, M. (2005). Photodegradation of common environmental pharmaceuticals and estrogens in river water. *Environmental Toxicology and Chemistry, 24*(6), 1303-1309.

Lister, A. L., & Van Der Kraak, G. (2008). An investigation into the role of prostaglandins in zebrafish oocyte maturation and ovulation. *General and Comparative Endocrinology, 159*(1), 46-57. doi: 10.1016/j.ygcen.2008.07.017

Matozzo, V., Rova, S., & Marin, M. G. (2012). The nonsteroidal anti-inflammatory drug, ibuprofen, affects the immune parameters in the clam Ruditapes philippinarum. *Marine Environmental Research, 79*, 116-121. doi: 10.1016/j.marenvres.2012.06.003

Matthews, J., & Gustafsson, J. A. (2006). Estrogen receptor and aryl hydrocarbon receptor signaling pathways. *Nuclear Receptor Signaling, 4*, e016. doi: 10.1621/nrs.04016

Mehinto, A. C., Martyniuk, C. J., Spade, D. J., & Denslow, N. D. (2012). Applications for next-generation sequencing in fish ecotoxicogenomics. *Frontiers in Genetics, 3*, 62. doi: 10.3389/fgene.2012.00062

Metcalfe, C. D., Miao, X., Koenig, B. G., & Struger, J. (2003). Distribution of acidic and neutral drugs in surface waters near sewage treatment plants in the lower great lakes, Canada. *Environmental Toxicology and Chemistry, 22*(12), 2881-2889. doi: 10.1897/02-627

Metzker, M. L. (2010). Sequencing technologies - the next generation. *Nature Reviews. Genetics, 11*(1), 31-46. doi: 10.1038/nrg2626

Milan, M., Pauletto, M., Patarnello, T., Bargelloni, L., Marin, M. G., & Matozzo, V. (2013). Gene transcription and biomarker responses in the clam Ruditapes philippinarum after exposure to ibuprofen. *Aquatic Toxicology (Amsterdam, Netherlands), 126*, 17-29. doi: 10.1016/j.aquatox.2012.10.007

Morita, M., Osoda, K., Yamazaki, M., Shirai, F., Matsuoka, N., Arakawa, H., & Nishimura, S. (2009). Effects of non-steroidal anti-inflammatory drugs on abeta deposition in abeta(1-42) transgenic C. elegans. *Brain Research, 1295*, 186-191. doi: 10.1016/j.brainres.2009.08.002

Mortensen, A. S., & Arukwe, A. (2007). Interactions between estrogen- and ah-receptor signalling pathways in primary culture of salmon hepatocytes exposed to nonylphenol and 3,3',4,4'-tetrachlorobiphenyl (congener 77). *Comparative Hepatology, 6*, 2. doi: 10.1186/1476-5926-6-2

Morthorst, J. E., Lister, A., Bjerregaard, P., & Van Der Kraak, G. (2013). Ibuprofen reduces zebrafish PGE(2) levels but steroid hormone levels and reproductive parameters are not affected. *Comparative Biochemistry and Physiology, Toxicology & Pharmacology : CBP, 157*(2), 251-257. doi: 10.1016/j.cbpc.2012.12.001

Naderi, M., Zargham, D., Asadi, A., Bashti, T., & Kamayi, K. (2013). Short-term responses of selected endocrine parameters in juvenile rainbow trout (Oncorhynchus mykiss) exposed to 4-nonylphenol. *Toxicology and Industrial Health,* doi: 10.1177/0748233713491806

Nallani, G. C., Paulos, P. M., Constantine, L. A., Venables, B. J., Huggett, D. B. (2011) Bioconcentration of ibuprofen in fathead minnow (*Pimephales promela*s) and channel catfish (*Ictalurus punctatus*). *Chemosphere, 84*(10), 1371-1377

Nuwaysir, E. F., Bittner, M., Trent, J., Barrett, J. C., & Afshari, C. A. (1999). Microarrays and toxicology: The advent of toxicogenomics. *Molecular Carcinogenesis, 24*(3), 153-159.

