Transcriptomic analysis of short-term 17A-ethynylestradiol exposure in two Californian sentinel fish species sardine (Sardinops sagax) and mackerel (Scomber japonicus)

Renaud, L., Agarwal, N., Richards, D. J., Falcinelli, S., Hazard, E. S., Carnevali, O., Hyde, J., & Hardiman, G. (2019). Transcriptomic analysis of short-term 17A-ethynylestradiol exposure in two Californian sentinel fish species sardine (Sardinops sagax) and mackerel (Scomber japonicus). *Environmental Pollution*, 926-937. https://doi.org/10.1016/j.envpol.2018.10.058

Published in:
Environmental Pollution

Document Version:
Peer reviewed version

Transcriptomic analysis of short-term 17α-ethynylestradiol exposure in two Californian sentinel fish species sardine (Sardinops sagax) and mackerel (Scomber japonicus)

Ludivine Renaud[1$], Nisha Agarwal[2$], Dylan J. Richards[3], Silvia Falcinelli[4], E. Starr Hazard[5], Oliana Carnevali[4], John Hyde[6] and Gary Hardiman[1,2,5,7,8*.]

[1] *Department of Medicine, Nephrology, Medical University of South Carolina, Charleston, SC, USA*
[2] *Biomedical Informatics Research Center, San Diego State University, San Diego, CA, USA*
[3] *Bioengineering Department, Clemson University, Clemson, SC, USA*
[4] *Dipartimento di Scienze della Vita e dell'Ambiente, Università Politecnica delle Marche, 60131, Ancona, Italy*
[5] *MUSC Bioinformatics, Center for Genomics Medicine, Medical University of South Carolina, Charleston, SC, USA*
[6] *NOAA Fisheries, Southwest Fisheries Science Center, La Jolla, CA, USA*
[7] *Department of Public Health Sciences, Medical University of South Carolina, Charleston, SC, USA*
[8] *Laboratory for Marine Systems Biology, Hollings Marine Laboratory, Charleston, SC, USA*

[$] *Joint first authors*
[*] *Correspondence should be addressed to Gary Hardiman: hardiman@musc.edu*

Key words: Endocrine disrupting chemicals; Sardine; Mackerel; RNA sequencing; Xenobiotics; 17α-ethynylestradiol; Genomic biomarkers

ABSTRACT

Endocrine disrupting chemicals (EDCs) are substances which disrupt normal functioning of the endocrine system by interfering with hormone regulated physiological pathways. Aquatic environments provide the ultimate reservoir for many EDCs as they enter rivers and the ocean via effluent discharges and accumulate in sediments. One EDC widely dispersed in municipal wastewater effluent discharges is 17α-ethynylestradiol (EE2), an orally bio-active estrogen used in almost all modern formulations of combined oral contraceptive pills and one of the most commonly prescribed medicines. As evidence of the health risks posed by EDCs mount, there is an urgent need to improve diagnostic tools for monitoring the effects of pollutants. As the cost of high throughput sequencing (HTS) diminishes, transcriptional profiling of an organism in response to EDC perturbation presents a cost-effective way of screening a wide range of endocrine responses. Coastal pelagic filter feeding fish species analyzed using HTS provide an excellent tool for EDC risk assessment in the marine environment. Unfortunately, there are limited genome sequence data and annotation for many of these species including Pacific sardine (*Sardinops sagax*) and chub mackerel (*Scomber japonicus*), which limits the utility of molecular tools such as HTS to study disruption of the endocrine system. In this study, we carried out RNA sequencing (RNAseq) of liver RNA harvested from wild sardine and mackerel exposed under laboratory conditions to EE2 and developed an analytical framework for transcriptomic analyses of species with limited genomic information. EE2 exposure altered expression patterns of key genes involved in important metabolic and physiological processes. The systems approach presented here provides a powerful tool for obtaining a comprehensive picture of endocrine disruption in aquatic organisms.

1. INTRODUCTION

The coast of Southern California is home to more than 20 million people. Servicing this high density urban strip are a series of wastewater treatment plants (WWTPs), which discharge more than a billion gallons of treated wastewater into the coastal marine environment on a daily basis [1]. Unfortunately, many contaminants are not filtered at the WWTPs and end up in the marine environment. These contaminants include a broad suite of compounds collectively referred to as endocrine disruptors (EDs) in concentrations that have been shown to disrupt the endocrine system. The derivative of estradiol 17α-ethynylestradiol (EE2) has been a primary component of most modern formulations of combined oral contraceptive pills since the 1960s and qualifies as an ED due to its ability to bind and activate estrogen receptors (ERs) [2]. EE2 is a bio-active estrogen that is minimally metabolized and subsequently released into the environment via the urine and feces from individuals who take it as a medication. It has been detected in effluents from three WWTPs in Southern California at concentrations ranging from 31 to 43 ng/L (100 - 150 pM) [3] and is also detectable in surface waters in the USA (4 - 40 pM) [4] and in Europe (0.3 – 15 pM) [5, 6].

Active at very low concentration, EE2 has been shown to decrease reproductive potential by inhibiting gonad development in both male and female vertebrates, increasing ovarian follicle atresia and inducing complete sex reversal at concentrations ranging from 0.1 to 100 ng/L (0.3 - 340 pM) [7-11]. There is increasing global evidence that exposure to low levels of EDs can lead to disruption of hormone-mediated responses in fish [12-15].

The Pacific sardine (*Sardinops sagax*) and Pacific chub mackerel (*Scomber japonicus*) are coastal pelagic fishes that are major components of commercial fisheries as well as forage base for marine mammals and fishes in California [16, 17]. Sardines (*Sardinops spp.*) are found in most temperate upwelling systems of the world and have supported large commercial fisheries, though their populations are known to historically go through boom and bust periods of abundance, likely due to climatic oscillations and to a lesser degree by industrialized fishing in the past century [18-20]. Along the west coast of North America, Pacific sardines occur across gradients of temperature, salinity, and anthropogenic pollution from southern Canada into the Gulf of California. As filter feeders in a coastal environment, sardines can readily accumulate toxins from the environment and food they eat [21]. Mackerels (*Scomber spp.*) are found in warm and temperate coastal areas throughout the Atlantic, Pacific and Indian oceans. The Pacific chub mackerel is found throughout the Pacific and largely overlaps in distribution with the Pacific sardine where both species school together near the coast and interact with other schooling coastal pelagic species such as northern anchovy (*Engraulis mordax*) and jack mackerel (*Trachurus symmetricus*) [22]. These species are widely consumed by humans, owing to their rich omega-3 fatty acids composition [23] imparting both a distinct flavor profile as well as health benefits. Due to their low trophic diet, cosmopolitan distribution, affinity for coastal habitats, and importance as forage, they serve as ideal sentinel species to monitor the health of the coastal marine environment.

Management of coastal pelagic fish species requires a detailed understanding of the effects of pollutants on the physiology of these species. This need motivated our study to characterize the effect of environmentally relevant levels of EE2 on sardine and mackerel liver transcriptome, the primary organ impacted by pollution and main site of detoxification [24]. Despite their commercial and ecological importance, little is known about the sardine and mackerel genomes, a limiting factor for the use of conventional genomics tools (e.g. microarrays and Q-RT PCR) to study

endocrine disruption in these species. However, recent advances in DNA sequencing technologies have enabled rapid production of transcriptomic data, which has enabled characterization of the expression of specific genes in these species [25, 26]. Furthermore, these new technologies allow for the development of biomarker fingerprints in sentinel species for different chemicals, such as EE2.

In this manuscript, we describe **(1)** a novel RNA sequencing (RNAseq) pipeline using scaffold transcriptomes from related species (**Supplemental Figure S1**) to streamline the data analysis of HTS for sardine and mackerel exposed to EE2, and **(2)** examine the effects of EE2 exposure on these two sentinel species by performing a systems level analysis of their hepatic transcriptome and provide common biomarkers to use as possible end points to predict the estrogen-like contamination in environmental monitoring providing the most relevant Adverse Outcome Pathways (AOP). This knowledge can help improve tests such as the USA EPS Office of Prevention, Pesticides and Toxic Substances (OPPTS) Fish Acute Toxicity Test, Freshwater and Marine and the Organization for Economic Co-operation and Development (OECD) Acute Toxicity Test for Estuarine and Marine Organisms.

2. MATERIALS & METHODS

2.1. Sampling of sardines and mackerels from Southern California coastal locations

Sardine and mackerel were obtained from the live bait receiver operated by Everingham Bros Inc. in Mission Bay, San Diego, California. Fish were allowed to acclimate to research tanks at the NOAA Southwest Fisheries Science Center Experimental Aquarium facility under ambient temperature (~17^{o}C) flow through seawater conditions and fed a pelleted feed (Bio-Oregon) to satiation. For exposure experiments (for both species), five males were exposed to 12.5 pM EE2 for five hours (environmentally relevant concentration in the range of surface water levels) and an additional unexposed five males served as controls. Fish were euthanized with an overdose of Tricaine Methanosulfonate (250mg/L) dissolved in seawater. Livers were harvested and flash-frozen in liquid nitrogen at -196^{o}C. These procedures followed an approved institutional IACUC protocol and all animals were treated humanely. RNA was extracted from liver samples using TRIzol reagent (Invitrogen, CA) and further purified using the RNeasy Mini kit with DNAse to remove DNA (Qiagen, Valencia, CA). RNA concentrations were determined at 260 nm using a ND1000 (Nanodrop, Wilmington, DE). RNA was tested for structural integrity with the 6000 Nano LabChip assay from Agilent, (Santa Clara, CA, USA). Only RNA samples with RIN scores > 7.0 were used for RNA-seq.

