Identification of conserved hepatic transcriptomic responses to 17β-estradiol using high-throughput sequencing in brown trout

Tamsyn M. Uren Webster, Janice A. Shears, Karen Moore & Eduarda M. Santos*

a. Biosciences, College of Life & Environmental Sciences, Geoffrey Pope Building, University of Exeter, Exeter, EX4 4QD

Corresponding Author:

* Eduarda M. Santos

Biosciences, College of Life & Environmental Sciences, Geoffrey Pope Building, University of Exeter, Exeter, EX4 4QD

E.Santos@exeter.ac.uk, Phone: +44 (0)1392 264607, Fax: +44 (0)1392 263434

Running Title:

RNA-seq in estrogen-exposed brown trout

Abstract

Estrogenic chemicals are major contaminants of surface waters and can threaten the sustainability of natural fish populations. Characterisation of the global molecular mechanisms of toxicity of environmental contaminants has been conducted primarily in model species rather than species with limited existing transcriptomic or genomic sequence information. We aimed to investigate the global mechanisms of toxicity of an endocrine disrupting chemical of environmental concern (17β-estradiol; E2) using high-throughput RNA sequencing (RNA-seq) in an environmentally-relevant species, brown trout (*Salmo trutta*). We exposed mature males to measured concentrations of 1.94, 18.06 and 34.38 ng E2/L for four days and sequenced three individual liver samples per treatment using an Illumina HiSeq 2500 platform. Exposure to 34.4 ng E2/L resulted in 2113 differentially-regulated transcripts (FDR<0.05). Functional analysis revealed up-regulation of processes associated with vitellogenesis, including lipid metabolism, cellular proliferation and ribosome biogenesis, together with a down regulation of carbohydrate metabolism. Using RT-QPCR, we validated the expression of eight target genes, and identified significant differences in the regulation of several known estrogen-responsive transcripts in fish exposed to the lower treatment concentrations (including *esr1* and *zp2.5*). We successfully used RNA-seq to identify highly conserved responses to estrogen, and also identified some estrogen-responsive transcripts which have been less well characterised, including *nots* and *tgm2l*. These results demonstrate the potential application of RNA-seq as a valuable tool for assessing mechanistic effects of pollutants in ecologically-relevant species for which little genomic information is available.

Keywords: RNA-seq, Illumina, transcriptomics, sequencing, salmonid

Introduction

The major endogenous estrogen in vertebrates, 17β-estradiol (E2), is a significant contributor to the estrogenic contamination of surface waters, and E2 equivalent concentrations (EEQs) of up to10 ng/L have been reported in rivers worldwide(10, 16, 26). In addition to input via wastewater treatment work effluents, E2 enters rivers in livestock and poultry waste, and can cause pulses of contamination (50, 51). In water bodies, E2 can also act in conjunction with other natural and synthetic estrogenic chemicals (i.e. estrone; ethynylestradiol, phytoestrogens, alkylphenols and other industrial chemicals) to cause adverse effects in natural populations of fish. Reported effects of environmental estrogens include the induction of intersex in many species including roach (28) and gudgeon (58), decreased reproductive success in wild fish (21, 27) and population collapses (31), providing evidence for the risks that estrogens pose to the sustainability of wild fish populations.

The effects of E2, and other estrogenic contaminants, are mediated predominantly via genomic pathways through binding and activation of nuclear estrogen receptors, which are ligand-dependent transcription factors (48). Through this mechanism, estrogen exposure is associated with a highly conserved induction of a well characterised suite of responsive genes. Of these, vitellogenin induction in male and juvenile fish has been the most widely described. In addition, induction of the transcripts encoding for estrogen receptor 1 (*esr1*) and zona pellucida proteins are also well characterised responses to estrogen exposure (e.g. 18, 48, 55, 60). The transcription of these genes in the liver is known to be strongly associated with the stage of vitellogenesis in females, and regulated via estrogen signalling (1). In addition to regulating the reproductive system, estrogens play a crucial role in a diverse range of other physiological processes including skeletal, muscular, cardiovascular, immune and ion-regulatory systems, all of which are therefore potential targets for disruption following exposure to estrogenic contaminants in fish (19, 48).

Transcriptomic approaches have been employed to characterise both the normal endogenous effects of estrogen signalling in females, and the effects of exposure to a number of estrogenic chemicals in male and juvenile fish using microarrays (e.g. 4, 18, 30, 34) and high-throughput sequencing (RNA-sage) (60). These studies have reported extensive transcriptional changes, reflecting the diverse range of genes and processes regulated by estrogens, including a number of broadly conserved pathways. High-throughput RNA sequencing (RNA-seq) has recently emerged as a robust, accurate and reproducible tool for conducting transcriptomics (36, 43, 49) but, as yet, this approach has rarely been applied to ecotoxicology. A major advantage of this technique is that it can be used to conduct non-biased, global mechanistic analysis in any species of interest without a requirement for prior sequence information.

In this project, we employed RNA-seq on an Illumina HiSeq 2500 platform to characterise the global hepatic transcriptomic responses of sexually mature male brown trout following exposure to E2, including an environmentally relevant concentration. Brown trout are an ecologically and economically important native European species, known to be sensitive to environmental stressors, but studies conducting mechanistic evaluations of its response to chemical toxicity are scarce. E2, originating from agricultural pollution, may be one of the environmentally relevant chemicals potentially affecting brown trout populations, which typically inhabit, and spawn in, smaller streams within farmland catchments. Additionally, by investigating the response to an estrogen, we aimed to discuss the suitability of RNA-seq to identify a conserved mechanistic response and its role as a valuable and robust tool in ecotoxicology.

Materials and methods

Fish maintenance

All experiments were conducted under approved protocols according to the UK Home Office regulations for use of animals in scientific procedures.

A mixed sex population of brown trout (2 years old) including mature and immature fish of both sexes were obtained from a local aquaculture facility (Hooke Springs Trout Farm, Dorset, UK) in late September, to correspond with the latter stages of reproductive maturation in this species, and maintained in 215 L tanks to allow for acclimation to laboratory conditions for three weeks prior to the start of the exposure, at the University of Exeter, UK. Each tank was aerated, supplied with 430 L/day de-chlorinated tap water, and maintained at 12 ± 0.2 °C, pH 7.5. Fish were kept under a 16:8 h light:dark cycle (with 30 minute dawn/dusk transitional periods) and fed with pellet feed (8 mm, Biomar, Grangemouth, UK) at a rate of 2% body weight per day. In order to exclude sexually mature females, which would be naturally excreting estrogens that may have influenced the exposure, we measured plasma calcium concentrations in all fish prior to the start of the exposure. Concentration of plasma calcium is known to be a good indicator of vitellogenin, and maturity status of female fish (40). Fish used in the exposure experiment included mature males and immature fish of both sexes.

Chemical exposures and sampling

Chemical exposure was conducted via a flow through system for a period of four days. This exposure duration was chosen because short term estrogen exposures are known to induce considerable and rapid transcriptional change (19) and to limit any possible confounding secondary effects of estrogen exposure (38). Four days can also be expected to simulate pulses of environmental estrogenic exposure associated with agricultural pollution (50, 51). Fish were exposed to three nominal concentrations, 2.5, 25 and 250 ng E2/L (17β-estradiol ≥98% purity, Sigma) or a dilution water control. The lowest concentration is in the range of EEQ concentrations reported in the environment (16, 26), while the higher concentrations were selected to facilitate mechanistic analysis.

Each treatment group consisted of one tank containing 8 individual fish (mature males and sexually immature fish of both sexes), and the control treatment was run in duplicate. Water samples were collected from each tank on day 3 of the exposure period and stored at -20 °C

prior to chemical analysis, using an Enzyme Immunoassay for Estradiol kit (Oxford Biomedical Research, Oxford, MI, USA) according to the manufacturer's instructions. Samples were diluted or concentrated (using an appropriate ratio of ethyl acetate and exposure water) in order to fall within the range of assay detection (0.02-2 μg/L) and measured in duplicate. The cross-reactivity of this assay was 100% for 17β-estradiol and ≤1.00% for testosterone and other sex steroids. The measured concentrations of E2 in the water were 1.94, 18.06 and 34.38 ng E2/L. The relatively low concentration of E2 measured in the 250 ng E2/L treatment group is likely due to its poor water solubility, given that we performed the exposure without the use of solvents to increase its environmental relevance. Concentrations of E2 in the concentrated stock solutions used to prepare the exposure concentrations were measured in parallel with exposure water samples and were 85, 77 and 43 % of the nominal values for the 1.9, 18.1 and 34.4 ng/L treatments, respectively. This suggests that the poor recovery of E2 in the highest treatment concentration predominantly resulted from poor solubility in the stock solution. Throughout this paper, we refer to the measured concentrations of E2 to indicate the exposure concentrations.

Fish were humanely sacrificed on day four of the exposure period by a lethal dose of benzocaine (0.5 g/L; Sigma-Aldrich) followed by destruction of the brain by pithing with a blade, in accordance with UK Home Office regulations. Wet weight and fork length were recorded and the condition factor [k= (weight (g) x 100)/ (fork length $(cm)^3$)] was calculated for individual fish. Sex and maturity of all fish was confirmed by observation of the gonads, and gonadsomatic index (GSI) [(gonad weight (mg)/ total weight (mg)) x 100)] was determined. Livers were dissected and weighed, and the hepatosomatic index (HSI) [(liver weight (mg)/ total weight (mg)) x 100)] was determined for individual fish. Portions of the liver were then snap frozen in liquid nitrogen and stored at -80°C prior to RNA extraction. Statistical analysis of morphological parameters was conducted using SigmaStat (version 12.0). All morphometric data met assumptions of normality (verified using the Shapiro-Wilk