Osachoff, H. L., Shelley, L. K., Furtula, V., van Aggelen, G. C., & Kennedy, C. J. (2013). Induction and recovery of estrogenic effects after short-term 17beta-estradiol exposure in juvenile rainbow trout (Oncorhynchus mykiss). *Archives of Environmental Contamination and Toxicology, 65*(2), 276-285. doi: 10.1007/s00244-013-9890-8

Oshlack, A. and Wakefield, M.J. (2009). Transcript length bias in RNA-Seq data confounds systems biology. *Biology Direct* 4, 14. doi: 10.1186/1745-6150-4-14

Palstra, A. P., Beltran, S., Burgerhout, E., Brittijn, S. A., Magnoni, L. J., Henkel, C. V., . . . Planas, J. V. (2013). Deep RNA sequencing of the skeletal muscle transcriptome in swimming fish. *PloS One, 8*(1), e53171. doi: 10.1371/journal.pone.0053171

Parolini, M., Binelli, A., & Provini, A. (2011). Chronic effects induced by ibuprofen on the freshwater bivalve dreissena polymorpha. *Ecotoxicology and Environmental Safety, 74*(6), 1586-1594. doi: 10.1016/j.ecoenv.2011.04.025

Phillips, R. B., DeKoning J. J., Brunelli, J. P., Faber-Hammond, J. J., Hansen, J. D., Christensen, K. A., Renn, S. C. P., Thorgaard, G. H. (2013) Characterization of the OmyY1 Region on the Rainbow Trout Y Chromosome. *International Journal of Genomics 2013,* doi: 10.1155/2013/261730

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics (Oxford, England), 26*(6), 841-842. doi: 10.1093/bioinformatics/btq033

Ragugnetti, M., Adams, M. L., Guimarães, A. T. B., Sponchiado, G., de Vasconcelos, E. C., & de Oliveira, C. M. R. (2011). Ibuprofen genotoxicity in aquatic environment: An experimental model using Oreochromis niloticus. *Water Air and Soil Pollution, 218*, 361.

Rainsford, K. D. (2009). Ibuprofen: Pharmacology, efficacy and safety. *Inflammopharmacology, 17*(6), 275-342. doi: 10.1007/s10787-009-0016-x

Rao, P. N. P., & Knaus, E. E. (2008). Evolution of nonsteroidal anti-inflammatory drugs (NSAIDs): Cyclooxygenase (COX) inhibition and beyond. *Journal of Pharmacy & Pharmaceutical Sciences : A Publication of the Canadian Society for Pharmaceutical Sciences, Societe Canadienne Des Sciences Pharmaceutiques, 11*(2), 81s-110s.

Rau, A., Gallopin, M., Celeux, G., & Jaffrezic, F. (2013). Data-based filtering for replicated high-throughput transcriptome sequencing experiments. *Bioinformatics (Oxford, England), 29*(17), 2146-2152. doi: 10.1093/bioinformatics/btt350

Fisheries Act, c. F-14 (1985).

Rapaport, F., Khanin, R., Liang, Y., Pirun, M., Krek, A., Zumbo, P., Mason, C.E., Socci, N.D., Betel, D. (2013) Comprehensive evaluation of differential gene expression analysis methods for RNA-Seq data. *Genome Biology*, *14*(9), R95. doi: 10.1186/gb-2013-14-9-r95

Richards, J. A., & Brueggemeier, R. W. (2003). Prostaglandin E2 regulates aromatase activity and expression in human adipose stromal cells via two distinct receptor subtypes. *The Journal of Clinical Endocrinology and Metabolism, 88*(6), 2810-2816.

Riddle, D. L., Blumenthal, T., Meyer, B. J., & Priess, J. R. (1997). Section I, the biological model. In D. L. Riddle, T. Blumenthal, B. J. Meyer & J. R. Priess (Eds.), *C. elegans II* (2nd edition ed.) Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press.

