Sex-specific and conserved gene co-expression networks underlie phenotypes within the paleopolyploid *Salvelinus* genus (Family Salmonidae)

Ben J. G. Sutherland[1†*], Jenni M. Prokkola[2], Céline Audet[3], and Louis Bernatchez[1]

[1] Institut de Biologie Intégrative et des Systèmes (IBIS), Université Laval, Québec, QC G1V 0A6, Canada

[2] Institute of Integrative Biology, University of Liverpool, L69 7ZB Liverpool, UK

[3] Institut des Sciences de la Mer de Rimouski, Université du Québec à Rimouski, Rimouski, QC G5L 3A1, Canada

[†]Current Address: Pacific Biological Station, Fisheries and Oceans Canada, Nanaimo, BC, V9T 6N7, Canada

*Author for correspondence: BJGS

Pacific Biological Station, Fisheries and Oceans Canada, Nanaimo, BC, Canada

Phone: 1 (250) 751-4676

Email: ben.sutherland.1@ulaval.ca (BS)

Data Deposition: Brook Charr raw sequence data has been uploaded to SRA under BioProject PRJNA445826, accession SRP136537. Arctic Charr raw sequence data is available under BioProject PRJNA307980, accession SRP068854.

ABSTRACT

Networks of co-expressed genes produce complex phenotypes that are associated with sexual dimorphism, tissue-specificity, and functional novelty. Modules of potentially co-regulated genes can be compared across individuals to identify differences associated with phenotypic divergence. Here we use RNA-sequencing in the paleopolyploid salmonid Brook Charr *Salvelinus fontinalis* to characterize sex-specific co-expression networks within the liver of 100 full-sib offspring. Module conservation in each sex-specific network was tested in the alternate sex and in the congener Arctic Charr *Salvelinus alpinus*. Modules were further characterized by functional and chromosomal enrichment, hub gene identification, and correlation with 15 unique growth, reproductive and stress-related phenotypes. The majority of modules were conserved between sexes and species, including those involved in conserved cellular processes (e.g. translation, immune pathways). The few identified sex-specific modules may contribute to sexual dimorphism and resolution of sexual antagonism through gene regulation, including male-specific modules correlated with fish size or enriched for transcription factor activity. Most modules, including sex-specific modules, contained genes randomly dispersed throughout the genome, although several were overrepresented on specific chromosomes, suggesting an infrequent, but existing, chromosomal-based co-regulation. This comprehensive analysis provides network and chromosomal context to give insight into the transcriptome regulatory architecture of the paleopolyploid salmonids.

Keywords: Chromosome; Co-regulation; Salmonid; Transcriptomics; Whole Genome Duplication; Weighted Gene Co-expression Network Analysis (WGCNA)

INTRODUCTION

Characterizing the genes and alleles that contribute to phenotypic variation is a central goal in genetics, and can be used to identify mechanisms underlying local adaptation or markers used for selective breeding. Typically this goal is addressed using association studies to identify genomic regions contributing to phenotypes (Mackay 2001; Bush and Moore 2012). Associating genotypes directly to phenotypes bypasses important intermediate regulatory steps such as gene transcription, which can be addressed in part with expression QTL (eQTL) mapping (Mackay et al. 2009). However, when considering the transcriptome, for example in eQTL studies, it is also important to consider the underlying structure of the network in which genes are co-regulated (Mähler et al. 2017).

Network context can aid in understanding gene-specific selective constraints based on network statistics (Mähler *et al.* 2017). Additionally, putting genes into a network can reduce many genes into a smaller set of modules that can be characterized for phenotype correlations (Filteau *et al.* 2013; Rose *et al.* 2015). Comparative network analyses can provide insight on sex differences (Langfelder *et al.* 2011), tissue differences, and even potential drivers of phenotypic change associated with adaptive divergence that may lead to ecological speciation (Filteau *et al.* 2013; Thompson *et al.* 2015). Comparing modules across species can advance the understanding of species-specific innovations. For example, comparison of human and chimp brain transcriptome networks identified low conservation for modules found in specific brain regions associated with human evolution such as the cerebral cortex (Oldham *et al.* 2006). In addition, cross-species gene co-expression analysis has been used to detect gene modules associated with disease (Mueller *et al.* 2017) or with seasonal phenotypic changes (Cheviron and Swanson 2017). These insights are often not possible to obtain through standard differential expression analysis, which captures a smaller proportion of the variation than differential co-expression analysis (Oldham *et al.* 2006; Gaiteri *et al.* 2014).