166 test) and equal variance, and were analysed using single factor one way analysis of variance
167 (ANOVA).
168
169 *RNA extraction, library preparation and sequencing*
170 Transcript profiling was conducted in the livers of three sexually mature males per treatment
171 group. RNA was extracted from livers using TRI reagent (Sigma-Aldrich) according to the
172 manufacturer's instructions, then further purified and treated with DNase on RNeasy Mini
173 extraction columns (Qiagen). The concentration, purity and integrity of RNA were determined
174 using a NanoDrop ND-1000 Spectrophotometer (NanoDrop Technologies, USA) and an
175 Agilent 2100 Bioanalyzer (Agilent Technologies, Inc., USA). All RNA input to library
176 construction was of high quality with 260/280 and 260/230 ratios > 1.8 and RIN scores >8.
177 External RNA Controls Consortium (ERCC) spike-in control mixes (Ambion) were added to
178 all individual RNA samples, according to the manufacturer's instructions. cDNA libraries from
179 all 15 samples were then prepared using the Illumina TruSeq RNA Sample Preparation kit,
180 multiplexed with 24 samples per lane (together with samples from another project) and
181 sequenced using an Illumina HiSeq 2500 to generate 100 bp paired-end reads, according to
182 the manufacturer's instructions.
183
184 *Transcriptome Assembly and Annotation*
185 To maximise sequence coverage depth and assemble an optimised male liver transcriptome
186 for brown trout, sequence reads from all samples from the current study were combined with
187 those from another project. Transcriptome assembly and annotation was conducted as
188 described previously (57), and this transcriptome was then used as a basis for expression
189 analysis in both projects. Contaminating Illumina adaptor sequences were removed and the
190 first 12 bp of all raw sequence reads were trimmed to remove 5' bias caused by random
191 hexamer priming using the FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit, July
192 2013). 3' sliding window quality trimming was performed
193 (http://wiki.bioinformatics.ucdavis.edu/index.php/ Trim.slidingWindow.pl , July 2013) and all

194 reads where < 90% bases had a Phred quality score >20, and those shorter than 15 bp,
195 were discarded. Digital normalisation was performed to remove highly duplicated reads
196 using the normalize-by-median.py script part of the khmer package described by Brown et
197 al. (7), with the recommended k-mer value of 20 and a coverage threshold of 200. This
198 process reduces the computer memory requirements of transcriptome assembly, and also
199 reduces the risk of potential sequencing error accumulation in abundant transcripts. All
200 retained reads were then paired, separated into forward and reverse fastq files before *de*
201 *novo* transcriptome assembly using Trinity (version r2013-02-25;(15), using the default
202 parameters and specifying a minimum contig length of 200 bp). All transcripts were
203 annotated using Blastx against Ensembl peptide databases (Release 71; April 2013) using
204 an e-value cut off < $1e^{-15}$ and assigned in the following preferential order; zebrafish (*Danio*
205 *rerio*); human (*Homo sapiens*) and mouse (*Mus musculus*); then all other available fish
206 species (stickleback (*Gasterosteus aculeatus*), medaka (*Oryzias latipes*), tilapia
207 (*Oreochromis niloticus*) and cod (*Gadus morhua*)). Additional annotation of previously un-
208 annotated differentially expressed transcripts was performed using Blast (< $1e^{-15}$) against
209 refseq, nr and nt databases.
210
211 *Transcriptomic Analysis*
212 Raw sequence reads from individual samples were mapped back against the assembled
213 transcripts using Bowtie2 (version 2.1.0 (32)), using the -k 1 parameter to report a single
214 best hit for each read and limit ambiguous mapping to redundant transcripts. Raw count data
215 for each transcript was extracted using idxstats in samtools (version 0.1.18 (35)) and input
216 into edgeR (45) for differential expression analysis. A criteria of at least one count in a
217 minimum of three individual samples (corresponding to the number of individuals per
218 treatment group) was imposed, and tagwise dispersion was applied with the recommended
219 prior.df =10. Comparisons were initially conducted between the two control groups to ensure
220 that our analysis did not identify differential expression as a result of random variation
221 between groups. Following this initial analysis, comparisons were conducted between the six

222 individual fish from the combined control groups and 3 individuals from each of the other
223 treatment groups. Transcripts were considered differentially expressed with a FDR < 0.05
224 (Benjamini-Hochberg correction). Hierarchical clustering was performed on all differentially
225 expressed transcripts for all samples using an Euclidean distance metric, in the Pheatmap
226 package for R. Functional analysis was then performed for differentially expressed genes
227 from each treatment using the Database for Annotation, Visualisation and Integrated
228 Discovery (DAVID v6.7; (23)), with the newly assembled brown trout male liver transcriptome
229 as a background. KEGG pathways and Gene Ontology (GO) terms for Biological Processes,
230 Cellular Components and Molecular Functions were considered significantly over-
231 represented when P < 0.05.Canonical pathway and network analysis was conducted using
232 Ingenuity Pathways Analysis (IPA; Ingenuity Systems, http://www.ingenuity.com) based on
233 the list of differentially expressed transcripts.
234
235 The raw sequence data, and processed results from the expression analysis have been
236 deposited in NCBI's Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo), and are
237 available via the GEO series accession number GSE57490.
238
239 *RT- QPCR analysis*
240 To validate the results of the differential expression analysis, real time quantitative PCR
241 (RT-QPCR) was used to quantify the expression of a selection of eight transcripts (*vtg1,*
242 *nots, esr1, zp2.5, zp3a.2, crot, tat, tgm2l*) in the liver of all individual fish, including both
243 sexually mature and immature males (n=7, 5, 7 and 4 in the control, 1.9, 18.1 and 34.4 ng
244 E2/L treatment groups, respectively). Transcript expression was also conducted for the
245 remaining, immature female fish (n= 6, 2 and 2 in the control, 1.9, and 34 ng E2/L treatment
246 groups, respectively), and analysis was conducted separately due to the very large sex
247 differences observed for some transcripts (including *vtg1* and *nots*). Primers were designed
248 using Beacon Designer 3.0 (Premier Biosoft International, Paulo Alto, CA), purchased from
249 MWG-Biotech (Ebersburg, Germany) and optimised as previously described (56). The

primer sequences, PCR product sizes, annealing temperatures and PCR efficiencies for each optimised primer pair are shown in Table 1. cDNA was synthesised from 2 µg of total RNA treated with RQ1 DNase (Promega, Southampton, UK) using random hexamers (MWG-Biotech) and M-MLV reverse transcriptase (Promega), according to the manufacturer's instructions. RT-QPCR was performed using 1:2 diluted cDNA in duplicate, using SYBR green chemistry, with an iCycler iQ Real-time Detection System (Bio-Rad Laboratories, Hercules, CA), including a negative control run in duplicate on each plate to verify the absence of cDNA contamination. Efficiency-corrected relative expression levels for each transcript were determined by normalising to a control transcript, *vapa,* which was selected based on its highly consistent expression across all individuals in the RNA-seq dataset.

Statistical analyses of RT-QPCR data were conducted with SigmaStat (version 12.0). Transcript expression data that did not meet normally-distributed criteria was log transformed before statistical analysis. All data was analysed using single factor one way analysis of variance (ANOVA), followed by the Holm-Sidak post hoc test. Data were considered to be significant when $P < 0.05$.

Results

Morphological parameters

Visual examination of the gonads revealed there were 17 mature males, defined by the presence of large white testis and milt, across all treatment groups (n= 3-7 fish per treatment). The remaining fish were sexually immature males and females. The mean mass and length of all mature males was 472.3 ± 9.1 g and 34.3 ± 0.2 cm, of immature males was 430.5 ± 44.5 and 32.6 ± 0.9, and of females was 420.0 ± 15.9 g and 33.5 ± 0.5 cm. The mean condition factor, HSI and GSI were, respectively, 1.16 ± 0.01, 1.11 ± 0.03 and 3.95 ± 0.32 for mature males, 1.24 ± 0.12, 0.85 ± 0.04 and 0.06 ± 0.01 for immature males and 1.11 ± 0.02, 1.00 ± 0.03 and 0.31 ± 0.02 for immature females, and there were no significant

277 differences for these parameters between treatment groups. Additionally, we observed no
278 alteration of general health or behaviour during the exposure period.
279
280 *Sequencing and transcriptome assembly*
281 In total, we sequenced 225.3 million paired 100 bp reads from male brown trout liver
282 samples, and 208.1 million (92.4 %) of these were retained after processing and quality
283 filtering. As described in Uren Webster et al. (57), highly duplicated reads were then
284 removed by digital normalisation, and 46.73 million paired reads were retained for input into
285 the *de novo* transcriptome assembly. The final transcriptome assembly consisted of 172,688
286 transcripts (107,095 loci) with a mean length of 767.5 bp and a N50 of 1292 bp. Of these,
287 62,236 transcripts were annotated using Blastx ($e < e^{-15}$) against Ensembl peptide
288 databases, and these included representation of 16,121 unique zebrafish transcripts.
289
290 *Transcript expression analysis*
291 A total of 137.6 million raw reads were obtained from the libraries generated from liver
292 samples of E2 exposed and control male fish, averaging 9.2 million reads per individual
293 sample, and 83.1 % of these were re-mapped against the transcriptome assembly.
294 Differential expression analysis between the control groups revealed only 3 differentially
295 regulated transcripts. Comparisons between each E2 treatment group and the combined
296 control group were conducted and resulted in only 4 (*bcl6a, spns1,* NM_001124310,
297 uncharacterised transcript) and 2 (NM_001124310, uncharacterised transcript) differentially
298 expressed transcripts for fish exposed to 1.9 and 18.1 ng E2/L, respectively. This may
299 correspond to changes associated with tank effects, given that similar numbers of
300 differentially expressed transcripts were found between the two control tanks and no
301 treatment associated trends were observed for any of these transcripts. Exposure to 34.4 ng
302 E2/L, however, resulted in 2113 differentially expressed transcripts (Figure 1a; Table S1),
303 including 808 unique annotations. Multidimensional scaling (MDS) plots and Euclidean
304 cluster analysis based on all differentially regulated transcripts show that all three individual

fish exposed to 34.4 ng E2/L have a very similar and consistent expression profile, clearly distinct from all other fish, whereas the control fish and those exposed to the lower concentrations of E2 cluster together (Figures 1b and c).

A list of the 20 most up- and down-regulated transcripts following exposure to 34.4 ng E2/L is shown in Table 2. The greatest fold-changes in expression were associated with up-regulated transcripts, and were dominated by well characterised estrogen-responsive genes, including a number of vitellogenin transcripts (*vtg1, vtg1l, vtg2, vtg3, vtg6, vtg7*), of which *vtg1* was the most highly expressed. Additionally, a transcript encoding nothepsin (*nots*), was similarly strongly induced in fish exposed to 34.4 ng E2/L. Transcripts encoding zona pellucida proteins (*zp2.2, zp2.5, zp3a.1, zp3a.2*) were also up-regulated (up to 70-230 fold) and estrogen receptor 1 (*esr1*) was up-regulated by up to 27 fold. There was a trend towards up-regulation of *esr1, zp3a.2* and *zp2.5* (2-4 fold) in the lower treatment groups compared to the control, but these results were not statistically significant, likely due to the low number of replicates in the RNA-Seq dataset.