Roberts, A., Pimentel, H., Trapnell, C., & Pachter, L. (2011). Identification of novel transcripts in annotated genomes using RNA-seq. *Bioinformatics (Oxford, England), 27*(17), 2325-2329. doi: 10.1093/bioinformatics/btr355

Roberts, A., Trapnell, C., Donaghey, J., Rinn, J.L., Pachter, L. (2011). Improving RNA-Seq expression estimates by correcting for fragment bias. *Genome Biology, 12*(3), R22. doi:10.1186/gb-2011-12-3-r22

Robertson, G., Schein, J., Chiu, R., Corbett, R., Field, M., Jackman, S. D., . . . Birol, I. (2010). De novo assembly and analysis of RNA-Seq data. *Nature Methods, 7*(11), 909-912. doi: 10.1038/nmeth.1517

Rowley, A. F., Vogan, C. L., Taylor, G. W., & Clare, A. S. (2005). Prostaglandins in non-insectan invertebrates: Recent insights and unsolved problems. *The Journal of Experimental Biology, 208*(Pt 1), 3-14. doi: 10.1242/jeb.01275

Safe, S., & Wormke, M. (2003). Inhibitory aryl hydrocarbon receptor-estrogen receptor alpha cross-talk and mechanisms of action. *Chemical Research in Toxicology, 16*(7), 807-816. doi: 10.1021/tx034036r

Salem, M., Rexroad, C. E.,3rd, Wang, J., Thorgaard, G. H., & Yao, J. (2010). Characterization of the rainbow trout transcriptome using sanger and 454-pyrosequencing approaches. *BMC Genomics, 11*, 564-2164-11-564. doi: 10.1186/1471-2164-11-564

Sánchez, C. C., Weber, G. M., Gao, G., Cleveland, B. M., Yao, J., & Rexroad, C. E. I. (2011). Generation of a reference transcriptome for evaluating rainbow trout responses to various stressors. *BMC Genomics, 12*, 626. doi: 10.1186/1471-2164-12-626

Santos, L. H., Araujo, A. N., Fachini, A., Pena, A., Delerue-Matos, C., & Montenegro, M. C. (2010). Ecotoxicological aspects related to the presence of pharmaceuticals in the aquatic environment. *Journal of Hazardous Materials, 175*(1-3), 45-95. doi:10.1016/j.jhazmat.2009.10.100

Saravanan, M., Devi, K. U., Malarvizhi, A., & Ramesh, M. (2012). Effects of ibuprofen on hematological, biochemical and enzymological parameters of blood in an Indian major carp, Cirrhinus mrigala. *Environmental Toxicology and Pharmacology, 34*(1), 14-22. doi: 10.1016/j.etap.2012.02.005

Simon, D. F., Domingos, R. F., Hauser, C., Hutchins, C. M., Zerges, W., & Wilkinson, K. J. (2013). Transcriptome sequencing (RNA-seq) analysis of the effects of metal nanoparticle exposure on the transcriptome of Chlamydomonas reinhardtii. *Applied and Environmental Microbiology, 79*(16), 4774-4785. doi:10.1128/AEM.00998-13

Scholz, S., & Mayer, I. (2008). Molecular biomarkers of endocrine disruption in small model fish. *Molecular and Cellular Endocrinology, 293*(1-2), 57-70. doi: 10.1016/j.mce.2008.06.008

Scholz, S., Renner, P., Belanger, S. E., Busquet, F., Davi, R., Demeneix, B. A., . . . Embry, M. R. (2013). Alternatives to in vivo tests to detect endocrine disrupting chemicals (EDCs) in fish and amphibians--screening for estrogen, androgen and thyroid hormone disruption. *Critical Reviews in Toxicology, 43*(1), 45-72. doi: 10.3109/10408444.2012.737762

Schug, T. T., Janesick, A., Blumberg, B., & Heindel, J. J. (2011). Endocrine disrupting chemicals and disease susceptibility. *The Journal of Steroid Biochemistry and Molecular Biology, 127*(3-5), 204-215. doi: 10.1016/j.jsbmb.2011.08.007