2.2. High Throughput Sequencing (HTS)

Libraries for RNA sequencing (RNAseq) were generated using Illumina TruSeq™ RNA Sample Preparation Kit (Illumina, USA) in accordance with the manufacturer's recommendations (**Figure 1A**). The library fragments were purified with AMPure XP system (Beckman Coulter, USA) to select cDNA fragments of approximately 300 bp in length. DNA fragments with ligated adaptor molecules on both ends were selectively enriched using Illumina PCR Primer Cocktail in a 10-cycle PCR reaction. Products were purified using AMPure XP system and quantified using the Agilent high sensitivity DNA assay on the Agilent Bioanalyzer 2100 system. Clustering was performed on a cBot Cluster Generation System using TruSeq PE Cluster Kit v3-cBot-HS (Illumina, USA). After cluster generation, the libraries were sequenced on an Illumina GAIIx

sequencer to a depth of approximately 5 million reads per sample and 101-bp single-end reads were generated. Data was formatted to FASTQ format (with sequences and their associated Phred quality scores) using CASAVA v1.8 (Illumina). RNA sequencing data has been submitted to the Gene expression Omnibus with the accession number **GSE113780**.

2.3. Transcript Level Analysis and Mapping to Reference Transcriptomes

Low quality reads were filtered to exclude those most likely to represent sequencing errors. Adaptor sequences were subsequently trimmed to generate clean full length reads. Only sequences which had a Phred quality score of Q30 (which indicated 99.9% base call accuracy) were used for downstream analysis.

To analyze the mackerel and sardine RNAseq data, an analytical pipeline was developed (**Figure 1B**) that starts with the conversion of fastq files to fasta format via a custom PERL script. As comprehensive genomic sequence data for both sardine and mackerel are presently not available, we exploited the phylogenetic similarity of sardine and zebrafish (**Supplemental Figure S1**), and mapped sardine reads to the zebrafish (*Danio rerio)* Genome Reference Consortium Zebrafish Build 10 (GRCz10) transcriptome using the Nucleotide-Nucleotide Basic Local Alignment Search Tool (BLAST) nucleotide to nucleotide option (blastn) [27]. For mackerel, we mapped reads to two reference transcriptomes, fugu (*Takifugu rubripes*) FUGU 4.0 and stickleback (*Gasterosteus aculeatus*) Gasterosteus aculeatus BROAD S1. Fastq files were converted to fasta files and blastn alignment was carried out against the reference transcriptomes of interest formatted as fasta files.

After determining the threshold alignment length (see below), the reads were assembled into transcript level expression summaries; the number of reads mapped to each gene or transcript, with alignment length greater than the threshold alignment length, were summed, yielding a count as a measure of transcript expression [28].

Each sample yielded a vector of counts or summed mapped reads, with length p equal to the number of transcripts present. Combining the results from n samples, the data were formatted as an (n × p) matrix of n transcripts and p samples. Log-log and square root plots were generated from the data using custom R scripts [29]. The base of the logarithm was arbitrarily selected as 2 as the choice facilitated convenient preliminary data QC (**Supplemental Figure S2**).

In parallel, human ortholog refseq IDs were appended to the fish sequence reads. This was achieved by aligning the reference fish transcriptome with the human proteome using blastx. Utilizing the NCBI gene2refseq function, each human protein ID was linked with its corresponding gene ID, such that the ($n \times p$) matrix of transcript counts, with n number of reference transcriptome genes contained the corresponding human gene IDs.

In order to infer differential gene expression with robust statistical power, we utilized DEseq2, a method which tests for differential expression based on a model using negative binomial distribution [30]. Transcript count data from DESeq2 analysis was ranked according to adjusted p-value (or q-value), the smallest false discovery rate (FDR) at which a transcript is called significant. FDR was calculated using the Benjamini-Hochberg multiple testing adjustment procedure and the cut-off was set at $q \leq 0.1$.

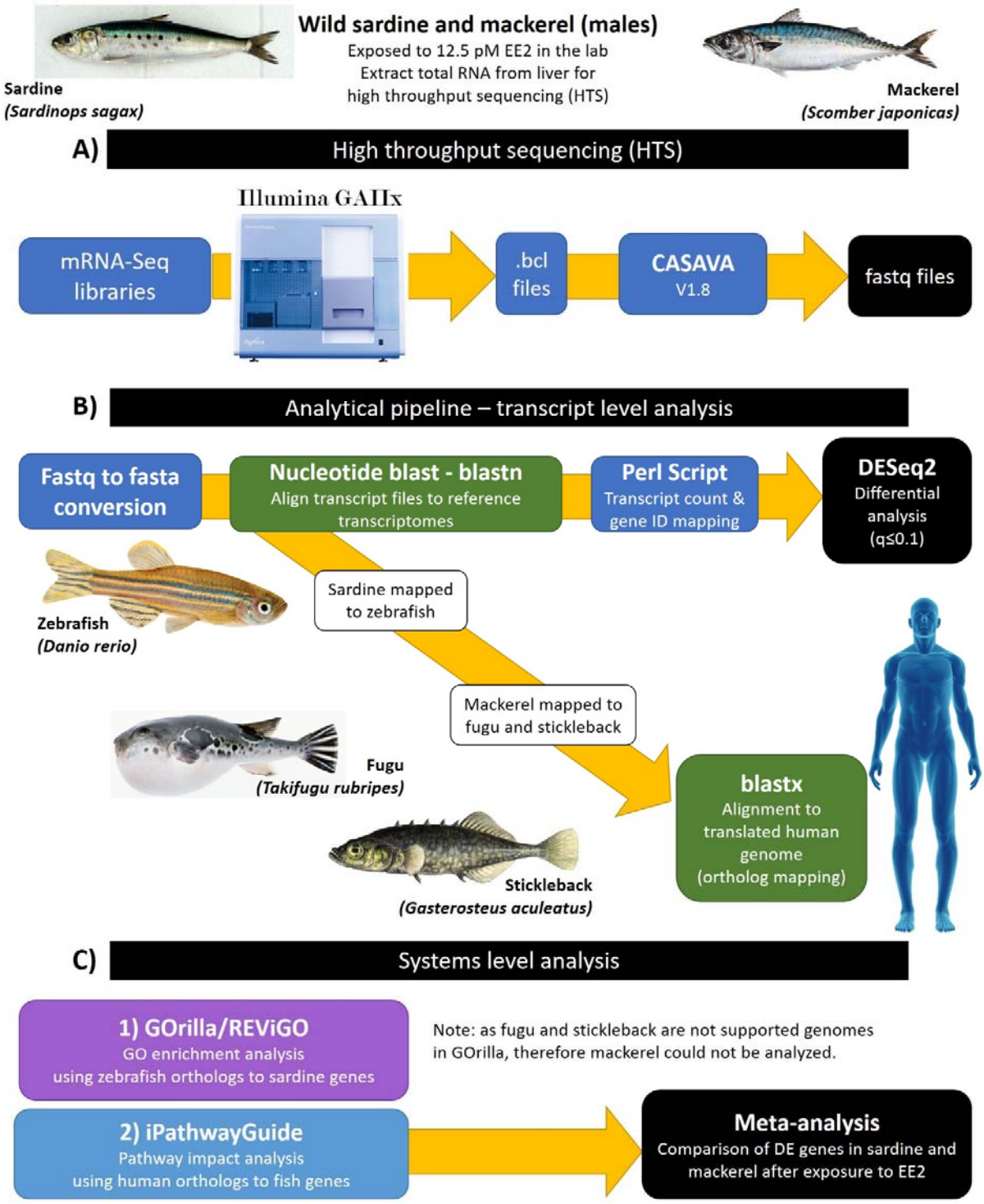

Figure 1: *Workflow followed in this study starting with A) high throughput sequencing (HTS), B) transcript level analysis using a novel analytical pipeline, and C) systems level analysis of the GO enrichment and pathway impact.*

2.4. Systems Level Analysis

Gene ontology (GO) enrichment analysis was performed on the sardine dataset mapped onto zebrafish genome as previously described [31, 32]. Briefly, GO terms were obtained using the online tool Gene Ontology enRIchment anaLysis and visuaLizAtion (GOrilla) (**Figure 1C-1**) [33]. GO terms were summarized using REduce & VIsualize Gene Ontology (REViGO) that combines redundant terms into a single, representative term based on a simple clustering algorithm relying on semantic similarity measures [34]. As fugu and stickleback are not supported genomes in GOrilla at this point in time, mackerel mapped onto these references could not be analyzed using this tool.