The development of sex differences in phenotypes and behavior can be attributed to differences in the expression of both sex-specific and autosomal genes, which can be mediated by both hormonal and epigenetic regulation (Ellegren and Parsch 2007; Wijchers and Festenstein 2011). Several aspects of sexual dimorphism remain under investigation, including the role of co-expression networks in solving sexual antagonism (i.e. conflicting selection pressures on phenotypes in each sex). Differing structure of networks between the sexes could provide a solution to sexual antagonism through gene regulation. Other genetic architecture solutions to this conflict may include sex-dependent dominance (Barson *et al.* 2015) or maintaining alleles associated with sexual antagonism on sex chromosomes (Blackmon and Brandvain 2017). However, the analyses of co-expression networks are complicated by the observation that sexual dimorphism in networks can be tissue-specific. In mouse *Mus musculus*, modules found in the liver and adipose tissue were more different between sexes than those in the brain and muscle (van Nas *et al.*

2009). Similarly, networks were very similar between sexes in zebrafish *Danio rerio* brain (Wong *et al.* 2014). Although the extent of differences may depend on the tissue of study, network comparisons between sexes can provide new insight into the regulatory underpinnings of sexual dimorphism and antagonism.

Constructing gene co-expression networks is often based on correlating transcript abundance (Langfelder and Horvath 2008). A network is comprised of nodes (i.e. genes) and their adjacencies (i.e. correlations with other genes), which together form modules (i.e. groups of correlated genes). Genes within a module may have similar functions or regulatory pathways, although this is not always the case (Gillis and Pavlidis 2012; van Dam *et al.* 2017). After network construction, genes that are highly connected and central to a module can be designated as hub genes. Hub genes may be more related to the function of the module than other genes within the module (van Dam *et al.* 2017), they can be more related to phenotypes correlated with the module (Langfelder *et al.* 2013), and can also be upstream regulators of other genes within the module. Hub genes can also be under higher selective constraint, and as a result may show lower genetic variation and higher phylogenetic conservation (Mähler *et al.* 2017). In summary, network information provides novel insight into both gene activity and evolution.

Functional characterization of modules can be achieved by correlating module principal components, i.e., "eigengenes", with phenotypes of profiled individuals. Together with functional enrichment of modules, these correlations can provide some information on unknown genes in non-model systems (Pavey et al. 2012). Likewise, co-expression metrics can provide information about the stimuli under which a gene responds, or the other annotated genes with which the unknown gene is correlated. When transcriptomics is used to classify samples based on a panel of co-expressed genes, it is not necessary to know the function of all of the genes within the panel. For example, the viral disease development (VDD) marker suite was identified to classify the probable viral infection state of Pacific salmon, and the larger panel of genes within this suite contains unknown genes (Miller *et al.* 2017). Considering that the unannotated genes respond together with antiviral response genes, this is valuable information regarding the potential function of the unannotated genes for other studies.

Network evolution after a large-scale genome perturbation, such as whole genome duplication, remains poorly understood in vertebrates, and the structure of existing networks after this event may provide insight into the evolution of novel phenotypes. The salmonids (Family Salmonidae) are a valuable system for studying genome evolution post genome duplication given that the salmonid ancestor underwent a whole genome duplication approx. 60-88 million years ago (Allendorf and Thorgaard 1984; Crête-Lafrenière *et al.* 2012; Macqueen and Johnston 2014). The duplicated genome was then retained and largely rediploidized prior to the salmonid lineage diversification, although some regions remain in a residual tetraploid state (Kodama *et al.* 2014; Lien *et al.* 2016). Within this context, characterizing

networks of genes in the salmonids not only will provide new information on specific genes but also begins to characterize the potential evolution and rewiring of gene regulatory networks post genome duplication. Given the high economic, cultural, and environmental value of salmonids, additional information regarding their responses to selection, stress and pathogens is invaluable.