Analysis of ERCC spike-in control data were conducted to determine the accuracy and dynamic range of the transcript expression measurements in this study. For all individual samples, there was a strong correlation between the calculated FPKM (Fragments Per Kilobase of transcript per Million mapped reads) values and the expected concentration of control transcripts (mean R^2 = 0.902±0.005). The dynamic range was calculated for all samples individually, using the control transcripts that were detected in a minimum of three libraries as the lower cut-off limit. The mean dynamic range in expression level for all 15 libraries was 26,753 FPKM. There was also a good correlation between the calculated and expected changes in transcript expression level between samples spiked with ERCC mix 1 and mix 2 (R^2=0.58). Together, these results provide strong technical validation for the quantitative expression profiling conducted in this study.

Functional analysis

Enriched Gene Ontology terms and KEGG pathways among up- and down-regulated transcripts following exposure to 34.4 ng E2/L are illustrated in Figure 2. GO terms including *translation*, *ribosome*, *lipid metabolic processes* and *growth factor binding* were over-represented in the list of up-regulated transcripts. Regulated transcripts within these Gene Ontologies included RNA polymerases (*polr1a, polr3a)* for *transcription*; translation initiation factors (*eif1ad, eif3s10, eif4a2*) for *translation*; and ribosomal components and binding proteins (*rpl5a, rpl12, rpl15, rpl36a, rpl39, rplp0, rpp21, rps2, rps9, rps23, rpsa, rrbp1a*) for *ribosome.* Within *lipid metabolism*, differentially regulated transcripts included apolipoproteins (*apob, apobb, apof, apoc2*), lipoprotein receptor (*lrpap1*), glycolipid transfer proteins (*gltpd2*) and transcripts involved in PPAR signalling (*ppardb, acoxl*). In addition, insulin-like growth factor (IGF) signalling was also affected and transcripts encoding IGF binding proteins (IGFBPs) were up-regulated in some cases (*igfbp5a, igfbp2a, igfbp2b*) and down-regulated in others (*igfbp1a, igfbp1b*).

For down-regulated transcripts, the most over-represented GO terms related to amino acid metabolism and biosynthesis and associated processes including *organic acid biosynthesis*, *transaminase activity* and *pyridoxal phosphate binding*. Of note, a number of processes involved in cysteine and methionine metabolic pathways were enriched, whereby differentially regulated transcripts included betaine-homocysteine methyltransferase (*bhmt*), S-adenosylmethionine synthase (*sash1),* methionine adenosyltransferase (*mat2aa*) and cysteine dioxygenase (*cdo1*). *Apoptosis* and *programmed cell death* were also over-represented in the list of down-regulated transcripts.

Ingenuity pathway analysis identified a gene network involved in the response to E2 with functions relating to amino acid metabolism, cell death and survival, endocrine system development and small molecule biochemistry, and with *esr1* and the myelocytomatosis oncogene (*myc*) as central nodes (Figure 3). In particular, a number of genes and processes

in this network can be associated with transcription and translation including histones, RNA polymerase and several translation elongation factors.

RT-QPCR validation

RT-QPCR analysis, performed for eight transcripts using all of the male fish, fully validated the results of the RNA-seq expression analysis (Figure 4a). Five up-regulated transcripts (*vtg1, nots, esr1, zp2.5, zp3a.2)* and two down-regulated transcripts (*tat, tgm2l*) were confirmed as being significantly differentially expressed in fish exposed to 34.4 ng/L E2 compared to those in the control group. Furthermore, the expression of *esr1* was also significantly increased in both the 1.9 and 18.1 ng/L groups, while *zp2.5* and *tgm2l* were significantly up- and down-regulated, respectively, following exposure to 18.1 ng/L, confirming the trends identified in the RNA-Seq dataset. For *crot,* there was a significant increase in expression of this transcript in male fish exposed to 1.9 ng/L E2, and increasing, but non-significant, trends in expression in the two higher treatment groups. RT-QPCR analysis was also performed on the same transcripts in the immature female fish, and identified very similar patterns of expression including significant up-regulation of *vtg1,nots, zp2.5, zp3a.2, crot* in the highest treatment group, and significant up-regulation of *esr1* in fish exposed to both 1.9 and 34.4 ng/L E2 (females were not present in the group of fish exposed to 18.1 ng/L; Figure 4b). *tgm2l* expression was only detected using RT-QPCR in sexually mature males (n=15 across all groups) and not in any immature males (n=6) or females (n=10).

Discussion

Despite the ecological and economic importance of brown trout, relatively little is known about the responses of this species to key stressors affecting its freshwater habitat, including endocrine disrupting chemicals. Here, we have conducted global transcriptional profiling using RNA-seq in the liver of sexually mature males exposed to E2, and identified very

389 significant transcriptional changes at the highest concentration tested (34.4 ng E2/L), that
390 were very consistent between the three individual fish in this group. In contrast,
391 concentrations of up to 18.1 ng E2/L did not induce significant changes following 4 days of
392 exposure. RT-QPCR analysis, using a greater number of individual fish per treatment, fully
393 validated the results of the RNA-seq analysis for eight target genes (including both up- and
394 down-regulated transcripts), and also identified significant differences in expression of
395 several of these in the 1.9 ng E2/L group (*esr1* and *crot*) and 18.1 ng E2/L group (*esr1,*
396 *zp2.5* and *tgm2l*). Importantly, this demonstrates that short-term exposures to low,
397 environmentally-relevant concentrations of E2 induce significant changes in the expression
398 of some of the most sensitive estrogen-responsive genes in brown trout. Although 34.4 ng/L
399 is far higher than EEQ concentrations regularly reported in surface waters, it is within a
400 range reported to occur in treated sewage effluent (16, 26), and may be associated with
401 short-term peaks of E2 contamination in streams inhabited by this species that occur as a
402 result of agricultural pollution (51). Therefore, the extent of transcriptional change found in
403 this study after a four day exposure to 34.4 ng E2/L may also be of interest for evaluating the
404 potential impacts of estrogens on populations of brown trout in the most contaminated
405 environments.
406
407 *Conserved estrogen-responsive transcripts*
408 Transcripts encoding six vitellogenin isoforms were very strongly induced in males exposed
409 to 34.4 ng E2/L, but were not detected using RNA-seq in fish exposed to 18.1 ng/L and
410 below. RT-QPCR quantified very low levels of *vtg1* expression in males in the control group,
411 and confirmed a very significant up-regulation in 34.4 ng E2/L in exposed males (by 66,000
412 fold), but no significant difference in expression in the two lower treatment groups. These
413 results are similar to that reported in previous transcriptomic studies where *vtg* transcripts
414 are generally the most strongly up-regulated following estrogen exposure (e.g. 34, 60). The
415 threshold for induction of vitellogenin transcript expression in mature males in this study was
416 slightly higher than previously reported values for transcript and protein induction in

salmonids. In juvenile rainbow trout exposed to E2 for 14 days, the median effective treatment concentration for plasma Vtg protein induction was in the range of 19-26 ng/L (55), while the lowest effective concentration for both plasma Vtg protein and hepatic *vtg1* transcript induction was found to be 14 ng/L (54). In juvenile brown trout, the median EC50 for plasma Vtg protein induction following 7-day E2 exposure was 15 ng/L (5). Here, this relatively higher threshold level for vitellogenin induction compared to previous reports for E2 exposure may reflect the shorter exposure period in this study of only four days and/or differences in maturity status.

Transcripts encoding four zona pellucida proteins and *esr1* were also amongst the most up-regulated transcripts in fish exposed to 34.4 ng E2/L, similarly to previous reports showing strong up-regulation in vitellogenic females and induction by E2 in males (18, 34, 60). There were also non-significant trends towards up-regulation of *zp*s and *esr1* in the lower treatment groups (by 2-4 fold), and RT-QPCR confirmed significant up-regulation of *esr1* (in males and females) from 1.9 ng E2/L, and of *zp2.5* (males only) from 18.1 ng E2/L. This suggests that, compared to vitellogenin, zona pellucida protein and *esr1* transcription is particularly sensitive to estrogen exposure in brown trout, as has been previously reported in other species (18, 30, 54). This corresponds with previous reports that, in the liver, *esr1* is the most responsive estrogen receptor to estrogen exposure, and is associated with vitellogeneis (13, 30). There are reports that *esr2a* and/or *esr2b* have a major role in regulating vitellogenesis (33, 52), but our results showed no significant difference in the expression of either of these transcripts, or any apparent trends towards this, suggesting that within our experiment, the potential involvement of these genes in vitellogenesis was not regulated by E2-induced changes at the transcriptional level.

Novel estrogen-responsive transcripts

The second most significantly up-regulated transcript in fish exposed to 34.4 ng E2/L encoded nothepsin (*nots*). This is a liver-specific aspartic proteinase normally exclusively

445 expressed in the livers of females, and has been linked to the proteolytic cleavage of the
446 vitellogenin precursor (44). Although there are several reports of an increase in nothepsin
447 expression following estrogen exposure in fish (44, 60), it has not been widely considered as
448 a particularly estrogen-responsive gene. The degree of induction observed in all fish treated
449 with 34.4 ng E2/L using RNA-seq, and confirmed in all individuals using RT-QPCR (36,000
450 fold increase for males), is extraordinary, and is in the same order of magnitude as the
451 increase in *vtg1* expression. This suggests that nothepsin could serve as a useful indicator
452 of estrogen exposure. However, similarly to *vtg1,* there was no increase in *nots* expression
453 quantified using RT-QPCR or RNA-Seq in the lower treatment groups, suggesting nothepsin
454 may be less sensitive to lower estrogen exposure concentrations compared to *esr1* and
455 zona pellucida proteins, at least following this short-term 4 day exposure.
456
457 The most down-regulated transcript (by 186 fold) in fish exposed to 34.4 ng/L encoded
458 transglutaminase 2-like (*tgm2l*). This strong decrease in expression was confirmed using
459 RT-QPCR (calculated as 250 fold), and *tgm2l* was also found to be significantly down-
460 regulated in mature males exposed to 18.1 ng/L. Transglutaminases are a family of enzymes
461 responsible for a diverse range of post-translational protein modifications by catalysing the
462 formation of isopeptide bonds (17). *tgm2l* has only been characterised in fish species, and its
463 specific function is not well defined. Here, the striking expression of hepatic *tgm2l* exclusively
464 in sexually mature male trout suggests a role of this gene in male reproductive function. It is
465 possible that, for example, this role may be similar to that of mammalian prostate
466 transglutaminase *tgm4,* which is important in the formation and function of seminal fluid, and
467 subsequently influences male fertility (9), but this gene has not been characterised in the
468 majority of fish species. Furthermore, human transglutaminases *tgm2* and *tgm4* are known
469 to have upstream androgen regulatory elements, and to be regulated by androgen treatment
470 (12, 25), while gonadal *tgm2* was also up-regulated following androgen exposure in juvenile
471 female rainbow trout (3). Recently, we also found reduced expression of *tgm2l* in mature
472 male brown trout exposed to the anti-androgen, linuron (57). Together, this suggests that

tgm2l may be regulated by androgens and therefore susceptible to disruption by estrogens, which disrupt the androgen to estrogen ratio, highlighting a potentially important endocrine biomarker.