Pathway impact analysis (**Figure 1C-2**) was performed using iPathwayGuide (Advaita Bioinformatics) [35], an approach that retrieves information from the Kyoto Encyclopedia of Genes and Genomes (KEGG) database for pathway analysis and diseases [36], from the Gene Ontology Consortium database for GO enrichment analysis [37], and from miRBase and TARGETSCAN databases for miRNA analysis [38, 39].

Area-proportional Venn diagrams were created using VENNY 2.1 [40].

3. RESULTS

3.1. Cross species transcriptome mapping

Due to the lack of reference genomes for sardine and mackerel, reads were mapped to the transcriptomes of phylogenetically similar fish species: sardine to the zebrafish transcriptome and mackerel to the fugu and stickleback transcriptomes (**Supplemental Table S1**). The mean of the total reads obtained with the sardine and mackerel samples were 4 and 3.7 million reads respectively. For sardine, about 1.4 million reads mapped to the zebrafish transcriptome, which represents 34% cross species mapping efficiency. For mackerel, about 1.5 and 3.5 million reads mapped to fugu and stickleback respectively, representing cross species mapping efficiencies of 40.5% for fugu and 94% for stickleback.

3.2. Calculation of threshold alignment length to provide maximal alignment stringency

Sequence read lengths were 101bp. In order to determine the minimum alignment length that provided the highest number of unique reads, we followed a read binning strategy so that a given read was placed in bin 0 if there was no alignment with length $\geq L$. When there were n alignments $\geq L$, the read was placed into bin $min\ (n,10)$. Utilizing this binning approach, based on alignment lengths a comma-separated values (csv) file was created with row numbers equal to $L + 1$ (L = threshold alignment length) and columns corresponding to bins 0 to 10. A series of QC plots were generated with the alignment length plotted on the x-axis, and the individual and fraction of combination of bins on the y-axis. From the plot of (bin1+bin2+bin3)/(bin4+..+bin10) on the y-axis and an alignment length L on the x-axis, which was essentially the fraction of unique (U) to non-unique ($1-U$) versus alignment length (L), it was deduced for sardine and mackerel with the $U/(1-U)$ ratio of 20, that the optimal threshold alignment length was 30 nucleotides (nt), i.e. 95% of the aligned reads were unique with an alignment length of 30 nt (**Figures 2A and 2B respectively**).

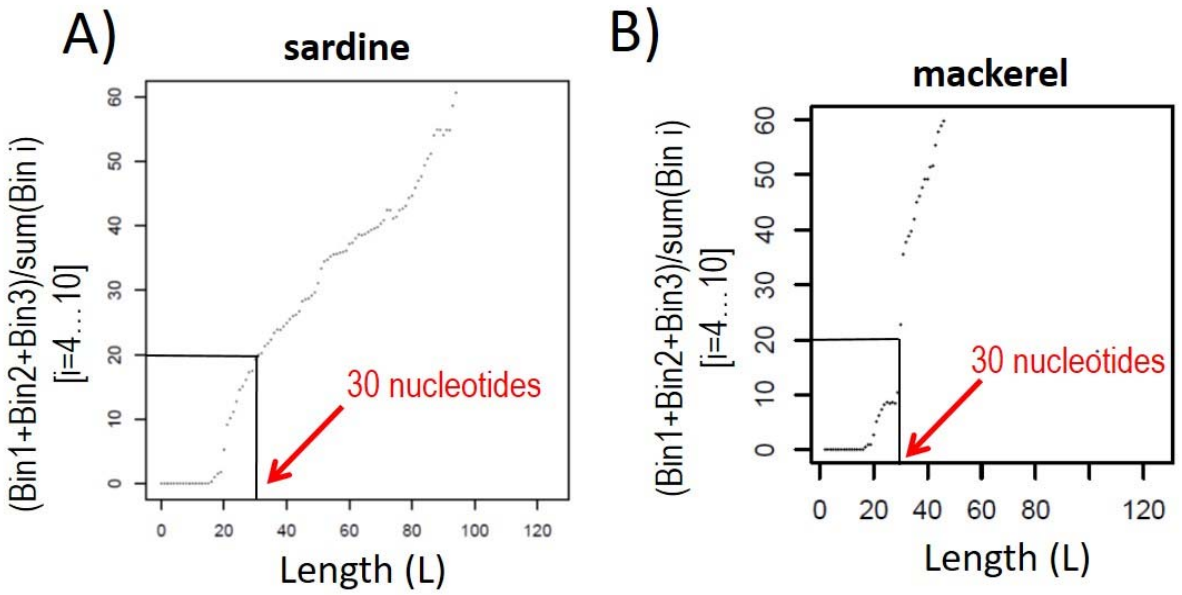

Figure 2: *Threshold alignment length (n=30) for sardine* ***(A)*** *and mackerel* ***(B)****. Alignment lengths plotted on the x-axis (labeled "L"), and the individual and fraction ratio of bin combinations on the y-axis. For sardine and mackerel with the U/(1-U) ratio of 20, the optimal threshold alignment length is 30 nucleotides, i.e. 95% of aligned reads are unique (see red arrows).*

3.3. GO analysis of fish genes

Gene level analysis performed with DESeq2 revealed 4,779 and 2,211 DE hepatic mRNAs in sardine and mackerel exposed to EE2 respectively. For systems level analysis, we first entered the sardine/zebrafish DE gene list in GOrilla since annotations for zebrafish are available. GOrilla GO enrichment analysis showed that translation, muscle contraction, cellular amide metabolism, muscle system, organonitrogen compound metabolism and biosynthesis, multicellular organismal process, regulation of body fluid levels and blood coagulation were some of the most enriched BP terms after exposure to EE2 in the liver of sardines (**Figure 3, deep blue to light green bubbles, and Supplemental Table S2**).

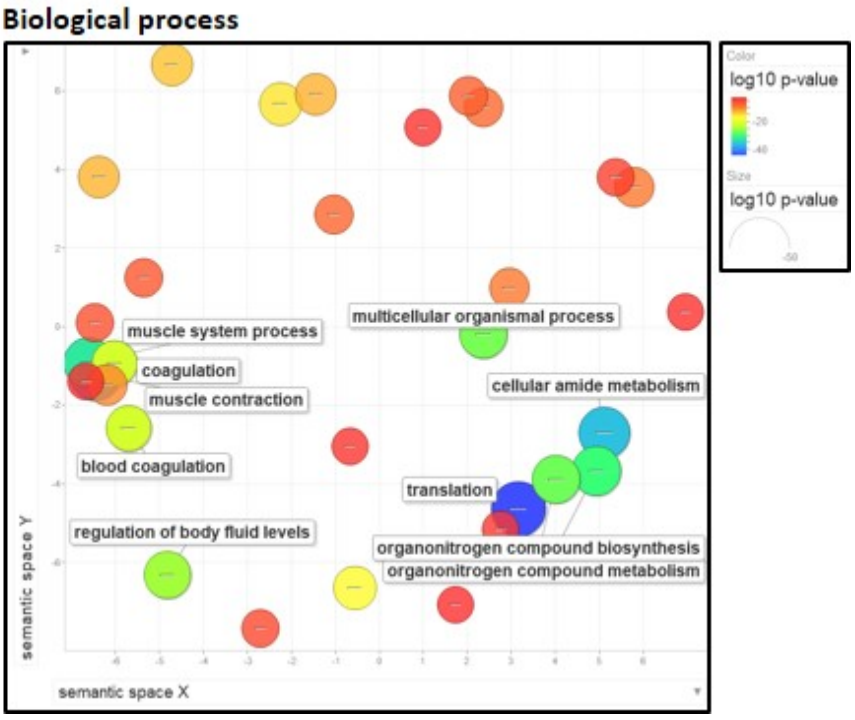

Figure 3: EE2 exposed sardine: *GOrilla GO enrichment analysis - Biological process. The REViGO scatterplots show cluster members in a 2-dimensional space obtained by applying multi-dimensional scaling to a matrix of the GO terms' semantic similarities (the axes have no intrinsic meaning). Bubble color indicates the -log10 p-value (legend in upper right-hand corner, blue and green bubbles are GO terms with greater significant p-values than the orange and red bubbles), the size indicates the GO term frequency in the GO database (bubbles of more general terms are bigger), and the proximity on the plot reflects the semantic similarity.*

3.4. Pathway impact analysis (iPathwayGuide)

Since the zebrafish annotations are not as detailed as human annotations [32], we entered the human orthologs of the sardine/zebrafish and mackerel/fugu DE genes in iPathwayGuide for a content richer GO and pathway impact analysis. Impact analysis uses two types of evidence: i) the overrepresentation of differentially expressed (DE) genes in a given pathway and ii) the perturbation of that pathway computed by propagating the measured expression changes across the pathway topology. Volcano plots revealed a greater dynamic range in the sardine dataset as compared to mackerel (**Supplemental Figure S3**).