Charr (*Salvelinus* spp.) are an under-characterized and phenotypically diverse group within Family Salmonidae. Brook Charr *Salvelinus fontinalis* is a primarily freshwater species native to Eastern North America, whereas Arctic Charr *S. alpinus* has a circumpolar distribution mainly in the Arctic and has freshwater and anadromous life-history types with multiple ecotypes. These two lineages diverged approximately 10 million years ago (Horreo 2017) but are adapted to different environments. Brook Charr are sensitive to low oxygen and are limited to sub-Arctic regions whereas Arctic Charr are highly tolerant to cold and hypoxia but intolerant to high temperatures (Anttila *et al.* 2015) and stay in ice-covered lakes and ponds through long Arctic winters. Brook Charr are an important fish for sports fishing in Eastern North America, and Arctic Charr are an important food source for Northern communities. With a changing climate, both of these species may be at risk due to higher temperatures and changing ecosystem structures.

Here we profile liver transcriptomes of 100 Brook Charr offspring from a single family by RNA-sequencing to characterize co-expression patterns. Transcriptome profiling was conducted shortly (3 h) after the application of an acute handling stressor to all individuals during the reproductive season, increasing variance among individuals. The goals of this study are to i) characterize the underlying modular structure of gene co-expression in Brook Charr; ii) identify which modules are conserved or specific between sexes and within the congener Arctic Charr *S. alpinus*; and iii) link phenotypic correlation, functional gene enrichment and chromosomal information to the identified networks. This large characterization of a co-expression network provides valuable insight on transcription regulatory architecture in a non-model salmonid and provides network metrics for unannotated and annotated genes within the liver transcriptome of Brook Charr.

METHODS

Animals and sample collection

Juvenile Brook Charr used in this study were originally used by Sauvage *et al*. to construct a low-density genetic map for reproductive (2012a), growth and stress response QTL analyses (2012b), and to construct a high-density genetic map that was integrated with the other salmonids (Sutherland *et al.* 2016) then used to identify QTL, recombination rate differences between the sexes, and the Brook Charr sex chromosome (Sutherland *et al.* 2017). The 192 F_2 individuals were full sibs from a single family resulting from a cross

of an F_1 female and F_1 male that were from an F_0 female from a wild anadromous population (Laval River, near Forestville, Québec) and an F_0 male from a domestic population (Québec aquaculture over 100 years). F_2 offspring were raised until 65-80 g and then 21 phenotypes were collected along with several repeat measurements to determine growth rate. Full details on these phenotypes are previously described (Sauvage *et al.* 2012a, 2012b), including full details on sex-specific averages and standard deviations of phenotype values in all 192 offspring (see Table S1 from Sutherland et al. 2017). Note that the present study only includes 100 offspring, but these values provide information on the general trends in sex differences and phenotype value ranges. Correlations between phenotypes are also previously reported (see Figure S2 of Sutherland et al. 2017). The 15 phenotypes used in the present study were maturity, length, weight, growth rate, condition factor, liver weight, post-stress cortisol, osmolality and chloride, change in cortisol, osmolality and chloride between one week before and three hours after an acute handling stress, egg diameter, sperm concentration, and sperm diameter. Fish were anaesthetized with 3-aminobenzoic acid ethyl ester and killed by decapitation as per regulations of Canadian Council of Animal Protection recommendations approved by the University Animal Care Committee, as previously reported (Sauvage *et al.* 2012a). Most individuals were in a reproductive state at the time of the dissection. Phenotypic sex was determined by gonad inspection (Sauvage *et al.* 2012b). Immediately after decapitation, liver tissue was excised, flash frozen then kept at -80 ^{o}C until RNA extraction.

RNA extraction and library preparation

A total of 100 of the 192 individuals were used for liver transcriptome profiling. Prior to extraction, samples were assigned random order to reduce batch effects on any specific group of samples. Total RNA was extracted from equal sized pieces of liver tissue from approximately the same location on the liver for all samples (0.4 x 0.2 x 0.2 cm; ~1 mg). This piece was rapidly immersed in 1 ml TRIzol (Invitrogen), then placed on dry ice until all samples per batch were prepared (6-12 per extraction round). When all samples were ready, the samples immersed in frozen TRIzol were allowed to slightly thaw for approximately 1 min until beads within the vials were able to move, then the samples were homogenized for 3 min at 20 hz, rotated 180^{o}, and homogenized again for 3 min at 20 hz on a MixerMill (Retsch). The homogenate was centrifuged at 12,000 x g for 10 min at 4 ^{o}C. The supernatant was transferred to a new 2 ml tube and incubated for 5 min at room temperature. Chloroform (200 µl) was added to the tube, the tube was shaken vigorously for 15 s and incubated 3 min at room temperature, then centrifuged at 12,000 x g for 15 min at 4 ^{o}C. Finally the aqueous layer was carefully transferred to a new centrifuge tube, as per manufacturer's instructions. This aqueous layer was put onto an RNeasy spin column (QIAGEN) as per manufacturer's instructions with the optional on-column DNase treatment. All samples were quality