Estrogen-regulated hepatic processes

A number of signalling pathways and processes enriched in the list of differentially regulated transcripts can be broadly related to vitellogenesis. Functional analysis revealed enrichment of lipid transport, and also differential regulation of many other transcripts involved in lipid, fatty acid and cholesterol metabolism. These processes have been previously associated with vitellogenesis in females and E2 exposure in male fish (34, 60), and are likely to reflect the incorporation of lipids into vitellogenins as they are synthesised in the liver.

Additionally, we found evidence of altered regulation of cellular signalling pathways involved in the regulation of hepatic cellular growth and proliferation, in particular insulin growth factor (IGF) signalling and myelocytomatosis oncogene (MYC) signalling. Previously, vitellogenesis in maturing females and estrogen-exposed males has been extensively linked with cellular growth and proliferation in the liver. Transcripts encoding IGF binding protein (IGFBP) types 2 and 5 were up-regulated, while those encoding IGFBP type 1 were down-regulated in fish exposed to 34.4 ng E2/L, suggesting regulation of the transport and bioavailability of IGF1 to bind to its receptors at target cells (24). Crosstalk between IGF and estrogen signalling pathways has also been previously demonstrated, and the transcription of IGFBPs is known to be directly regulated by E2 (20, 29). MYC signalling has been proposed as the dominant regulator of estrogen-induced cellular growth, and estrogen exposure induces *myc* transcription via upstream enhancer activation (39). We observed an up-regulation of *myc* by up to 10 fold following exposure to 34.4 ng E2/L. Additionally, pathway analysis highlighted its role as a central regulator, alongside *esr1*, of other differentially expressed genes involved in cell proliferation. Tissue homeostasis depends on a balance between cell death and cell survival, growth and proliferation, which are often controlled by the same interacting

501 signalling pathways, including regulation by MYC and IGFs (41). In parallel, *apoptosis* was
502 among the down-regulated cellular processes, suggesting that E2 exposure induced liver
503 growth and proliferation, and suppressed apoptosis.
504
505 Differential regulation of processes and transcripts involved in methionine and cysteine
506 metabolism were also observed. This pathway plays an important role in regulating DNA
507 methylation, whereby S-adenosylmethionine (SAM) acts as the key methyl group donor.
508 Modulation of DNA methylation has been implicated in cell proliferation and tumourgenesis,
509 and reported to be altered by estrogen exposure (2, 37). Additionally, studies in human cell
510 lines have shown that reactive estrogen metabolites (quinones) bind homocysteine, which is
511 a key intermediate in methionine and cysteine metabolism (14). Plasma concentrations of
512 free homocysteine are also regulated by estrogen, and are lower in women of reproductive
513 age (11). Therefore, a reduction in homocysteine might contribute to the observed
514 differential-regulation of these associated metabolic enzymes.
515
516 Exposure to E2 also resulted in up-regulation of a number of transcripts with roles in
517 transcription and translation, as well as an over-representation processes and pathways
518 involved in their regulation. Furthermore, *ribosome* and *endoplasmic reticulum* were
519 amongst the most enriched GO and KEGG pathway terms. Ribosome biogenesis in
520 response to estrogen exposure has been previously linked to increased cell growth and
521 proliferation, reflecting a general up-regulation of translation (39, 60). In fish, the observed
522 induction of transcription and translation machinery is also likely to reflect the very significant
523 increase in the synthesis and post-translational modification of vitellogenins and zona
524 pellucida proteins. Ribosomal constituent over-expression has been previously reported in
525 male zebrafish exposed to E2 (47) and in female vitellogenic livers (60). Together, induction
526 of the expression of growth regulators, and of transcription and translation pathways,
527 illustrates the very significant stimulatory effect of E2 on cell proliferation and protein
528 synthesis in the livers of male brown trout.

529

530 A number of processes involved in carbohydrate and amino acid metabolism dominate the
531 over-represented GO terms in the list of down-regulated transcripts. In particular, there was
532 a striking down-regulation of transcripts associated with gluconeogenesis, including tyrosine,
533 alanine, aspartate and glutamate metabolic pathways. Transcripts encoding a key
534 gluconeogenic enzyme, tyrosine aminotransferase (*tat*), were amongst the most significantly
535 down-regulated transcripts in the 34.4 ng E2/L exposed fish and this was also confirmed by
536 RT-QPCR analysis. Additionally, a large number of other transcripts involved in
537 gluconeogenesis were also down-regulated, including a transcript encoding the rate limiting
538 enzyme phosphoenolpyruvate carboxylate 1 (*pck1*). A decrease in liver glucose
539 concentration has previously been reported in vitellogenic female fish, as well as immature
540 trout treated with estrogen, and has been shown to be primarily associated with reduced
541 gluconeogenesis rather than increased glucose utilisation (59). In mammalian studies,
542 estrogen signalling has been widely shown to regulate energy metabolism, including through
543 both glucose and lipid metabolic pathways (8, 22, 46). In the present dataset, there was also
544 evidence of up-regulation of lipid synthesis and transport, possibly suggesting a shift towards
545 lipid metabolism as a preferable energy source for vitellogenesis.
546

547 *Application of RNA-seq in ecotoxicology*

548 Together with a number of previous studies (see review by Qian et al. (42)), the present
549 study provides evidence that RNA-seq has very significant potential for mechanistic analysis
550 of chemical exposures in (non-) model organisms, and also offers a number of technical
551 advantages over other global methodologies to measure global transcript expression. Here,
552 we successfully identified highly conserved responses to estrogen, compared to other
553 species, together with several more novel estrogen-responsive transcripts. This highlights
554 the potential for RNA-seq to investigate mechanisms of toxicity for less studied chemical
555 pollutants, and also demonstrates the feasibility of conducting global gene expression
556 profiling in species of environmental interest for which previous sequence information is

limited, without the investment required to develop a specific microarray, or the need to use surrogate model species.

Analysis of spike-in controls provided strong technical validation for the accuracy of the expression analysis, and the mean calculated dynamic range in expression measured in our experimental data based on the quantified expression of control transcripts was 26,753, which far exceeds that typically found in microarray experiments (up to several hundred fold)(6). The very small number of differentially expressed transcripts between control treatments, and also between the control group and groups exposed to 1.9 and 18.1 ng E2/L reflect the ability of the technique to avoid false positives. This is consistent with other studies that have reported that RNA-seq is accurate and reproducible. However, there is also some evidence that the stringent statistical thresholds imposed during RNA-seq analysis can reduce its relative sensitivity, particularly for rare transcripts or those with small fold changes between treatment groups (36, 43, 49). Here, we found evidence that RNA-seq was less sensitive than RT-QPCR (using more individual fish per treatment) for detecting significant changes in expression of transcripts in the lower concentration groups, although clear trends towards differential regulation were apparent in some cases in the RNA-Seq dataset. It is important to note that we used only three individual fish per treatment group for the RNA-Seq analysis and this is likely to have contributed to the low sensitivity of the technique in this study.

Using individual fish within RNA-seq experimental design provides a considerable advantage for statistical power, compared to the use of pooled replicates, and maximising the number of biological replicates is essential to reduce the impact of biological variation between individuals, and likely to increase sensitivity for detection of transcripts with small fold changes in expression. In the present study we analysed three individual fish per treatment group, and we would expect sensitivity to improve considerably with the use of more replicates. Maximising the sequence coverage depth per sample is also likely to increase the

sensitivity of RNA-seq, particularly for rare transcripts, and is therefore equally important to consider. The feasibility of maximising both the number of replicates and coverage in RNA-seq experiments is rapidly improving with developments in sequencing technology, and is likely to considerably improve sensitivity in the future. Overall, our data highlights the potential of RNA-seq as a valuable tool in mechanistic ecotoxicology which, crucially, is not reliant on pre-existing genomic resources for the species of interest.

Acknowlegements

We wish to thank Ronny van Aerle for advice on the bioinformatics, and Audrey Farbos and Konrad Paszkiewicz for facilitating the sequencing experiments. We also thank Charles Tyler, Rod Wilson, Erin Reardon and Jo Rabineau for critical discussions and contributions towards setting up the fish exposure experiments. This work was supported by a Natural Environment Research Council CASE PhD studentship (Grant number NE/I528326/1) and the Salmon & Trout Association. Karen Moore was supported by a Wellcome Trust Institutional Strategic Support Award (WT097835MF).