DE analysis in sardine exposed to EE2 identified 3,918 DE genes out of a total of 5,321 genes with measured expression (**Supplemental Figure S3A**). Significant differences between the EE2 exposed and control sardines in metabolic, steroid biosynthesis and cancer pathways emerged (**Supplemental Table S3 – Biological Pathways**) as well as enrichment in several signaling pathways such as Fc epsilon RI, cAMP, oxytocin and VEGF. Additionally, complement & coagulation cascades and focal adhesion biological pathways were also significantly perturbed.

The GO analysis revealed that multicellular organismal (**Supplemental Table S3 – Biological Process**), muscle, steroid, metabolic (includes alcohol and cholesterol metabolic processes) related biological processes were highly enriched after exposure to EE2. Response to chemical and drug, negative regulation of peptidase and endopeptidase activity as well as positive regulation of immune system process were also highly represented BP terms in this analysis. Other general BP GO terms were also listed including regulation of biological quality, ion transport, developmental process and system process. Together these results suggest that exposure to EE2 in sardine had a strong impact on transcripts related to metabolism, steroid biosynthesis, several enzymes and signaling pathways as well as the immune system, response to drug/chemical, and cancer pathways.

Exposure to EE2 in mackerel perturbed 2,024 DE genes out of 5,300 genes with measured expression (**Supplemental Figure S3B**). The most enriched biological pathways in EE2 exposed mackerel are fat digestion & absorption, metabolic, steroid biosynthesis and cancer pathways (**Supplemental Table S4 – Biological Pathways**) as well as enrichment in several signaling pathways such as PI3K-Akt and Rap1. ECM-receptor interaction, necroptosis, EGFR tyrosine kinase inhibitor resistance and proteasome pathways were also highly significant in this analysis.

Several BP terms related to metabolic (**Supplemental Table S4 – Biological Process**) and steroid biosynthetic processes were highly enriched in the GO analysis. More specifically, enriched metabolic processes can be sub-divided into alcohol, cholesterol and lipid related processes. Together, these results suggest that exposure to EE2 in mackerel disrupts digestion, steroid biosynthesis, several metabolic processes and signaling pathways, and enhances cancer pathways.

3.5. Meta-analysis (iPathwayGuide)

In order to compare the effects of EE2 exposure on the hepatic transcriptomes of sardine and mackerel we carried out a meta-analysis. This revealed that 910 DE genes (**Figure 4A**), 13 enriched pathways (**Figure 4B and Supplemental Table S5**) and 88 BP (**Figure 4C**) are shared by sardine and mackerel. Additionally, this analysis highlighted that each species has a unique signature after EE2 exposure; 3,008 DE genes, 21 pathways and 405 BP are unique to sardines (**pink in panels A, B, C**) while 1,114 DE genes, 24 pathways and 300 BP are unique to mackerel (**green in panels A, B, C**).

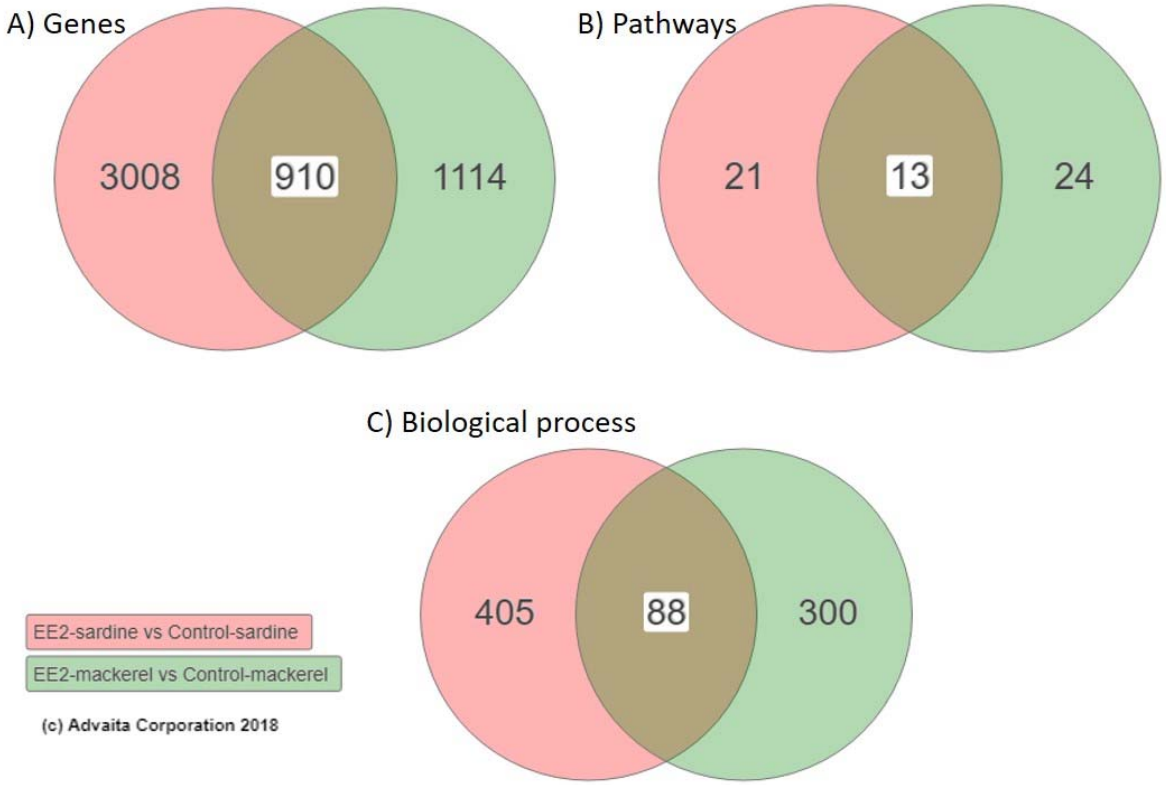

Figure 4: *Venn diagrams from iPathwayGuide.* ***A)*** *Intersection between liver DE genes expressed in sardine and mackerel in response to EE2 exposure; 910 DE genes are commonly expressed in both species while 3,008 and 1,114 DE genes are unique to sardine and mackerel respectively.* ***B)*** *Intersection between enriched pathways in sardine and mackerel; 13 pathways are common to both species while 21 and 24 pathways are unique to sardine and mackerel respectively.* ***C)*** *Intersection between enriched Biological Process (BP) in sardine and mackerel; 88 BP are commonly expressed in both species while 405 and 300 BP are unique to sardine and mackerel respectively.*

4. DISCUSSION

4.1. Development of a novel analytical pipeline for sardine and mackerel

Despite the commercial and ecological importance of sardine and mackerel, very little genomic information is available at the present time. The first goal of this study was to develop a novel analytical pipeline to facilitate **1)** mapping reads to reference transcriptomes using blastn when no reference transcriptome is available for a given species, **2)** calculation of threshold alignment length to provide maximal alignment stringency, **3)** assembly of reads with transcript-level counts or expression summaries, **4)** sequence quality control, **5)** annotation of sequence reads with human GO identifiers, **6)** data normalization (to adjust the read counts for variance in sequencing depths between different samples), differential expression analysis and FDR estimation using DEseq2, **7)** generation of significant DE transcript lists, and **8)** GO and pathway impact analysis.

Based on fish phylogenetic relationships and available reference transcriptomes, we aligned sardine fasta files to the zebrafish transcriptome using blastn, which provided a cross species mapping efficiency of 34%. In the case of mackerel, the fasta files were aligned to both the fugu and stickleback transcriptomes (note that a draft genome sequence has been released for stickleback which is a closer evolutionary relative of mackerel than fugu). The number of mapped sequence tags was considerably higher when stickleback was chosen as the reference (94% cross species mapping efficiency) compared with fugu (40.5%). However, fugu was used as the reference (for all downstream analyses) as this genome is currently better annotated than stickleback [41, 42]. Therefore, zebrafish and fugu transcriptomes served as scaffolds for mapping purposes in our study.

Analysis of the BLAST data output revealed multiple hits for each single query read of varying alignment length to the cognate transcripts. As sequences were being compared across fish species, full length alignment was not expected and consequently many of the small alignment length hits were not unique or meaningful. It was therefore necessary to parse out smaller sequences and determine the optimal alignment length. This was calculated for both fish species so that the only transcript retained was the one whose alignment length was the maximum and greater than the threshold alignment length for each read.