Simpson, J. T., Wong, K., Jackman, S. D., Schein, J. E., Jones, S. J., & Birol, I. (2009). ABySS: A parallel assembler for short read sequence data. *Genome Research, 19*(6), 1117-1123. doi: 10.1101/gr.089532.108

Stasinakis, A. S., Thomaidis, N. S., Arvaniti, O. S., Asimakopoulos, A. G., Samaras, V. G., Ajibola, A., . . . Lekkas, T. D. (2013). Contribution of primary and secondary treatment on the removal of benzothiazoles, benzotriazoles, endocrine disruptors, pharmaceuticals and perfluorinated compounds in a sewage treatment plant. *The Science of the Total Environment, 463-464*, 1067-1075. doi: 10.1016/j.scitotenv.2013.06.087

Subbaramaiah, K., Hudis, C. A., & Dannenberg, A. J. (2011). The prostaglandin transporter regulates adipogenesis and aromatase transcription. *Cancer Prevention Research (Philadelphia, Pa.), 4*(2), 194-206. doi: 10.1158/1940-6207.CAPR-10-0367

Sumpter, J. P., & Jobling, S. (1995). Vitellogenesis as a biomarker for estrogenic contamination of the aquatic environment. *Environmental Health Perspectives, 103*(Suppl. 7), 173-178.

Tixier, C., Singer, H. P., Oellers, S., & Muller, S. R. (2003). Occurrence and fate of carbamazepine, clofibric acid, diclofenac, ibuprofen, ketoprofen, and naproxen in surface waters. *Environmental Science & Technology, 37*(6), 1061-1068.

Trapnell, C., Pachter, L., & Salzberg, S. L. (2009). TopHat: Discovering splice junctions with RNA-seq. *Bioinformatics (Oxford, England), 25*(9), 1105-1111. doi: 10.1093/bioinformatics/btp120

Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform during cell differentiation. *Nature Biotechnology, 28*(5), 511-515. doi: 10.1038/nbt.1621

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., ... Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and cufflinks. *Nature Protocols, 7*(3), 562-578. doi: 10.1038/nprot.2012.016

Tyler, C.R., Sumpter, J.P., Campbell, P.M. (1991) Uptake of vitellogenin into oocytes during early vitellogenic development in the rainbow trout, *Oncorhynchus mykiss* (Walbaum). *Journal of Fish Biology, 38*(5), 681-689. doi: 10.1111/j.1095-8649.1991.tb03157.x

Tyler, C. R., Jobling, S., & Sumpter, J. P. (1998). Endocrine disruption in wildlife: A critical review of the evidence. *Critical Reviews in Toxicology, 28*(4), 319-361. doi: 10.1080/10408449891344236

Van Aggelen, G., Ankley, G. T., Baldwin, W. S., Bearden, D. W., Benson, W. H., Chipman, J. K., ... Yu, L. (2010). Integrating omic technologies into aquatic ecological risk assessment and environmental monitoring: Hurdles, achievements, and future outlook. *Environmental Health Perspectives, 118*(1), 1-5. doi: 10.1289/ehp.0900985

van Anholt, R. D., Spanings, T., Koven, W., & Wendelaar Bonga, S. E. (2003). Effects of acetylsalicylic acid treatment on thyroid hormones, prolactins, and the stress response of tilapia (Oreochromis mossambicus). *American Journal of Physiology. Regulatory, Integrative and Comparative Physiology, 285*(5), R1098-106. doi: 10.1152/ajpregu.00731.2002

van Hecken, A., Schwartz, J. I., Depré, M., de Lepeleire, I., Dallob, A., Tanaka, W., . . . de Schepper, P. J. (2000). Comparative inhibitory activity of rofecoxib, meloxicam, diclofenac, ibuprofen, and naproxen on COX-2 versus COX-1 in healthy volunteers. *Journal of Clinical Pharmacology, 40*(10), 1109-1120. doi: 10.1177/009127000004001005

Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A., & Speleman, F. (2002). Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biology, 3*(7), RESEARCH0034.