checked using a BioAnalyzer (Agilent), where all samples had RIN ≥ 8.3 (mean = 9.5), and were quantified using spectrophotometry on a Nanodrop-2000 (Thermo Scientific).

Libraries were prepared in the randomized order using TruSeq RNA Sample Prep Kit v2 (Illumina) to generate cDNA as per manufacturer's instructions using adapters from both Box A and Box B, AMPure XP beads (Agencourt) and a magnetic plate in batches of 8-16 samples per batch. A total of 1 µg total RNA was added to each reaction, and the manufacturer's instructions were followed to construct sequencing libraries. Fragmentation times of 2, 4 and 6 min were tested for optimal size fragmentation and consistency, and as a result of this test, all samples were processed using a 6 min fragmentation time to ensure inter-sample consistency. PCR amplification to enrich cDNA was performed using 15 cycles, as per manufacturer's instructions. All libraries were quantified using Quant-iT PicoGreen (ThermoFisher) and quality checked using the BioAnalyzer on High Sensitivity chips (Agilent). Once all samples were confirmed to be high quality and of approximately the same insert size, eight individually tagged samples were pooled in equimolar quantities (80 ng per sample) and sent to McGill Sequencing Center for 100 bp single-end sequencing on a HiSeq2000 (Illumina; total = 13 lanes). Parents (F_1 individuals) were sequenced in duplicate in two separate lanes.

RNA-seq mapping and normalization

Quality trimming and adapter removal was performed using Trimmomatic (Bolger et al. 2014), removing adapters with the *-illuminaclip* (2:30:10) option and removing low quality reads (< Q2) using *-slidingwindow* (20:2), *-leading* and *-trailing* options. Q2 was used for optimal quantification as previously demonstrated (MacManes 2014). A reference transcriptome for Brook Charr was obtained from the Phylofish database (Pasquier *et al.* 2016). Trimmed reads were mapped against the reference transcriptome with *bowtie2* (Langmead and Salzberg 2012) using *--end-to-end* mode reporting multiple alignments (*-k* 40) for optimal use with eXpress for read count generation (Roberts and Pachter 2013). The multiple alignment file was converted to bam format and sorted by read name using SAMtools (Li *et al.* 2009) and input to eXpress (see full bioinformatics pipeline in *Data Accessibility*).

Read counts were imported into edgeR (Robinson et al. 2010) for normalization and low expression filtering. The smallest sequence library (lib87, 8,373,387 alignments) was used to calculate a low-expression threshold, where transcripts with fewer than 10 reads mapping per transcript in this library was used as the count-per-million (cpm) threshold (1.19), as suggested in the edgeR documentation. Any transcript passing this threshold in at least five individuals was retained for downstream analysis. However, additional low expression filtering was conducted per sex separately (see below). The remaining reads were normalized using the weighted trimmed mean of M-values (TMM Robinson and Oshlack 2010). Finally, $\log_2$ cpm values were generated using normalized library sizes and exported as a

matrix file. Although transcripts were previously annotated in the Phylofish database (Pasquier *et al.* 2016), each transcript was re-annotated using *tblastx* against Swissprot (cutoff = $1e^{-5}$) to obtain as many identifiers as possible for Gene Ontology enrichment analysis. For individual gene descriptions, the re-annotated Swissprot identifier was used primarily, and the Phylofish annotation secondarily.

Weighted gene co-expression network analysis (WGCNA) in Brook Charr

Normalized $\log_2$ cpm expression levels for all individuals were used as an input for weighted gene correlation network analysis (WGCNA; Langfelder and Horvath 2008; 2012). To best associate modules with phenotypes of interest, sexes were analyzed separately and evaluated for conservation in the second sex (Langfelder *et al.* 2011). Due to the independent analysis of each sex, low expression filters (i.e. cpm > 0.5 in at least five individuals) were conducted separately in each sex. The average library size was 27,896,535 alignments, indicating cpm > 0.5 corresponds to at least 13.95 reads aligning to the transcript for this mean library size.