References

1. **Arukwe A, and Goksøyr A**. Eggshell and egg yolk proteins in fish: hepatic proteins for the next generation: oogenetic, population, and evolutionary implications of endocrine disruption. *Comparative Hepatology* 2: 4-12, 2003.
2. **Baccarelli A, and Bollati V**. Epigenetics and environmental chemicals. *Current opinion in pediatrics* 21: 243-250, 2009.
3. **Baron D, Montfort J, Houlgatte R, Fostier A, and Guiguen Y**. Androgen-induced masculinization in rainbow trout results in a marked dysregulation of early gonadal gene expression profiles. *BMC genomics* 8: 357, 2007.
4. **Benninghoff AD, and Williams DE**. Identification of a transcriptional fingerprint of estrogen exposure in rainbow trout liver. *Toxicological Sciences* 101: 65-80, 2008.
5. **Bjerregaard P, Hansen PR, Larsen KJ, Erratico C, Korsgaard B, and Holbech H**. Vitellogenin as a biomarker for estrogenic effects in brown trout, *Salmo trutta;* laboratory and field investigations. *Environ Toxicol Chem* 27: 2387-2396, 2008.
6. **Black MB, Parks BB, Pluta L, Chu T-M, Allen BC, Wolfinger RD, and Thomas RS**. Comparison of Microarrays and RNA-Seq for Gene Expression Analyses of Dose-Response Experiments. *Toxicological Sciences* 137: 385-403, 2014.
7. **Brown CT, Howe A, Zhang Q, Pyrkosz AB, and Brom TH**. A reference-free algorithm for computational normalization of shotgun sequencing data. *arXiv:12034802 [q-bioGN]* 2012.
8. **Chen J-Q, Brown TR, and Russo J**. Regulation of energy metabolism pathways by estrogens and estrogenic chemicals and potential implications in obesity associated with increased exposure to endocrine disruptors. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* 1793: 1128-1143, 2009.
9. **Dean MD**. Genetic Disruption of the Copulatory Plug in Mice Leads to Severely Reduced Fertility. *PLoS Genet* 9: e1003185, 2013.
10. **Desbrow C, Routledge EJ, Brighty GC, Sumpter JP, and Waldock M**. Identification of estrogenic chemicals in STW effluent.1. Chemical fractionation and in vitro biological screening *Environ Sci Technol* 32: 1549-1558, 1998.
11. **Dimitrova KR, DeGroot K, Myers AK, and Kim YD**. Estrogen and homocysteine. *Cardiovascular research* 53: 577-588, 2002.
12. **Dubbink HJ, de Waal L, van Haperen R, Verkaik NS, Trapman J, and Romijn JC**. The human prostate-specific transglutaminase gene (TGM4): genomic organization, tissue-specific expression, and promoter characterization. *Genomics* 51: 434-444, 1998.
13. **Filby AL, and Tyler CR**. Molecular characterization of estrogen receptors 1, 2a, and 2b and their tissue and ontogenic expression profiles in fathead minnow (Pimephales promelas). *Biol Reprod* 73: 648-655, 2005.
14. **Gaikwad NW**. Mass spectrometry evidence for formation of estrogens-homocysteine conjugates: Estrogens can regulate homocysteine levels. *Free Radical Biology and Medicine 65*: 1447-1454, 2013.
15. **Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, and Zeng Q**. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* 29: 644-652, 2011.
16. **Green C, Williams R, Kanda R, Churchley J, He Y, Thomas S, Goonan P, Kumar A, and Jobling S**. Modelling of steroid estrogen contamination in UK and south Australian rivers predict modest increases in concentrations in the future. *Environmental Science & Technology* 47: 7224-7232, 2013.
17. **Griffin M, Casadio R, and Bergamini C**. Transglutaminases: nature's biological glues. *Biochem J* 368: 377-396, 2002.
18. **Gunnarsson L, Kristiansson E, Farlin L, Nerman O, and Larsson DJ**. Sensitive and robust gene expression changes in fish exposed to estrogen - "a microarray approach. *BMC Genomics* 8: 149-155, 2007.

19. **Hall JM, Couse JF, and Korach KS**. The multifaceted mechanisms of estradiol and estrogen receptor signaling. *J Biol Chem* 276: 36869-36872, 2001.
20. **Hamelers I, and Steenbergh P**. Interactions between estrogen and insulin-like growth factor signaling pathways in human breast tumor cells. *Endocrine-Related Cancer* 10: 331-345, 2003.
21. **Harris CA, Hamilton PB, Runnalls TJ, Vinciotti V, Henshaw A, Hodgson D, Coe TS, Jobling S, Tyler CR, and Sumpter JP**. The consequences of feminization in breeding groups of wild fish. *Environmental Health Perspectives* 119: 306-311, 2011.
22. **Herzog B, Cardenas J, Hall RK, Villena JA, Budge PJ, Giguère V, Granner DK, and Kralli A**. Estrogen-related Receptor α Is a Repressor of Phosphoenolpyruvate Carboxykinase Gene Transcription. *Journal of Biological Chemistry* 281: 99-106, 2006.
23. **Huang DW, Sherman BT, and Lempicki RA**. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protocols* 4: 44-57, 2008.
24. **Hwa V, Oh Y, and Rosenfeld RG**. The insulin-like growth factor-binding protein (IGFBP) superfamily. *Endocrine reviews* 20: 761-787, 1999.
25. **Jia L, Berman BP, Jariwala U, Yan X, Cogan JP, Walters A, Chen T, Buchanan G, Frenkel B, and Coetzee GA**. Genomic Androgen Receptor-Occupied Regions with Different Functions, Defined by Histone Acetylation, Coregulators and Transcriptional Capacity. *PLoS ONE* 3: e3645, 2008.
26. **Jobling S, Burn RW, Thorpe K, Williams R, and Tyler C**. 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: 797-802, 2009.
27. **Jobling S, Coey S, Whitmore J, Kime D, Van Look K, McAllister B, Beresford N, Henshaw A, Brighty G, and Tyler C**. Wild intersex roach (Rutilus rutilus) have reduced fertility. *Biol Reprod* 67: 515-524, 2002.
28. **Jobling S, Nolan M, Tyler CR, Brighty GC, and Sumpter JP**. Widespread Sexual Disruption in Wild Fish. *Environ Sci Technol* 32: 2498-2506, 1998.
29. **Kamangar BB, Gabillard J-C, and Bobe J**. Insulin-like growth factor-binding protein (IGFBP)-1,-2,-3,-4,-5, and-6 and IGFBP-related protein 1 during rainbow trout postvitellogenesis and oocyte maturation: molecular characterization, expression profiles, and hormonal regulation. *Endocrinology* 147: 2399-2410, 2006.
30. **Katsiadaki I, Williams TD, Ball JS, Bean TP, Sanders MB, Wu H, Santos EM, Brown MM, Baker P, and Ortega F**. Hepatic transcriptomic and metabolomic responses in the Stickleback (*Gasterosteus aculeatus*) exposed to ethinyl-estradiol. *Aquatic Toxicology* 97: 174-187, 2010.
31. **Kidd KA, Blanchfield PJ, Mills KH, Palace VP, Evans RE, Lazorchak JM, and Flick RW**. Collapse of a fish population after exposure to a synthetic estrogen. *Proceedings of the National Academy of Sciences* 104: 8897, 2007.
32. **Langmead B, and Salzberg SL**. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9: 357-359, 2012.
33. **Leaños-Castañeda O, and Van Der Kraak G**. Functional characterization of estrogen receptor subtypes, ERα and ERβ, mediating vitellogenin production in the liver of rainbow trout. *Toxicology and Applied Pharmacology* 224: 116-125, 2007.
34. **Levi L, Pekarski I, Gutman E, Fortina P, Hyslop T, Biran J, Levavi-Sivan B, and Lubzens E**. Revealing genes associated with vitellogenesis in the liver of the zebrafish (Danio rerio) by transcriptome profiling. *BMC Genomics* 10: 141-158, 2009.
35. **Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, and Durbin R**. The sequence alignment/map format and SAMtools. *Bioinformatics* 25: 2078-2079, 2009.
36. **McCarthy DJ, Chen Y, and Smyth GK**. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Research* 40: 4288-4297, 2012.

37. **Mirbahai L, Southam AD, Sommer U, Williams TD, Bignell JP, Lyons BP, Viant MR, and Chipman JK**. Disruption of DNA Methylation via S-Adenosylhomocysteine Is a Key Process in High Incidence Liver Carcinogenesis in Fish. *Journal of Proteome Research* 12: 2895-2904, 2013.
38. **Moggs JG, Ashby J, Tinwell H, Lim FL, Moore DJ, Kimber I, and Orphanides G**. The need to decide if all estrogens are intrinsically similar. *Environmental health perspectives* 112: 1137, 2004.
39. **Musgrove EA, Sergio CM, Loi S, Inman CK, Anderson LR, Alles MC, Pinese M, Caldon CE, Schatte J, and Gardiner-Garden M**. Identification of functional networks of estrogen-and c-Myc-responsive genes and their relationship to response to tamoxifen therapy in breast cancer. *PLoS ONE* 3: e2987, 2008.
40. **Nagler JJ, Ruby SM, Idler DR, and So YP**. Serum phosphoprotein phosphorus and calcium levels as reproductive indicators of vitellogenin in highly vitellogenic mature female and estradiol-injected immature rainbow trout (Salmo gairdneri). *Canadian Journal of Zoology* 65: 2421-2425, 1987.
41. **Prendergast GC**. Mechanisms of apoptosis by c-Myc. *Oncogene* 18: 2967-2987, 1999.
42. **Qian X, Ba Y, Zhuang Q, and Zhong G**. RNA-Seq Technology and Its Application in Fish Transcriptomics. *Omics: a journal of integrative biology* 18: 98-110, 2014.
43. **Rapaport F, Khanin R, Liang Y, Pirun M, Krek A, Zumbo P, Mason CE, Socci ND, and Betel D**. Comprehensive evaluation of differential gene expression analysis methods for RNA-seq data. *Genome Biol* 14: R95, 2013.
44. **Riggio M, Scudiero R, Filosa S, and Parisi E**. Oestrogen-induced expression of a novel liver-specific aspartic proteinase in Danio rerio (zebrafish). *Gene* 295: 241-246, 2002.
45. **Robinson MD, McCarthy DJ, and Smyth GK**. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26: 139-140, 2010.
46. **Ropero AB, Alonso-Magdalena P, Quesada I, and Nadal A**. The role of estrogen receptors in the control of energy and glucose homeostasis. *Steroids* 73: 874-879, 2008.
47. **Ruggeri B, Ubaldi M, Lourdusamy A, Soverchia L, Ciccocioppo R, Hardiman G, Baker ME, Palermo F, and Polzonetti-Magni AM**. Variation of the genetic expression pattern after exposure to estradiol-17 beta and 4-nonylphenol in male zebrafish (Danio rerio). *General and Comparative Endocrinology* 158: 138-144, 2008.
48. **Segner H, Casanova-Nakayama A, Kase R, and Tyler CR**. Impact of environmental estrogens on fish considering the diversity of estrogen signaling. *General and Comparative Endocrinology* 191: 190-201, 2013.
49. **SEQC/MAQC-III_Consortium**. A comprehensive assessment of RNA-seq accuracy, reproducibility and information content by the Sequencing Quality Control Consortium. *Nat Biotechnol* 32: 903-914, 2014.
50. **Shappell NW, Elder KH, and West M**. Estrogenicity and nutrient concentration of surface waters surrounding a large confinement dairy operation using best management practices for land application of animal wastes. *Environmental Science & Technology* 44: 2365-2371, 2010.
51. **Shore LS, and Shemesh M**. Naturally produced steroid hormones and their release into the environment. *Pure and Applied Chemistry* 75: 1859-1871, 2003.
52. **Soverchia L, Ruggeri B, Palermo F, Mosconi G, Cardinaletti G, Scortichini G, Gatti G, and Polzonetti-Magni A**. Modulation of vitellogenin synthesis through estrogen receptor beta-1 in goldfish (Carassius auratus) juveniles exposed to 17-β estradiol and nonylphenol. *Toxicology and applied pharmacology* 209: 236-243, 2005.
53. **Supek F, Bošnjak M, Škunca N, and Šmuc T**. REVIGO summarizes and visualizes long lists of Gene Ontology terms. *PLoS ONE* 6: e21800, 2011.
54. **Thomas-Jones E, Thorpe K, Harrison N, Thomas G, Morris C, Hutchinson T, Woodhead S, and Tyler C**. Dynamics of estrogen biomarker responses in rainbow trout exposed to 17 beta-estradiol and 17 alpha-ethinylestradiol. *Environ Toxicol Chem* 22: 3001-3008, 2003.