4.2. Systems level analysis of transcriptomic changes using zebrafish annotations

Utilizing the online tools GOrilla and REViGO, we performed GO enrichment analysis using sardine mapped to zebrafish genes for the sardine dataset. This analysis showed that translation (including redundant GO terms peptide and amide biosynthetic process), muscle system & contraction, and cellular amide metabolism were some of the most enriched BP terms after exposure to EE2 in the liver of sardines, followed by organonitrogen compound metabolism and biosynthesis, regulation of body fluid levels and blood coagulation (fibrin clot formation/fibrinolysis, negative regulation of wound healing, hemostasis and blood coagulation). The enriched BP terms include genes that surprisingly mapped to GO BP terms related to muscle system and contraction. This may reflect physiological differences between fish and mammals. The histology of the fish liver for example differs from its mammalian counterpart in that there is a far less tendency of the hepatocytes to form distinct cords or lobules, and the typical portal triads are not obvious [43]. Many of these muscle system & contraction mRNAs are associated with disease signatures (e.g. mouse phenotype database MP: 0003141; cardiac fibrosis and MP:

0005608; cardiac interstitial fibrosis). EE2 exposure has been linked with hepatotoxicity and fibrosis [44] so it is conceivable that these mRNA signatures of fibrosis in other tissues are being detected in this analysis. The liver plays a central role in the clotting process, and acute and chronic liver diseases have been linked with coagulation disorders [45]. Our GO analysis highlighted that EE2 exposure impacted "blood coagulation" as mentioned above, reinforcing that the liver plays a role in the production of clotting factors as well as red blood cell production [46, 47].

4.3. Systems level analysis of transcriptomic changes using human orthologs

Another challenge is the fact that zebrafish and fugu have limited gene annotations when compared to human, further limiting the resolution of the analysis [32]. To remedy this short-coming, human refseq IDs and GO terms were appended to the fish sequence reads by aligning the reference fish transcriptomes with the human proteome using blastx. As 69% of zebrafish genes have at least one human ortholog [48], this approach allows a more sensitive comparative genomics based analysis of transcriptomic changes in these sentinel species. Utilizing the NCBI gene2refseq function, each human protein ID was linked with its corresponding gene ID. However, fish and mammalian species exhibit considerable diversity, each possessing unique biological characteristics. By "humanizing" the fish dataset, species specific genes are lost, *i.e.* vitellogenin, and artefactual data may result if the orthologs are not carefully mapped. However, in the absence of annotated genomes and GO identifiers for the majority fish species, this approach represents a valid analytical alternative that provides functional characterization of the fish transcriptome. Additionally, the BLAST program has an internal p-value threshold, ensuring that the homology between the human and reference fish proteome reported is deemed significant. Furthermore, information obtained about impacted human pathways also provide insights on the potential effects to human health of a particular contaminant.

4.3.1. Effects of EE2 on sardine

The most obvious result from our study is that exposure to 12.5 pM EE2 greatly affected sardine transcripts related to metabolism, including carbon metabolism (KEGG 01200, **Figures S4-S5**), glycolysis and gluconeogenesis (KEGG 00010, **Figures S6-S7**). In particular, the main metabolic perturbation we observed was on the biosynthesis of steroid hormone (KEGG 00140, **Figures S8-S9**). This is consistent with previous studies conducted in turbot [49] and zebrafish [50], and this is the first report to our knowledge of this hormonal disruption induced by EE2 in sardines. Given that steroid hormones are derived from cholesterol and that our GO analysis highlighted that cholesterol biosynthetic and metabolic processes were significantly impacted by exposure to EE2, it is consistent with our previous observation. In rat, EE2 has been shown to decrease testosterone production via a reduction of the cAMP pathway during steroidogenesis [51]. This is interesting as the cAMP signaling pathway (KEGG 04024, **Figures S10-S11**) was one of the most enriched Biological Process in our GO analysis. The cAMP pathway regulates pivotal physiologic processes including metabolism, secretion, calcium homeostasis, muscle contraction, cell fate and gene transcription [35]. cAMP is one of the most common and universal second messengers and its formation is promoted by adenylyl cyclase (AC) activation by ligands including hormones. Estrogenic hormones have been shown to have a non-genomic action that activates second-messenger signaling systems such as cAMP, explaining how steroid hormones can alter the expression of genes [52]. To our knowledge, no studies conducted in fish have examined the impact of EE2 on the cAMP signaling pathway.

Several other signaling pathways were enriched upon EE2 exposure, including VEGF, FcεRI, and oxytocin signaling pathways. The VEGF signaling pathway (KEGG 04370, **Figures S12-S13**) is a crucial signal transducer in both physiologic and pathologic angiogenesis [35]. Its activation leads to proliferation and migration of endothelial cells. Recent studies in human and primate have shown that estrogen regulates VEGF expression [53, 54], however the effects of EE2 on VEGF signaling pathway in fish have not been well characterized to date.

Activation of the FcεRI signaling pathway (KEGG 04664, **Figures S14-S15**) in mast cells induces the release of preformed granules containing histamines, proteoglycans (PGs), prostaglandins and cytokines that contribute to inflammatory responses [35]. Interestingly, many PGs have been shown to be key macromolecules that contribute to the development of various types of cancer including proliferation, adhesion, angiogenesis and metastasis [55]. Our analysis revealed that the pathway "proteoglycans in cancer" was highly enriched in EE2-treated sardines (KEGG 05205, **Figures S16-S17**), suggesting that exposure to EE2 can disrupt many cancer-associated pathways, such as FcεRI and proteoglycan pathways. Minimal research has been carried out on the effects of EE2 on the FcεRI signaling pathway and PGs, highlighting the need for further examination of this subject.

Additionally, the FcεRI signaling pathway plays a role in innate immunity as mast cells have been shown to play a critical role in defense against bacterial, viral and fungal pathogens [56]. Mast cells are activated during innate immune responses by multiple mechanisms, including well-established responses to complement components. We showed here that the "complement & coagulation cascades" pathway (KEGG 04610, **Figures S18-S19**) was highly perturbed by EE2 (consistent with our previous GO analysis with GOrilla using zebrafish annotations), suggesting that EE2 disrupted the innate immunity of exposed sardines. The effect of combined oral contraceptive pills containing EE2 on coagulation parameters has been shown in human [57] but more investigation is needed to understand the effect of EE2 on sardine's innate immunity.

4.3.2. Effects of EE2 on mackerel

In mackerel, the most enriched biological pathway upon EE2 exposure was "fat digestion & absorption" (KEGG 04975, **Figures S20-S21**). Given that fat is an important energy source from food, this emphasizes that EE2 even at low concentration can severely impair major basic pathways. The second most disturbed pathway in EE2-treated mackerel was "chemical carcinogenesis" (KEGG 05204, **Figures S22-S23**). In fact, exposure to environmental chemical carcinogens may account for the majority of human cancers [35] by interacting with a variety of genotoxic and non-genotoxic mechanisms. Genotoxic mechanisms can generate DNA adducts (segments of DNA bound to a cancer-causing chemical) either by direct impact on the DNA or indirectly by affecting metabolism. Non-genotoxic carcinogens induce inflammation, immunosuppression, formation of reactive oxygen species, activation of receptors and epigenetic silencing [35]. Our data suggest that EE2 acts as a carcinogen in mackerel potentially by altering signal-transduction pathways resulting in hypermutability, genomic instability, loss of proliferation control and resistance to apoptosis. This is consistent with previous studies conducted in rats [58], in hamsters [59] and zebrafish [60].

Moreover, EE2 also perturbed "central carbon metabolism in cancer", another pathway that is central to malignant transformation of cells (KEGG 05230, **Figures S24-S25**); as Otto Warburg discovered, a major shift in cellular metabolism is required in cancer cells in order to support growth and survival. This shift results in the consumption of large amounts of glucose,

maintaining high rates of glycolysis and converting glucose into lactic acid even under normal oxygen concentrations [61, 62].

Several signaling pathways were highly enriched upon EE2 treatment in mackerel, including PI3K-Akt and Rap1 signaling pathways. The PI3K-Akt signaling pathway (KEGG 04151, **Figures S26-S27**) is activated by many types of cellular stimuli or toxic insults and regulates fundamental cellular functions such as transcription, translation, proliferation, growth, and survival [35]. PI3K catalyzes the production of phosphatidylinositol-3,4,5-triphosphate (PIP3) at the cell membrane, a second messenger that helps to activate Akt. Once active, Akt can control key cellular processes by phosphorylating substrates involved in apoptosis, protein synthesis, metabolism, and cell cycle. Our data suggest that exposure to EE2 affects fundamental processes by regulating second messenger in mackerel, consistent with our previous conclusion that EE2 regulates cAMP and translation in sardines. Interestingly, estrogens have been shown to induce VEGF-A expression in rat pituitary tumor cells through PI3K-Akt dependent signaling pathway, independently of activation of estrogen receptors [63]. However, the effects of EE2 on the mackerel's PI3K-Akt pathway remain undefined.