Within each sex, transcriptome outliers were detected and removed by clustering samples based on transcript expression by Euclidean distance and visually inspecting relationships with the *hclust* average agglomeration method of WGCNA (Langfelder and Horvath 2008). Removal of outliers prevents spurious correlations of modules due to outlier values and improves network generation (Langfelder *et al.* 2011). Remaining samples were then correlated with trait data using *plotDendroAndColors*. Network parameters for both female and male networks were defined as per tutorials using unsigned correlation networks (Langfelder and Horvath 2008). Unsigned networks consider the connectivity between similarly positive or negative correlations to be equal. Thus genes in the same module may have similar or inverse expression patterns. An optimal soft threshold power (6) was identified by evaluating effects on the scale free topology model fit and mean connectivity by increasing the threshold power by 1 between 1-10 and by 2 between 12-20 (Figure S1), as suggested by Langfelder and Horvath (2008). An unsigned adjacency matrix was generated in WGCNA to identify the 25,000 most connected transcripts to retain to reduce computational load. Finally, to further minimize noise and spurious associations, adjacency relationships were transformed to the Topological Overlap Matrix using the *TOMdist* function (Langfelder and Horvath 2008).

Similarity between modules was evaluated using module eigengenes (i.e. the first principal component of the module). Dissimilarity between eigengenes was calculated by signed Pearson correlation as suggested by Langfelder and Horvath (2008) and plotted using *hclust.* When modules were more than 0.75 correlated (dissimilarity 0.25), they were merged as suggested by Langfelder and Horvath (2008). Merged module eigengenes were then correlated against phenotypes by Pearson correlations. Notably the sign of the correlation does not necessarily indicate the sign of correlation between the

253 expression of specific genes in each module and the phenotype because the modules were built using
254 unsigned networks.
255 Network metrics for individual genes were calculated, including Gene Significance (i.e. the
256 absolute value of the trait-gene correlation) and Module Membership (i.e. the module eigengene-gene
257 correlation). Module membership was used to define the top central genes for modules of interest (i.e. hub
258 genes) and the module membership values of each gene within its cluster were determined. Gene
259 Significance characteristics were calculated for traits weight, specific growth rate, condition factor,
260 hepatosomatic index, change in cortisol, osmolality and chloride from the brief handling stressor, female
261 egg diameter, and male sperm concentration and diameter.
262 Enrichment analysis of clusters was conducted using the re-annotated Swissprot identifiers in
263 DAVID Bioinformatics (Huang *et al.* 2009). Heatmaps were generated by using the package *gplots*
264 (Warnes *et al.*) using the normalized $\log_2$ cpm data. Expression values were standardized across samples
265 for each gene and Pearson correlation was used to cluster genes and samples.
266 To determine sex-specific or sex-conserved modules, module conservation was evaluated by
267 comparing male gene expression to the generated female modules, and visa-versa, using the
268 *modulePreservation* function of WGCNA. A total of 200 permutations of randomly assigned module
269 labels were used to calculate module preservation rank and Zsummary (Langfelder *et al.* 2011). Low
270 Zsummary scores indicate no preservation (≤ 2), intermediate indicate moderate preservation (2-10) and
271 high scores (≥ 10) indicate strong module preservation (Langfelder *et al.* 2011). Module quality was also
272 determined for each module as a measure of module robustness that is characterized by conducting the
273 analysis on multiple random subsets of the original data (Langfelder *et al.* 2011). In addition, cross-
274 tabulation of the proportions of female modules in male modules and visa-versa were performed in R (R
275 Core Team 2018) to determine when a majority of one module was found in a module of the opposite sex.
276 Cross-tabulation requires similar modular structures of the compared networks, whereas adjacency
277 comparisons directly compare co-expression independent of network topology. All pipelines to analyze
278 the current results are documented and available on GitHub (see *Data Accessibility*).
279
280 *Module conservation in Arctic Charr*
281 To compare module conservation between Brook Charr and Arctic Charr *S. alpinus* we used RNA-seq
282 data from 1+ year-old Arctic Charr. The broodstock of this population was reared in hatchery conditions
283 for three generations after being collected from a subarctic, land-locked population in Finland (Lake
284 Kuolimo, 61°16′ N; 27°32′ E). The data were collected from nine male liver samples from each of 8 ^{o}C
285 and 15 ^{o}C (total = 18 samples), but due to a large effect of temperature on the Arctic Charr transcriptome,
286 only the nine samples from the lower temperature were used here (8 ^{o}C, normal rearing temperature