55. **Thorpe KL, Cummings RI, Hutchinson TH, Scholze M, Brighty G, Sumpter JP, and Tyler CR**. Relative potencies and combination effects of steroidal estrogens in fish. *Environmental Science & Technology* 37: 1142-1149, 2003.
56. **Uren Webster TM, Laing LV, Florance H, and Santos EM**. Effects of Glyphosate and its Formulation, Roundup, on Reproduction in Zebrafish (Danio rerio). *Environmental science & technology* 48: 1271-1279, 2014.
57. **Uren Webster TM, Perry MH, and Santos EM**. The Herbicide Linuron Inhibits Cholesterol Biosynthesis and Induces Cellular Stress Responses in Brown Trout. *Environmental Science & Technology* 49: 3110-3118, 2015.
58. **van Aerle R, Nolanusan M, Jobling S, Christiansen LB, Sumpter JP, and Tyler CR**. Sexual disruption in a second species of wild cyprinid fish (the gudgeon, Gobio gobio) in United Kingdom freshwaters. *Environ Toxicol Chem* 20: 2841-2847, 2001.
59. **Washburn BS, Bruss ML, Avery EH, and Freedland RA**. Effects of estrogen on whole animal and tissue glucose use in female and male rainbow trout. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology* 263: R1241-R1247, 1992.
60. **Zheng W, Xu H, Lam SH, Luo H, Karuturi RKM, and Gong Z**. Transcriptomic Analyses of Sexual Dimorphism of the Zebrafish Liver and the Effect of Sex Hormones. *PLoS ONE* 8: 1-12, 2013.

Figure Legends

Figure 1. Differentially expressed transcripts following exposure to E2 in the liver of mature male brown trout. Multiple transcripts are included for each gene annotation, which potentially reflect the presence of different isoforms as well as redundant fragments within the list of differentially expressed transcripts. **A)** Number of up-regulated and down-regulated transcripts in each treatment group calculated using EdgeR (FDR <0.05). **B)** Multidimensional scaling plot illustrating the very significant effect of exposure to 34.4 ng E2/L on the hepatic transcriptome of male brown trout (presented within the blue circle, for visualisation purposes) compared to all other groups, based on the expression of all differentially-regulated transcripts. Individual fish are represented by the following codes: c1, c2, c3 c4, c5 and c6 represent the control individuals; le1, le2 and le3 represent individuals exposed to 1.9 ng E2/L; me1, me2 and me3 represent individuals exposed to 18.1 ng E2/L; he1, he2 and he3 represent individuals exposed to 34.4 ng E2/L. **C)** Heatmap illustrating the relative expression level of all differentially-regulated transcripts in all individual samples (individuals are represented by the same codes as in B). Data presented are log10 transformed read counts per transcript. The hierarchical clustering to generate gene and condition trees was conducted using a Euclidean distance metric in the R pheatmap package.

Figure 2. Over-represented Gene Ontology Terms and KEGG Pathways ($P < 0.05$) in the list of **A)** up-regulated and **B)** down-regulated transcripts in fish exposed to 34.4 ng/L E2. Values presented represent the P-value associated with over-representation. Darker shaded bars indicate GO terms where the adjusted P value was < 0.05 (Benjamini-Hochberg correction). Analysis was conducted using the Database for Annotation, Visualization and Integrated Discovery (DAVID) (Huang et al. 2008) v6 .7, using our brown trout liver transcriptome as a background, and using Reduce and Visualise Gene Ontology (Revigo)(53) to condense redundant terms.

828 **Figure 3.** Enriched gene network constructed using differentially expressed transcripts (FDR
829 <0.05) following exposure to 34.4 ng E2/L. This was the highest scoring enriched network
830 generated by IPA using default settings. Associated functions of this network include amino
831 acid metabolism, cell death and survival, endocrine system development and small molecule
832 biochemistry. Nodes represent genes and edges represent gene relationships. The intensity
833 of node shading represents degree of up-regulation (red) or down-regulation (green), while
834 uncoloured nodes represent genes that were not identified as being differentially expressed
835 in our experiment but were included in this network based on evidence stored in IPA
836 databases. Node shapes denote enzymes, phosphatases, kinases, peptidases, G-protein
837 coupled receptor, transmembrane receptor, cytokines, growth factor, ion channel,
838 transporter, translation factor, nuclear receptor, transcription factor and other. EEF1E1;
839 eukaryotic translation elongation factor 1 epsilon 1, EPRS; glutamyl-prolyl-tRNA synthetase,
840 HSPA8; heat shock 70kDa protein 8, GRWD1; glutamate-rich WD repeat containing 1, CKB;
841 creatine kinase, brain, TUB84B; tubulin alpha 1-like protein, MCM8; minichromosome
842 maintenance complex component 8, EZH2; enhancer of zeste homolog 2, TRIM35; tripartite
843 motif containing 35, EZH1; enhancer of zeste homolog 1, ATAD3A/ATAS3B; ATPase family,
844 AAA domain containing 3A/3B, GRIK5; glutamate receptor, ionotropic, kainate 5, KDM4A;
845 lysine (K)-specific demethylase 4A, Hsp90; 90-kDa heat shock protein, ESR1; estrogen
846 receptor 1, UBALD1; UBA-like domain containing 1, NIPBL; Nipped-B homolog, MYC; v-myc
847 avian myelocytomatosis viral oncogene homolog, ASS1; argininosuccinate synthase 1,
848 RPS23; ribosomal protein S23, NOC2L; nucleolar complex associated 2 homolog, YBX1; Y
849 box binding protein 1, NCL; nucleolin POLR1A; polymerase (RNA) I polypeptide A, 194kDa,
850 SUDS3; suppressor of defective silencing 3 homolog, RBMS2; RNA binding motif, single
851 stranded interacting protein 2, SRP72; signal recognition particle 72kDa.

852 **Figure 4.** Transcript profile analysis for a selection of target genes in all males **(A)** and
853 females **(B),** conducted using RT-QPCR**.** Data are presented as mean fold change relative
854 to expression in the control group. Relative expression was calculated as ratio of the

855 expression for each target gene / expression for *vapa* mRNA. Expression of *tgm2l* was only
856 quantified in mature males and was below the detection limit of the QPCR assay for both
857 immature males and females (n/d: non detectable). Data was collected from 3-7 males and
858 2-6 females per treatment group. Individuals for which the expression was below the
859 detection limit of the assay were excluded from the analysis. Asterisks represent significant
860 differences between each treatment group and the control group (*P<0.05 **P<0.01
861 ***P<0.001).

862

863

864

865

866

867

868

869

870

871

872

873

874

875

876 **Table 1:** Target genes, primer sequences and assay details for RT-QPCR analysis.

Target gene	Symbol	Forward Primer (5'-3')	Reverse Primer (5'-3')	Product size (bp)	Ta (°C)	PCR efficiency (%)
Vesicle-associated membrane protein-associated protein A	*vapa*	CACTGAACATTCCAACTC	TGAGCATTGATAACAGGT	118	59.5	95.5
Estrogen receptor 1	*esr1*	GCAGAACACTTCACAGCATT	ATCCACATAACAGCGACAGA	126	59.5	101.9
Carnitine O-octanoyltransferase	*crot*	GCTGGTAATGTGGTGTTG	ATGGTATCCTTGGTGACTC	83	53.5	105.0
Nothepsin	*nots*	ATGATGACAGGAGGTGAA	AGGAAGGAAAGAAGGAAGA	86	58.0	114.0
Tyrosine aminotransferase	*tat*	AGCATCGTAATCCTAGCAAGA	TCAAGCACCAGCACAGAT	83	56.0	95.7
Transglutaminase 2 like	*tgm2l*	CTGCCACCTAAACACAAA	ATCCAACACCTTCACAAC	75	56.0	99.9
Vitellogenin 1	*vtg1*	AACTTGATTGGAATTGAG	TAATACCTACTTGCTGAA	132	55.0	111.5
Zona pellucida glycoprotein 2.5	*zp2.5*	ATCAATAACCACAGCCACAATG	ACCAGGGACAGCCAATATG	75	55.0	101.2
Zona pellucida glycoprotein 3a.2	*zp3a.2*	AACTACACTCCACTTCATC	CACATCTCCTTCATCTTCA	86	54.5	112.6

877

878

879

Table 2. List of the 20 most up-regulated and down-regulated transcripts in fish exposed to 34.4 ng E2/L.