The Rap1 signaling pathway (KEGG 04015, **Figures S28-S29**) controls diverse processes such as cell adhesion, cell-cell junction formation and cell polarity [35]. Rap1 plays a dominant role in the control of cell-cell and cell-matrix interactions by regulating the function of integrins and other adhesion molecules in various cell types. Our analysis also revealed that the "ECM-receptor interaction" biological pathway was perturbed by EE2 exposure in mackerels (KEGG 04512, **Figures S30-S31**). The ECM serves an important role in the maintenance of cell and tissue structure and function. Specific interactions between cells and the ECM are mediated by transmembrane molecules, mainly integrins and also proteoglycans, CD36, or other cell-surface-associated components [35]. These interactions lead to a direct or indirect control of cellular activities such as adhesion, migration, differentiation, proliferation, and apoptosis. Together, this suggests that EE2 disrupted cell-cell and cell-matrix mechanisms by perturbing Rap1 signaling and ECM-receptor interaction, in agreement with previous *in vitro* studies [64]. *In vivo* data also suggest that estrogens can exert direct regulatory effects on endothelial cells by increasing surface expression of integrins and enhancing integrin-mediated signaling [65]. Since integrins have an important role in mediating endothelial cell attachment, migration and differentiation, the increase in integrin expression and function induced by estrogens may promote neovascularization and vessel repair. Moreover, our data show that EE2 diminished the expression of three collagen genes (*COL1A1*, *COL1A2* and *COL6A3*). This is consistent with Hansen *et al.* [66] who concluded that administration of EE2 in women suppressed synthesis of collagen in tendon in response to exercise. Our data also show an increase in expression of *ITGB1* in response to EE2. Castillo-Briceno *et al.* [67] have previously demonstrated that EE2 induces liver and kidney toxicity and injuries in fish via upregulation of ECM-related molecules, including ITGB1.

Our analysis also revealed that the "necroptosis" biological pathway was perturbed in EE2 exposed mackerel (KEGG 04217, **Figures S32-S33**). Necroptosis is a programmed form of necrosis. It can be initiated by different stimuli, such as tumor necrosis factor (TNF), TNF-related apoptosis-inducing ligand (TRAIL), Fas ligand (FasL), interferon (IFN), LPS, viral DNA or RNA, DNA-damage agent and requires the kinase activity of receptor-interacting protein 1 (RIPK1) and RIPK3 [35]. Its execution involves ROS generation, calcium overload, the opening of the mitochondrial permeability transition pore, mitochondrial fission, inflammatory response and chromatinolysis. Necroptosis participates in pathogenesis of many diseases, including

neurological diseases, retinal disorders, acute kidney injury, inflammatory diseases and microbial infections. This cell death process is evolutionarily conserved and a *caspase-3* gene has been identified in zebrafish [68] and sea bass [69], retaining the motifs that are functionally important, such as the active site and the cleavage site at the aspartic residue [70]. The examination of the effects of EE2 exposure on necroptosis in fish is in its infancy and further investigation is needed to fully understand its impact in mackerel.

Exposure to EE2 perturbed the "biosynthesis of amino acids" pathway as well (KEGG 01230, **Figures S34-S35**). This KEGG module is the most conserved one in the KEGG MODULE database and is found in almost all completely sequenced genomes [35]. Although human nutritional requirements are well-characterized, little is known of the specific nutrient (protein/amino acid, lipid, carbohydrate, mineral and vitamin) requirements of cultivated fish species; even in zebrafish, a commonly used model, this information is almost entirely lacking [71]. Requirements for amino acids should be delineated in feeding studies with different strains and life stages before the effect of a given contaminant can fully be assessed on this pathway. Nevertheless, our data are consistent with a previous study reporting that the most sensitive biological process to EE2 exposure was amino acid glycosylation [72]; in fact, the authors argued that the genes relating to amino acid glycosylation may be a more sensitive set of biomarkers than the current single standard biomarker of vitellogenin.

4.4. Meta-analysis (Common pathways and biomarkers in sardine and mackerel and extrapolation to humans).

We exploited iPathwayGuide to perform a meta-analysis of the effects of EE2 on the hepatic transcriptome of the sardine and mackerel. Meta-analysis revealed that several DE genes, biological pathways and BP are commonly affected in both species whereas others are unique to sardine or mackerel. "Carbon metabolism (**Figure S36**), steroid hormone biosynthesis (**Figure S37**), glycolysis & gluconeogenesis (**Figure S38**), Rap1 signaling pathway (**Figure S39**) and chemical carcinogenesis" (**Figure S40**) are amongst the 13 perturbed biological pathways shared in both species and discussed in the section above. "Glyoxylate & dicarboxylate metabolism, drug metabolism (other enzymes), retinol metabolism, metabolic pathways, arginine & proline metabolism, steroid biosynthesis, pathways in cancer and pentose phosphate pathway" are also commonly disrupted pathways in both species.

Taken together, the meta-analysis reveals that exposure to 12.5 pM EE2 has a significant impact on basic biological responses in both fish species, impacting essential metabolic processes and biological pathways. These results highlight that EE2 deregulates signaling pathways that are involved in the development and progression of cancer as well as the endocrine and immune systems. The literature on estrogen-induced carcinogenesis (and contaminants with estrogenic properties) is substantial [73-75] yet the exact mechanisms and pathways involved are still under investigation. One established mechanism involves direct genomic action by activation of estrogen-receptor signal transduction pathways associated with increased cell proliferation and inhibition of apoptosis [76-78] via perturbation of genes controlling cell cycle [76] and other tumor-promoting factors such as VEGF [53, 54, 79]. Our findings are consistent with these studies; in sardine, the VEGF signaling pathway was significantly perturbed after EE2 exposure and 4 Biological Process GO terms related to cell cycle were enriched, including "regulation of G1/S transition of mitotic cell cycle and positive regulation of cell cycle G2/M phase transition" (**Supplemental Figure S41**).

Estrogens and estrogenic contaminants can also interfere with cell cycle regulation by upregulating mitogen-activated protein kinases and secondary messengers, such as cAMP, via non-genomic action [77, 78, 80]. Accordingly, we determined that the "cAMP and PI3K-Akt signaling pathways" were significantly impacted in EE2-treated sardines and mackerels respectively. Estrogen metabolites have been shown to have genotoxic (formation of DNA adducts and oxidative DNA damage), mutagenic, transforming, and carcinogenic effects [81, 82]. This is consistent with the enrichment of the "chemical carcinogenesis" pathway in mackerels exposed to EE2 in our study.

Much effort has been expended in search of alternate model for estrogen-related human cancer research. Lam *et al.* have demonstrated that molecular conservation of estrogen-responsiveness between zebrafish and human cancer cell lines makes the zebrafish a great model and established common mechanisms for estrogen-induced carcinogenesis [83]; from an evolutionary perspective, their findings suggest that estrogen regulation of cell cycle is perhaps one of the earliest forms of steroidal-receptor controlled cellular processes. Upon further characterization of the sardine and mackerel transcriptomes, these two species have the potential for future use as additional cancer research models.

In summary the rationale for exploring liver tissue in this study stemmed from the fact that the liver is the primary organ for metabolism and detoxification. The motivation for carrying out EE2 exposure experiments was to facilitate extrapolation of toxicological insights from fish species to human and provide insights into ocean and human health. Our work has uncovered common biomarkers of low level EE2 exposure which provides a basis for improved risk assessment practices based on 'Big Omics Data'. Strategies have been developed by others to derive AOPs (for example, domoic acid exposure) [84]. Perkins and co-workers employed a reverse engineering strategy based on network inference to uncover key nodes (i.e., genes, proteins, metabolites) impacted by chemical perturbation [85, 86]. Kramer and colleagues successfully demonstrated that models can predict potential population impact, once the data are transformed into the prediction of an adverse outcome of demographic significance (e.g., reproduction) at the organism level [87]. Celander *et al.* developed a comparative genomics approach with CYP19 aromatase, and predicted potential aromatase toxicity of new compounds across a multiplicity of species [88]. Systematic organization of Big Data Science into AOP frameworks likely can improve regulatory decision-making through greater integration and more meaningful exploitation of mechanistic data. However, in order to develop a useful knowledgebase that encompasses toxicological contexts of concern to human health risk assessment, novel approaches such as the RNAseq pipeline we described must be developed in accordance with AOP core principles.

5. CONCLUSION

The primary goal of this study was to develop a novel analytical pipeline to elucidate the effects of environmental levels of EE2 on the hepatic transcriptome in sardine and mackerel, two species that do not have a well-characterized genome. Using scaffold transcriptomes from related species (zebrafish for sardine, and fugu for mackerel), we were able to streamline the HTS data analysis and examine the effects of low dose EE2 exposure on these species by performing a systems level analysis of their hepatic transcriptome. We determined that even at low doses, EE2 disrupted basic biological processes and pathways, leading to a signature of metabolic, hormonal and immune dysfunction, as well as carcinogenesis in exposed fish. This work uncovered common dysregulated genes and biological pathways in both species providing biomarkers for EE2

583 exposure in the marine environment, and facilitating monitoring of ocean environmental health.
584 As many of these processes and pathways are well conserved from an evolutionary perspective,
585 and as our analytical pipeline projected the fish genes onto their human orthologs we can infer the
586 impact that EE2 exposure could have on human-health. Our results call for further examination of
587 the effects of estrogenic contaminants on fish and human health at concentrations that represent
588 relevant levels of environmental exposure.
589
590

ACKNOWLEDGEMENTS

We thank J. Sprague, C. Ludka and E. Ricciardelli at the UCSD Biomedical Genomics Facility (BIOGEM) for assistance with high throughput sequencing. We thank Drs. Roman Sasik, Faramarz Valafar, Jienwei Chen, Sean M. Courtney and Jeremy Davis-Turak for bioinformatics and biostatistical insights. We thank Dr. Michael E. Baker for many useful discussions on steroid hormones and endocrine disruption. No writing assistance was utilized in the production of this manuscript.