during summer at the fish hatchery; Figure S2) (Prokkola *et al.* 2018). Fish body mass at 8°C was on average 24.2 g ± standard deviation (S.D.) 10.4 g. Sample processing was explained fully by Prokkola et al (2018) and briefly described here. In August 2013, fish were euthanized using 200 ppm sodium bicarbonate-buffered tricaine methanesulfonate (MS-222), after which liver samples were collected and flash frozen in liquid nitrogen (Prokkola *et al*. 2018). RNA was extracted from approximately 10 mg of liver tissue using Tri-reagent (Molecular Research Center), and quality checked using a BioAnalyzer (Agilent), with an average identified RNA integrity number of 9.95. Strand-specific cDNA library preparation and sequencing were conducted at Beijing Genomics Institute (BGI Hong Kong) using TruSeq RNA Sample Prep Kit v2 (illumina) and sequenced on an Illumina HiSeq2000 instrument to generate paired-end 100 bp reads. All samples were pooled with unique barcodes across four sequencing lanes. Adapters were removed at BGI, and reads trimmed with Trimmomatic (Bolger et al. 2014) using options *leading* and *trailing* (5) *slidingwindow* (4:15) and *minlen* (36). From samples included in this study, on average 41.7 ± S.D. 7.4 million reads were obtained. Transcript expression was calculated as above, including using the Brook Charr reference transcriptome for ease of cross-species comparisons. Data filtering, normalization and WGCNA input was conducted as above. However, modules were not generated from these samples due to the smaller sample size within only the 8°C temperature relative to the Brook Charr samples. Using samples from both temperatures, modules were previously identified (Prokkola *et al.* 2018). Once normalized and input to WGCNA, read counts in Arctic Charr 8°C samples were used to build a gene adjacency matrix, which was then compared against modules generated for female and male Brook Charr samples using the *modulePreservation* function as described above. Caveats regarding this data should be noted, including the smaller sample size, minor differences in rearing environments (albeit both were reared in hatchery conditions), and the maturity stage of the fish.

Identifying one transcript per gene and assigning chromosome positions

The reference transcriptome (see *Data Accessibility*) possibly contains multiple isoforms for a single gene, and therefore to investigate chromosome enrichment of identified modules, an approach to limit the transcriptome to a single transcript per gene was applied here. First, the Brook Charr reference transcriptome was aligned to the Atlantic salmon *Salmo salar* chromosome-level genome assembly RefSeq GCF_000233375.1 (Lien *et al.* 2016) using GMAP (Wu and Watanabe 2005). Output alignments were converted to an indexed bam file with only high quality (-q 30) alignments retained using samtools (Li *et al.* 2009). The indexed bam was converted to a bed file using bedtools *bamtobed* (Quinlan and Hall 2010). Total lengths of each Brook Charr transcript were determined using python scripts (see *Data Accessibility*). Using alignments against the Atlantic Salmon genome, a single Brook Charr transcript was retained for every group of transcripts that aligned in a continuous overlapping block on the Atlantic

Salmon genome using a custom R script (see *Data Accessibility*). This script preferentially retained the longest expressed transcript of the contiguous block, when possible, and discarded all other redundant transcripts as well as those that did not align. In some cases, a single transcript can align to multiple locations with high mapping quality (MAPQ $\geq$ 30). Since there was no reason to retain one alignment over another, both alignments were retained in the baseline set for these cases. In the case that both alignments were to the same chromosome, which could occur due to tandem duplication of genes, this was noted during chromosome enrichment analysis (see below).