UP-REGULATED				DOWN-REGULATED			
Symbol	**Name**[a]	**Fold change**[b]	**FDR**	**Symbol**	**Name**[a]	**Fold change**[b]	**FDR**
vtg1	vitellogenin 1	↑ >5438	4.6E-119	*tat*	tyrosine aminotransferase	↓ 4.6	8.9E-9
nots	nothepsin	↑ >4475	5.4E-107	*tgm2l*	transglutaminase 2, like	↓ 186.1	3.3E-8
vtg6	vitellogenin 6	↑ >2000	1.3E-102	*cbln8*	cerebellin 8	↓ 5.7	2.6E-7
vtg2	vitellogenin 2	↑ >1100	7.3E-92	*hsd3b7*	hydroxy-delta-5-steroid dehydrogenase, 3beta- and steroid delta-isomerase	↓ 5.4	2.8E-7
si:dkey-4c23.3	vitellogenin 1-1	↑ >220	9.6E-59	*errfi1*	ERBB receptor feedback inhibitor 1	↓ 6.3	5.5E-7
vtg3	vitellogenin 3	↑ >825	1.4E-56	*igfbp1a*	insulin-like growth factor binding protein 1a	↓ 50.7	6.2E-7
zp3a.2	zona pellucida 3a.2	↑ 149	1.6E-53	*slc3a2a*	solute carrier family 3, member 2a	↓ 10.4	6.8E-7
si:dkey-179j5.2	family with sequence similarity 20, member C	↑ >185	5.0E-52	*faxdc2*	chromosome 5 open reading frame 4	↓ 4.8	1.8E-6
zp2.5	zona pellucida 2.5	↑ 77.6	1.1E-42	*pnp5a*	purine nucleoside phosphorylase 5a	↓ 27.1	2.3E-6
zp3a.1	zona pellucida 3a.1	↑ 161.5	1.1E-42	*epha8*	eph receptor A8	↓ >21	7.5E-6
vtg7	vitellogenin 7	↑ >107	8.3E-42	*pfkfb1*	6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 1	↓ 30.1	8.1E-6
crot	carnitine o-octanoyltransferase	↑ 54.4	4.9E-40	*pptc7a*	PTC7 protein phosphatase homolog a	↓ 6.3	9.2E-6
esr1	estrogen receptor 1	↑ 25.7	1.6E-37	si:dkey-238o13.4	si:dkey-238o13.4	↓ 4.5	1.8E-5
zp2.2	zona pellucida 2.2	↑ 160.7	5.1E-31	*st3gal3b*	ST3 beta-galactoside alpha-2,3-sialyltransferase 3b	↓ 3.8	2.4E-5
aqp12	aquaporin 12	↑ 28.6	8.4E-31	*ret*	ret proto-oncogene receptor tyrosine kinase	↓ 8.5	4.3E-5
lrrc58b	leucine rich repeat containing 58b	↑ 20.6	1.0E-30	*ntng2a*	netrin g2a	↓ 4.9	4.9E-5
igfbp5a	insulin-like growth factor binding protein 5a	↑ >49	2.6E-30	*ulk1a*	unc-51-like kinase 1a	↓ 4.7	4.9E-5
rdh10a	retinol dehydrogenase 10a	↑ >108	8.8E-28	*grb7*	growth factor receptor-bound protein 7	↓ 26.3	1.2E-4
slc7a11	solute carrier family 7, member 11	↑ >51	8.9E-27	*cldn11a*	claudin 11a	↓ 6.5	1.3E-4
lpgat1	lysophosphatidylglycerol acyltransferase 1	↑ 26.0	3.0E-25	*slc25a29*	solute carrier family 25, member 29	↓ 4.8	1.4E-4

a where there were multiple differentially regulated transcripts assigned the same annotation, only the most significantly regulated transcript is included in this list. ***b*** for transcripts where no read counts were detected in any of the individuals in one of the groups, a nominal value of 1 count was given to each individual in that group in order to calculate a fold change value, for visualisation purposes, in this table.

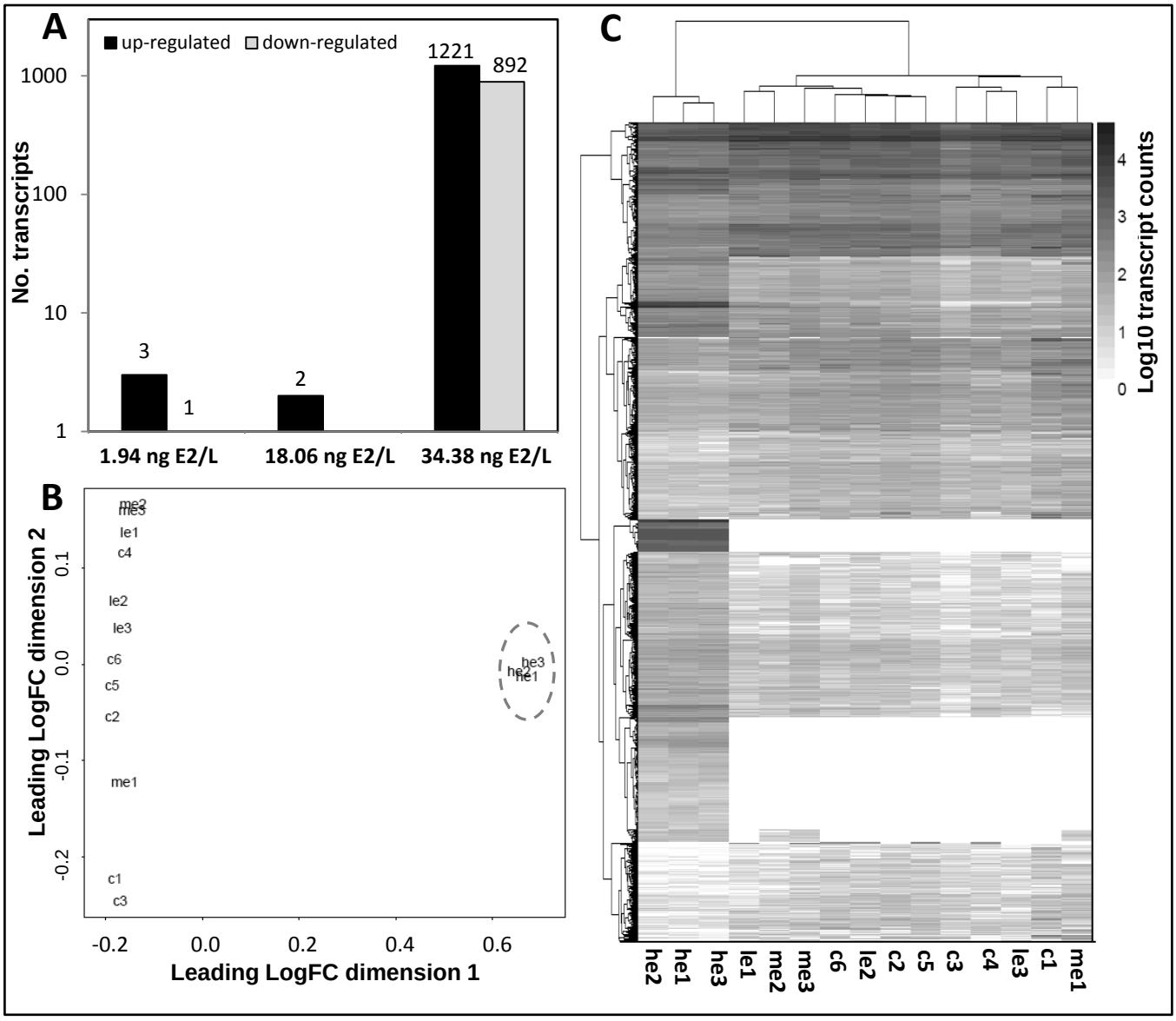
A
up-regulated
down-regulated
No. transcripts
1000
100
10
1
3
1
2
1221
892
1.94 ng E2/L
18.06 ng E2/L
34.38 ng E2/L
B
Leading LogFC dimension 2
Leading LogFC dimension 1
me2
me3
le1
c4
le2
le3
c6
c5
c2
me1
c1
c3
he3
he2
he1
C
Log10 transcript counts
he2
he1
he3
le1
me2
me3
c6
le2
c2
c5
c3
c4
le3
c1
me1

Figure 1

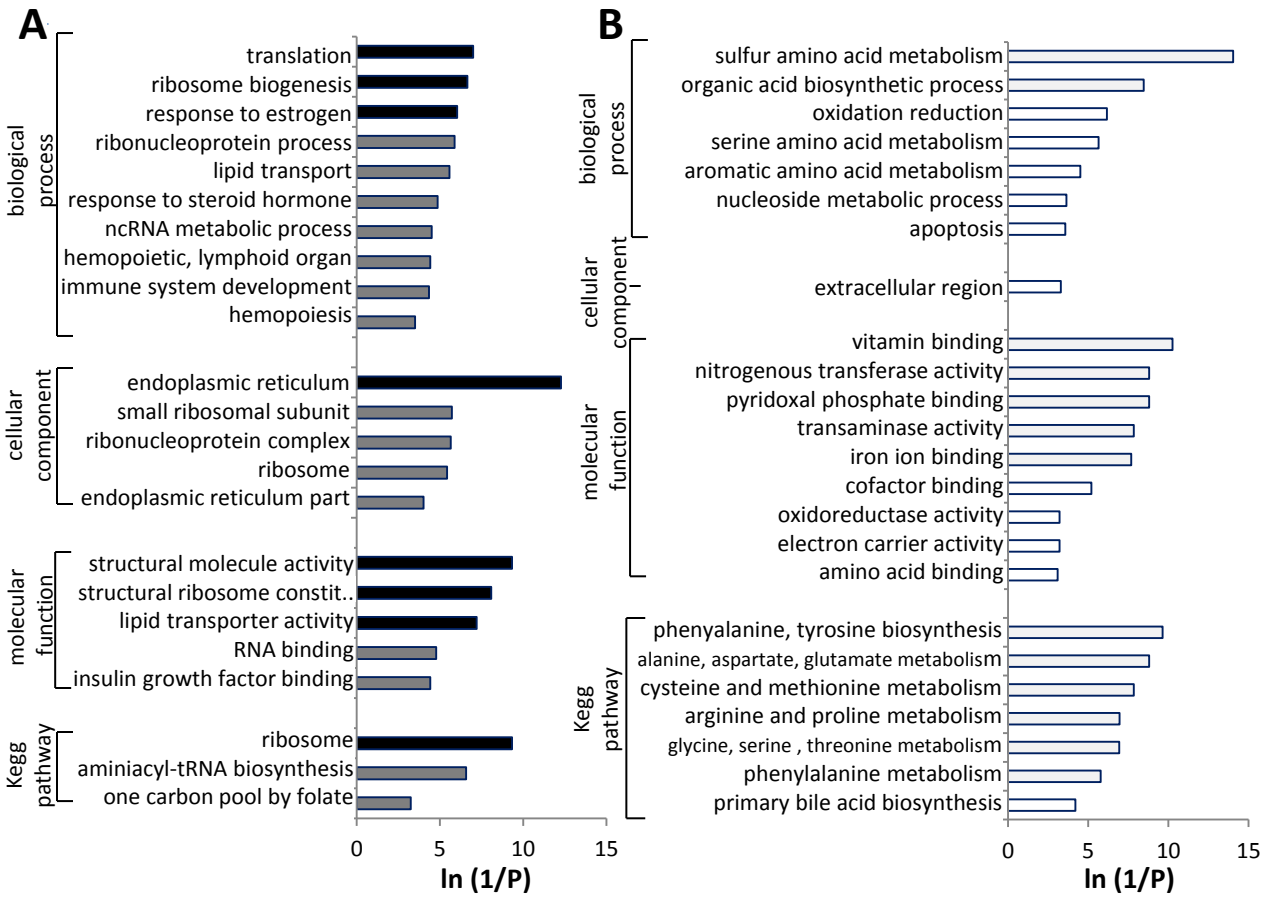
A
biological process
translation
ribosome biogenesis
response to estrogen
ribonucleoprotein process
lipid transport
response to steroid hormone
ncRNA metabolic process
hemopoietic, lymphoid organ
immune system development
hemopoiesis
cellular component
endoplasmic reticulum
small ribosomal subunit
ribonucleoprotein complex
ribosome
endoplasmic reticulum part
molecular function
structural molecule activity
structural ribosome constit..
lipid transporter activity
RNA binding
insulin growth factor binding
Kegg pathway
ribosome
aminiacyl-tRNA biosynthesis
one carbon pool by folate
0
5
10
15
ln (1/P)
B
biological process
sulfur amino acid metabolism
organic acid biosynthetic process
oxidation reduction
serine amino acid metabolism
aromatic amino acid metabolism
nucleoside metabolic process
apoptosis
cellular component
extracellular region
molecular function
vitamin binding
nitrogenous transferase activity
pyridoxal phosphate binding
transaminase activity
iron ion binding
cofactor binding
oxidoreductase activity
electron carrier activity
amino acid binding
Kegg pathway
phenyalanine, tyrosine biosynthesis
alanine, aspartate, glutamate metabolism
cysteine and methionine metabolism
arginine and proline metabolism
glycine, serine , threonine metabolism
phenylalanine metabolism
primary bile acid biosynthesis
0
5
10
15
ln (1/P)