FUNDING SOURCES

GH acknowledges funding from SC EPSCoR and start-up funding from College of Medicine at the Medical University of South Carolina. SF was a recipient of a training grant from The Campus World Program.

611
612 **REFERENCES**
613
614 1. Baker, M.E.;Vidal-Dorsch, D.E. Molecular analysis of endocrine disruption in hornyhead
615 turbot at wastewater outfalls in southern california using a second generation multi-species
616 microarray. *PLoS One* **2013**. *8,*e75553.
617 2. Escande, A.;Pillon, A. Evaluation of ligand selectivity using reporter cell lines stably
618 expressing estrogen receptor alpha or beta. *Biochemical Pharmacology* **2006**. *71,*1459-
619 1469.
620 3. Soliman, M.A.;Pedersen, J.A.Suffet, I.H. Rapid gas chromatography-mass spectrometry
621 screening method for human pharmaceuticals, hormones, antioxidants and plasticizers in
622 water. *J Chromatogr A* **2004**. *1029,*223-237.
623 4. Filali-Meknassi, Y.;Auriol, M. Quantification of steroid sex hormones using solid-phase
624 extraction followed by liquid chromatography-mass spectrometry. *Water Environ Res*
625 **2007**. *79,*687-696.
626 5. Belfroid, A.;Van der Horst, A. Analysis and occurrence of estrogenic hormones and their
627 glucuronides in surface water and waste water in The Netherlands. *Science of the Total*
628 *Environment* **1999**. *225,*101-108.
629 6. Benijts, T.;Dams, R. Analysis of estrogenic contaminants in river water using liquid
630 chromatography coupled to ion trap based mass spectrometry. *Rapid Communications in*
631 *Mass Spectrometry* **2002**. *16,*1358-1364.
632 7. Hill, R.L.Janz, D.M. Developmental estrogenic exposure in zebrafish (Danio rerio): I.
633 Effects on sex ratio and breeding success. *Aquatic Toxicology* **2003**. *63,*417-429.
634 8. Petersen, G.;Andersen, L. Baseline test studies and development of ELISA. *Zebra Fish for*
635 *Testing Endocrine Disrupting Chemicals. TemaNord* **2000**. *555,*35-46.
636 9. Van den Belt, K.;Verheyen, R.Witters, H. Effects of 17α-ethynylestradiol in a partial life-
637 cycle test with zebrafish (Danio rerio): effects on growth, gonads and female reproductive
638 success. *Science of the Total Environment* **2003**. *309,*127-137.
639 10. Van den Belt, K.;Verheyen, R.Witters, H. Comparison of vitellogenin responses in
640 zebrafish and rainbow trout following exposure to environmental estrogens. *Ecotoxicology*
641 *and environmental safety* **2003**. *56,*271-281.
642 11. Weber, L.P.;Hill, R.L.Janz, D.M. Developmental estrogenic exposure in zebrafish (Danio
643 rerio): II. Histological evaluation of gametogenesis and organ toxicity. *Aquatic Toxicology*
644 **2003**. *63,*431-446.
645 12. Filby, A.L.;Neuparth, T. Health impacts of estrogens in the environment, considering
646 complex mixture effects. *Environ Health Perspect* **2007**. *115,*1704-1710.
647 13. Jin, Y.;Shu, L. Temperature and photoperiod affect the endocrine disruption effects of
648 ethinylestradiol, nonylphenol and their binary mixture in zebrafish (Danio rerio). *Comp*
649 *Biochem Physiol C Toxicol Pharmacol* **2010**. *151,*258-263.
650 14. Jobling, S.;Burn, R.W. Statistical modeling suggests that antiandrogens in effluents from
651 wastewater treatment works contribute to widespread sexual disruption in fish living in
652 English rivers. *Environ Health Perspect* **2009**. *117,*797-802.
653 15. Tyler, C.R.;Lange, A. The roach (Rutilus rutilus) as a sentinel for assessing endocrine
654 disruption. *Environ Sci* **2007**. *14,*235-253.

16. Cardona, L.;Martínez-Iñigo, L. The role of sardine as prey for pelagic predators in the western Mediterranean Sea assessed using stable isotopes and fatty acids. *Marine Ecology Progress Series* **2015**. *531,*1-14.
17. Trites, A.Donnelly, C. The decline of Steller sea lions Eumetopias jubatus in Alaska: a review of the nutritional stress hypothesis. *Mammal review* **2003**. *33,*3-28.
18. Baumgartner, T. Reconstruction of the history of the Pacific sardine and northern anchovy populations over the past two millenia from sediments of the Santa Barbara basin, California. *CalCOFI Rep* **1992**. *33,*24-40.
19. Hill, K.T.;Crone, P.R.Zwolinski, J.P. Assessment of the Pacific sardine resource in 2017 for U.S. management in 2017-18. Pacific Fishery Management Council, April 2017 Briefing Book. 2017: Portland, Oregon. p. 146 p.
20. McClatchie, S.;Hendy, I. Collapse and recovery of forage fish populations prior to commercial exploitation. *Geophysical Research Letters* **2017**. *44,*1877-1885.
21. Lefebvre, K.;Silver, M. Domoic acid in planktivorous fish in relation to toxic Pseudo-nitzschia cell densities. *Marine Biology* **2002**. *140,*625-631.
22. Zeng, L.Cheng, Q. Thirty novel microsatellite markers for the coastal pelagic fish, Scomber japonicus (Scombridae). *J Genet* **2012**. *91,*e64-68.
23. Kris-Etherton, P.M.;Harris, W.S.Appel, L.J. Fish consumption, fish oil, omega-3 fatty acids, and cardiovascular disease. *Circulation* **2002**. *106,*2747-2757.
24. Mitra, V.Metcalf, J. Metabolic functions of the liver. *Anaesthesia & Intensive Care Medicine* **2012**. *13,*54-55.
25. Hardiman, G. Application of ultra-high throughput sequencing and microarray technologies in pharmacogenomics testing. *In: Dasgupta A, editor. Therapeutic Drug Monitoring. London: Elsevier; 2012.* **2012**144–159.
26. Fox, S.;Filichkin, S.Mockler, T.C. Applications of ultra-high-throughput sequencing. *Methods Mol Biol* **2009**. *553,*79-108.
27. Altschul, S.F.;Gish, W. Basic local alignment search tool. *J Mol Biol* **1990**. *215,*403-410.
28. Li, J.Tibshirani, R. Finding consistent patterns: A nonparametric approach for identifying differential expression in RNA-Seq data. *Stat Methods Med Res* **2013**. *22,*519-536.
29. Team, R.C. R: A language and environment for statistical computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2014. 2015.
30. Love, M.I.;Huber, W.Anders, S. *Moderated estimation of fold change and dispersion for RNA-Seq data with DESeq2*Country, 2014.
31. Huff, M.;da Silveira, W.A. Systems Analysis of the Liver Transcriptome in Adult Male Zebrafish Exposed to the Plasticizer (2-Ethylhexyl) Phthalate (DEHP). *Sci Rep* **2018**. *8,*2118.
32. Renaud, L.;Silveira, W.A.D. The Plasticizer Bisphenol A Perturbs the Hepatic Epigenome: A Systems Level Analysis of the miRNome. *Genes (Basel)* **2017**. *8*.
33. Eden, E.;Navon, R. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics* **2009**. *10,*48.
34. Supek, F.;Bosnjak, M. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One* **2011**. *6,*e21800.
35. Draghici, S.;Khatri, P. A systems biology approach for pathway level analysis. *Genome Res* **2007**. *17,*1537-1545.