Vane, J. R., & Botting, R. M. (1998). Anti-inflammatory drugs and their mechanism of action. Inflammation Research : Official Journal of the European Histamine Research Society ...[Et Al.], 47 Suppl 2, S78-87.

Verenitch, S. S., Lowe, C. J., & Mazumder, A. (2006). Determination of acidic drugs and caffeine in municipal wastewaters and receiving waters by gas chromatography-ion trap tandem mass spectrometry. *Journal of Chromatography.A, 1116*(1-2), 193-203. doi: 10.1016/j.chroma.2006.03.005

Vijay, N., Poelstra, J. W., Kunstner, A., & Wolf, J. B. (2013). Challenges and strategies in transcriptome assembly and differential gene expression quantification. A comprehensive in silico assessment of RNA-seq experiments. *Molecular Ecology, 22*(3), 620-634. doi: 10.1111/mec.12014

Wahli, W., Dawid, I. B., Ryffel, G. U., & Weber, R. (1981). Vitellogenesis and the vitellogenin gene family. *Science, 212*(4492), 298-304. doi: 10.1126/science.7209528

Wang, J., Gardinali P. R. (2013) Uptake and depuration of pharmaceuticals in reclaimed water by mosquito fish (Gambusia holbrooki): a worst-case, multipole-exposure scenario. *Environmental Toxicology and Chemistry, 32*(8), 1752-1758.

Wang, Z., Gerstein, M., & Snyder, M. (2009). RNA-seq: A revolutionary tool for transcriptomics. *Nature Reviews.Genetics, 10*(1), 57-63. doi: 10.1038/nrg2484

Wang, R., Li, C., Stoeckel, J., Moyer, G., Liu, Z., & Peatman, E. (2012). Rapid development of molecular resources for a freshwater mussel, *Villosa lienosa*, using an RNA-seq-based approach. *Freshwater Science, 31*(3), 695-708. doi:10.1899/11-149.1

Wang, J., Duncan, D., Shi, Z., & Zhang, B. (2013). WEB-based GEne SeT AnaLysis toolkit (WebGestalt): Update 2013. *Nucleic Acids Research, 41*(Web Server issue), W77-83. doi:10.1093/nar/gkt439

Watts, J. L. (2009). Fat synthesis and adiposity regulation in Caenorhabditis elegans. *Trends in Endocrinology and Metabolism: TEM, 20*(2), 58-65. doi: 10.1016/j.tem.2008.11.002

Williams, P. L., & Dusenbery, D. B. (1990). Aquatic toxicity testing using the nematode, caenorhabditis elegans. *Environmental Toxicology and Chemistry, 9*(10), 1285-1290. doi: 10.1002/etc.5620091007

Wiseman, S. B., He, Y., Gamal-El Din, M., Martin, J. W., Jones, P. D., Hecker, M., & Giesy, J. P. (2013). Transcriptional responses of male fathead minnows exposed to oil sands process-affected water. *Comparative Biochemistry and Physiology, Toxicology & Pharmacology : CBP, 157*(2), 227-235. doi:10.1016/j.cbpc.2012.12.002

World Health Organization (WHO). (2002). Global assessment of the state-of-the-science of endocrine disruptors. Retrieved 10/01, 2013, from http://www.who.int/ipcs/publications/new_issues/endocrine_disruptors/en/

World Health Organization (WHO). (March 2011). WHO model list of essential medicines - 17th list. Retrieved from http://www.who.int/medicines/publications/essentialmedicines/en/index.html

Yano, A., Guyomard, R., Nicol, B., Jouanno, E., Quillet, E., Klopp, C., Cabau, C., Bouchez, O., Fostier, A., Guiguen, Y. (2012). An Immune-Related Gene Evolved into the Master Sex-Determining Gene in Rainbow Trout, *Oncorhynchus mykiss*. *Current Biology, 22*, 1423-1428