Figure 2

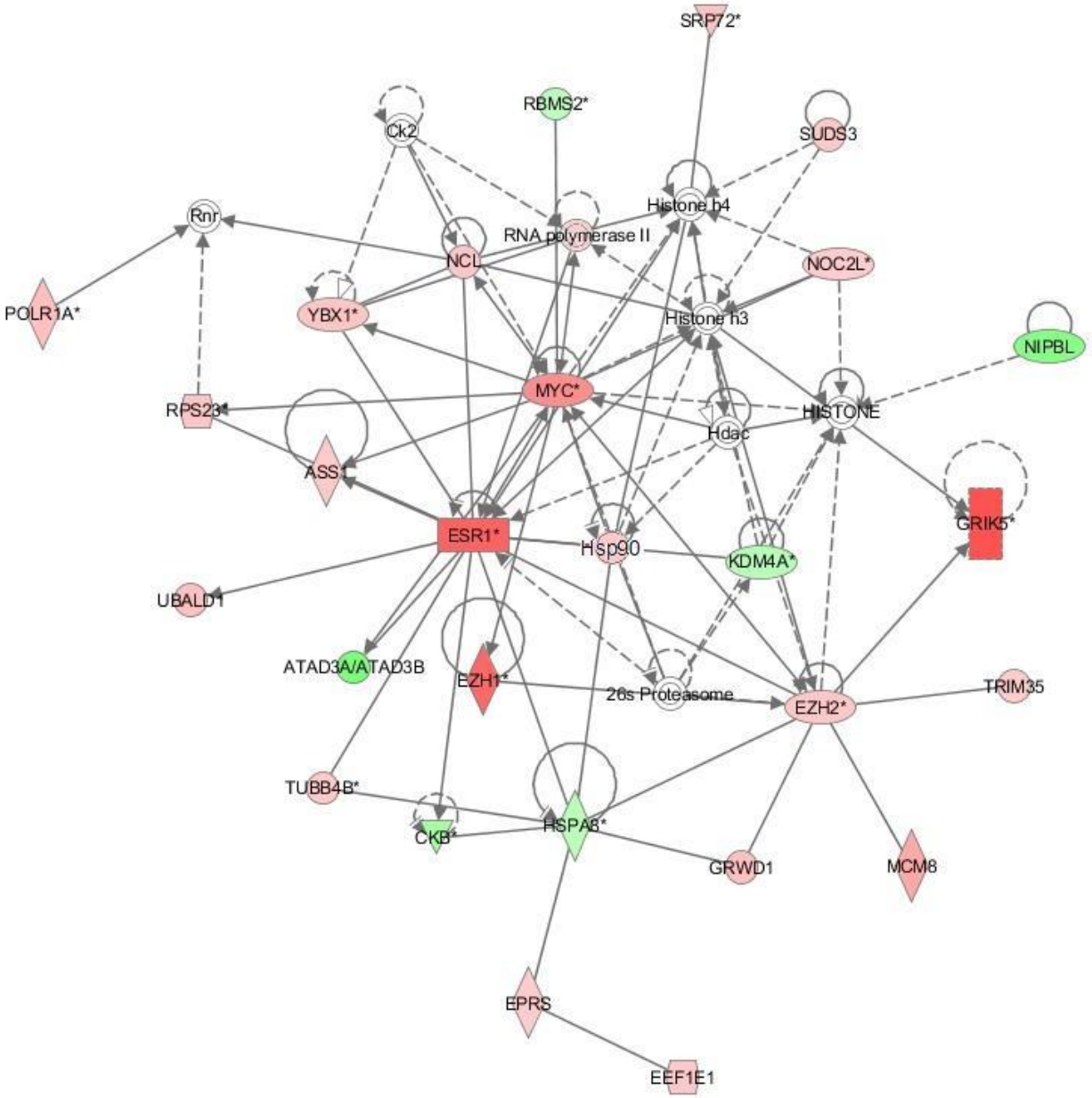
SRP72*
RBMS2*
Ck2
SUDS3
Rnr
Histone h4
RNA polymerase II
NCL
NOC2L*
POLR1A*
YBX1*
Histone h3
NIPBL
MYC*
RPS23*
HISTONE
Hdac
ASS1
GRIK5*
ESR1*
Hsp90
KDM4A*
UBALD1
ATAD3A/ATAD3B
EZH1*
26s Proteasome
TRIM35
EZH2*
TUBB4B*
CKB*
HSPA8*
GRWD1
MCM8
EPRS
EEF1E1

Figure 3

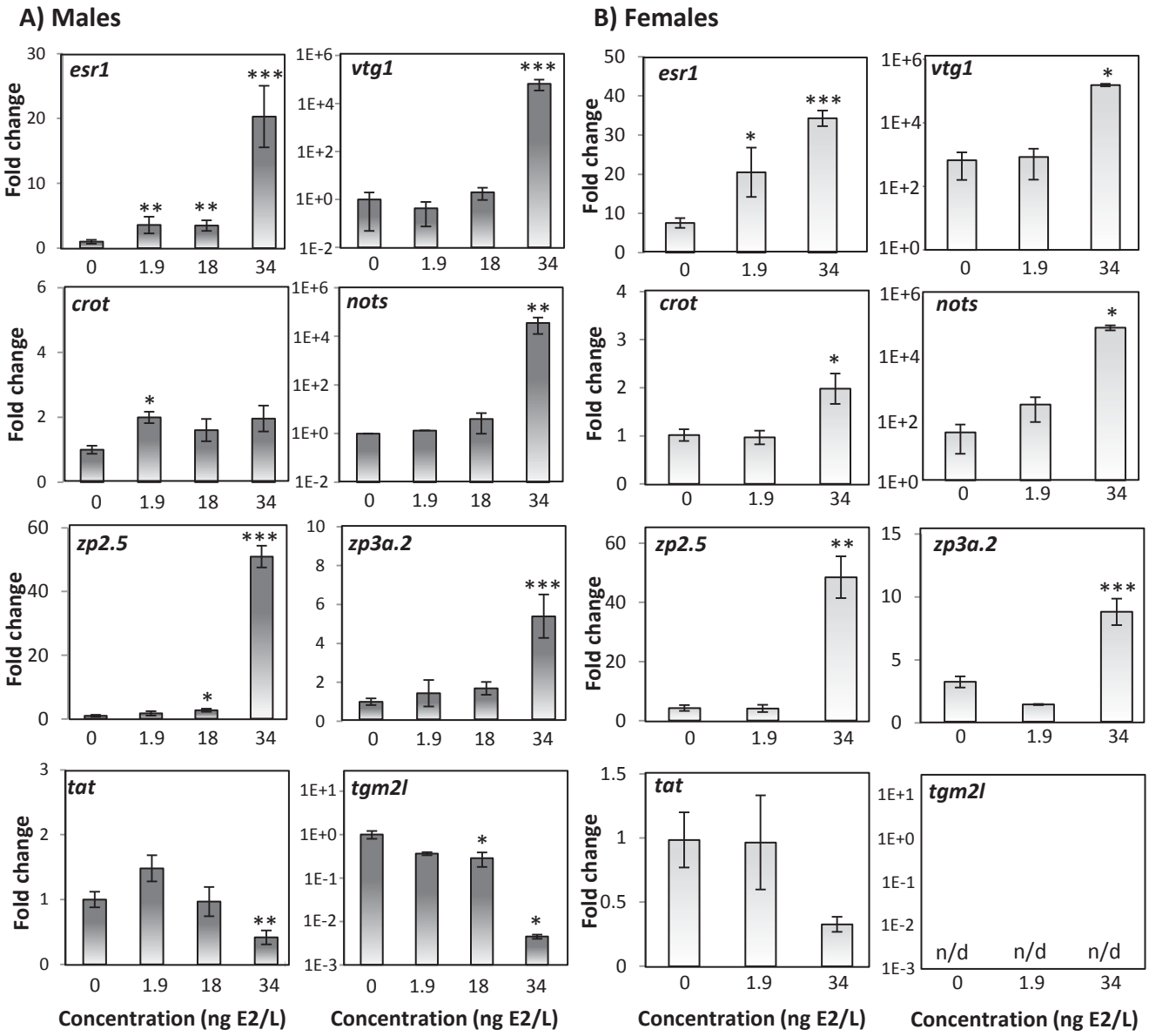
A) Males
B) Females
esr1
vtg1
crot
nots
zp2.5
zp3a.2
tat
tgm2l
Fold change
Concentration (ng E2/L)
0
1.9
18
34
n/d

Figure 4