36. Kanehisa, M.Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* **2000**. *28*,27-30.
37. Ashburner, M.;Ball, C.A. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **2000**. *25*,25-29.
38. Agarwal, V.;Bell, G.W. Predicting effective microRNA target sites in mammalian mRNAs. *Elife* **2015**. *4*.
39. Kozomara, A.Griffiths-Jones, S. miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Research* **2013**. *42*,D68-D73.
40. Oliveros, J.C. VENNY. An interactive tool for comparing lists with Venn Diagrams. **2007**.
41. Aparicio, S.;Chapman, J. Whole-genome shotgun assembly and analysis of the genome of Fugu rubripes. *Science* **2002**. *297*,1301-1310.
42. Jones, F.C.;Grabherr, M.G. The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* **2012**. *484*,55.
43. NCTC. Fish Histology and Histopathology Manual. 2014 03/06/2018]; Available from: https://training.fws.gov/resources/course-resources/fish-histology/index.html.
44. Pandey, G.;Pandey, S.Sharma, M. Experimental hepatotoxicity produced by ethinyl estradiol. *Toxicology international* **2011**. *18*,160.
45. Amitrano, L.;Guardascione, M.A. Coagulation disorders in liver disease. *Semin Liver Dis* **2002**. *22*,83-96.
46. Bourdeau, V.;Deschenes, J. Genome-wide identification of high-affinity estrogen response elements in human and mouse. *Mol Endocrinol* **2004**. *18*,1411-1427.
47. Cleuren, A.C.;Van der Linden, I.K. 17alpha-Ethinylestradiol rapidly alters transcript levels of murine coagulation genes via estrogen receptor alpha. *J Thromb Haemost* **2010**. *8*,1838-1846.
48. Howe, K.;Clark, M.D. The zebrafish reference genome sequence and its relationship to the human genome. *Nature* **2013**. *496*,498-503.
49. Labadie, P.Budzinski, H. Alteration of steroid hormone profile in juvenile turbot (Psetta maxima) as a consequence of short-term exposure to 17α-ethynylestradiol. *Chemosphere* **2006**. *64*,1274-1286.
50. Hoffmann, J.;Torontali, S. Hepatic gene expression profiling using Genechips in zebrafish exposed to 17α-ethynylestradiol. *Aquatic Toxicology* **2006**. *79*,233-246.
51. Kuo, T.-H.;Wang, K.-L. Inhibitory Effects of 17 Alpha-Ethynylestradiol on the Production of Testosterone by Rat Leydig Cells. 2012, Oxford University Press.
52. Aronica, S.M.;Kraus, W.L.Katzenellenbogen, B.S. Estrogen action via the cAMP signaling pathway: stimulation of adenylate cyclase and cAMP-regulated gene transcription. *Proceedings of the National Academy of Sciences* **1994**. *91*,8517-8521.
53. Chennazhi, K.P.Nayak, N.R. Regulation of angiogenesis in the primate endometrium: vascular endothelial growth factor. *Semin Reprod Med* **2009**. *27*,80-89.
54. Okada, H.;Tsuzuki, T. Regulation of decidualization and angiogenesis in the human endometrium: mini review. *J Obstet Gynaecol Res* **2014**. *40*,1180-1187.
55. Theocharis, A.D.Karamanos, N.K. Proteoglycans remodeling in cancer: Underlying molecular mechanisms. *Matrix Biol* **2017**.
56. Marshall, J.S.Jawdat, D.M. Mast cells in innate immunity. *Journal of Allergy and Clinical Immunology* **2004**. *114*,21-27.

57. Al-Obaidy, M.J.Al-Youzbaki, W.B. Effect of combined oral contraceptive pills on some coagulation parameters. *Effect of educational level of pregnant women on compliance with antenatal care services in Erbil city* 164.
58. Yager, J.D.;Roebuck, B.D. Effects of ethinyl estradiol and tamoxifen on liver DNA turnover and new synthesis and appearance of gamma glutamyl transpeptidase-positive foci in female rats. *Carcinogenesis* **1986**. *7,*2007-2014.
59. Li, J.J.Li, S.A. Estrogen carcinogenesis in Syrian hamster tissues: role of metabolism. *Federation proceedings* **1987**. *46,*1858-1863.
60. Notch, E.G.;Miniutti, D.M.Mayer, G.D. 17α-Ethinylestradiol decreases expression of multiple hepatic nucleotide excision repair genes in zebrafish (Danio rerio). *Aquatic Toxicology* **2007**. *84,*301-309.
61. Diaz-Ruiz, R.;Rigoulet, M.Devin, A. The Warburg and Crabtree effects: On the origin of cancer cell energy metabolism and of yeast glucose repression. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **2011**. *1807,*568-576.
62. Warburg, O. On the origin of cancer cells. *Science* **1956**. *123,*309-314.
63. Banerjee, S.;Saxena, N. 17α-Estradiol-induced VEGF-A expression in rat pituitary tumor cells is mediated through ER independent but PI3K-Akt dependent signaling pathway. *Biochemical and Biophysical Research Communications* **2003**. *300,*209-215.
64. Hofbauer, R.;Frass, M. Oral contraceptives that contain ethinyl estradiol (0.035 mg) and cyproterone acetate (2 mg) inhibit leukocyte transmigration through endothelial cell monolayers. *Fertility and sterility* **1999**. *72,*652-656.
65. Cid, M.C.;Esparza, J. Estradiol enhances endothelial cell interactions with extracellular matrix proteins via an increase in integrin expression and function. *Angiogenesis* **1999**. *3,*271-280.
66. Hansen, M.;Koskinen, S.O. Ethinyl oestradiol administration in women suppresses synthesis of collagen in tendon in response to exercise. *The Journal of physiology* **2008**. *586,*3005-3016.
67. Castillo-Briceño, P.;Arizcun-Arizcun, M. Correlated expression profile of extracellular matrix-related molecules during the inflammatory response of the teleost fish gilthead seabream. *Developmental & Comparative Immunology* **2010**. *34,*1051-1058.
68. YABU, T.;KISHI, S. Characterization of zebrafish caspase-3 and induction of apoptosis through ceramide generation in fish fathead minnow tailbud cells and zebrafish embryo. *Biochemical Journal* **2001**. *360,*39-47.
69. Reis, M.I.R.;Nascimento, D.S. Molecular cloning and characterisation of sea bass (Dicentrarchus labrax L.) caspase-3 gene. *Molecular Immunology* **2007**. *44,*774-783.
70. Rojas, V.;Guzmán, F. Development of a caspase-3 antibody as a tool for detecting apoptosis in cells from rainbow trout (Oncorhynchus mykiss). *Electronic Journal of Biotechnology* **2012**. *15,*12-12.
71. Lawrence, C. The husbandry of zebrafish (Danio rerio): a review. *Aquaculture* **2007**. *269,*1-20.
72. Yang, L.;Allen, B.C.Thomas, R.S. BMDExpress: a software tool for the benchmark dose analyses of genomic data. *BMC Genomics* **2007**. *8,*387.
73. Paulose, T.;Speroni, L. Estrogens in the wrong place at the wrong time: Fetal BPA exposure and mammary cancer. *Reproductive Toxicology* **2015**. *54,*58-65.

74. Rutkowska, A.Z.;Szybiak, A. Endocrine disrupting chemicals as potential risk factor for estrogen-dependent cancers. *Pol Arch Med Wewn* **2016**. *126*,562-570.
75. Trevino, L.S.;Wang, Q.Walker, C.L. Hypothesis: Activation of rapid signaling by environmental estrogens and epigenetic reprogramming in breast cancer. *Reprod Toxicol* **2015**. *54*,136-140.
76. Pearce, S.T.Jordan, V.C. The biological role of estrogen receptors α and β in cancer. *Critical reviews in oncology/hematology* **2004**. *50*,3-22.
77. Russo, J.Russo, I.H. The role of estrogen in the initiation of breast cancer. *The Journal of steroid biochemistry and molecular biology* **2006**. *102*,89-96.
78. Yager, J.D.Davidson, N.E. Estrogen carcinogenesis in breast cancer. *New England Journal of Medicine* **2006**. *354*,270-282.
79. Mueller, M.D.;Vigne, J.-L. Regulation of vascular endothelial growth factor (VEGF) gene transcription by estrogen receptors α and β. *Proceedings of the National Academy of Sciences* **2000**. *97*,10972-10977.
80. Bonkhoff, H. Estrogen receptor signaling in prostate cancer: Implications for carcinogenesis and tumor progression. *Prostate* **2018**. *78*,2-10.
81. Roy, D.;Cai, Q. Estrogen-induced generation of reactive oxygen and nitrogen species, gene damage, and estrogen-dependent cancers. *Journal of Toxicology and Environmental Health, Part B* **2007**. *10*,235-257.
82. Russo, J.;Lareef, M.H. Estrogen and its metabolites are carcinogenic agents in human breast epithelial cells. *The Journal of steroid biochemistry and molecular biology* **2003**. *87*,1-25.
83. Lam, S.H.;Lee, S.G. Molecular conservation of estrogen-response associated with cell cycle regulation, hormonal carcinogenesis and cancer in zebrafish and human cancer cell lines. *BMC Medical Genomics* **2011**. *4*,41.
84. Watanabe, K.H.;Andersen, M.E. Defining and modeling known adverse outcome pathways: Domoic acid and neuronal signaling as a case study. *Environmental toxicology and chemistry* **2011**. *30*,9-21.
85. Perkins, E.J.;Chipman, J.K. Reverse engineering adverse outcome pathways. *Environmental toxicology and chemistry* **2011**. *30*,22-38.
86. Perkins, E.J.;Antczak, P. Adverse outcome pathways for regulatory applications: examination of four case studies with different degrees of completeness and scientific confidence. *Toxicological Sciences* **2015**. *148*,14-25.
87. Kramer, V.J.;Etterson, M.A. Adverse outcome pathways and ecological risk assessment: Bridging to population - level effects. *Environmental Toxicology and Chemistry* **2011**. *30*,64-76.
88. Celander, M.C.;Goldstone, J.V. Species extrapolation for the 21st century. *Environmental toxicology and chemistry* **2011**. *30*,52-63.