The alignment information per retained Brook Charr transcript was used to assign an Atlantic Salmon chromosome identifier to each retained unique transcript, and this was combined with the co-expression network information of the transcript, specifically its assigned module. This analysis was conducted separately for females and males, as expressed genes are in some cases different between the two sexes and therefore so would the selection of which transcript to retain. Finally, for each co-expression module, the proportions of Brook Charr genes belonging to each Atlantic Salmon chromosome were characterized and compared to the total list of all non-redundant Brook Charr transcripts identified. Fisher exact tests were then used to determine significance for each module-chromosome combination ($p \leq 0.01$). The correspondence between the Atlantic Salmon genome assembly accession identifier and the Atlantic Salmon chromosome identifier were obtained from the NCBI genome assembly website (see *Data Accessibility*).

The approach taken here and described below is conservative and has some drawbacks that would reduce power, but is superior to including the same gene multiple times for multiple isoforms. The main drawback is that this method of redundancy identification depends on the alignment against the reference genome and therefore any genes not aligning were removed. Further, the non-target reference genome will have differences from our focal species. However, even though large scale fusions may differ among the salmonids, much of the within chromosome arm synteny is retained, as previously demonstrated (Sutherland *et al.* 2016). Using the chromosome correspondence between Atlantic Salmon and Brook Charr within previous work (Sutherland *et al.* 2016), the Brook Charr chromosomes enriched for specific modules can be identified. It is probable that resulting enrichments would be stronger if all genes were successfully aligned to the reference genome, and if the reference genome was within the focal species. Nonetheless, this approach reduced many isoforms down to a fewer number of unique genes for the purpose of characterizing chromosome enrichment from each module in this non-model species without an intraspecific reference genome. Other approaches have been used to reduce redundancy in *de novo* reference transcriptomes of salmonids, although these are often conducted during the process of assembling the transcripts. For example, four reference transcriptomes were reduced by approximately 4-

fold in transcript number (Carruthers *et al.* 2018) using the transdecoder pipeline (Haas *et al.* 2013) then reducing marginally further using CD-hit-EST (Li and Godzik 2006).

RESULTS

Transcriptome overview

Of the total 69,440 transcripts in the Brook Charr reference transcriptome, 51,911 passed initial low expression filters (counts-per-million mapped reads (cpm) > 1.19 in at least 5 individuals when using all samples together including females, males, and male Arctic Charr). This corresponded to at least 10 reads mapping per transcript in the smallest library size. Subsequent filtering within each sex individually (cpm > 0.5 in at least 5 individuals; corresponding to > 13 reads mapping based on average library size) resulted in 50,748 and 50,530 transcripts passing filters in females and males, respectively. When considering each sex individually, most of the expressed genes were expressed in a majority of the samples, with females expressing 35,461 transcripts in > 90% of the samples, and males expressing 35,714 transcripts in > 90% of the samples (see Figure S3).

Hierarchical clustering of samples by gene expression indicated a large effect of sex, where 35 of 47 F_2 females were all clustered together with the F_1 female, and 52 of 53 F_2 males were clustered together with the F_1 male (Figure 1). As described in the Methods, outliers were removed to avoid spurious network correlations (Langfelder *et al.* 2011), and this included the removal of one male leaving 52 males remaining, and the removal of one group of females that had large liver weight, leaving 35 females remaining (see phenotype liver weight in Figure 1; Figure S4). When these samples were included while constructing the female network, many modules correlated with the liver weight phenotype, suggesting that these samples were having a large impact on the network. Interestingly, females displayed higher inter-individual variance than males, as indicated by the multiple smaller clusters of females in the hierarchical clustering relative to the fewer and larger clusters of the males (Figure 1).

Network construction and phenotype correlations: female Brook Charr

Highly correlated module eigengenes (r > 0.75) were merged, combining 81 assigned modules into 14 (Figure 2A; Figure S5A). Assigned female modules each contained a range of 77-10,533 transcripts (Table 1). The largest module was *darkred*, with 10,533 transcripts (see Table 1), which included more transcripts than even the unassigned *grey* module (second largest; 5,892 transcripts).

Phenotype correlations with module eigengenes indicate potential functional associations of the modules (Table 1; Figure 3). Only non-redundant phenotypes were used (n = 15). The strongest associations of phenotypes to modules were with maturity index, for example with *thistle2* (r > 0.81), and *coral1* (r = -0.83). Although the large liver weight outlier samples were removed prior to network

generation, liver weight remained highly correlated with *indianred4* (r = 0.73), *salmon*, *coral* (r = 0.52), and *blue2* (r = -0.64). Growth rate showed similar module correlations to liver weight (Figure 3). Osmolality change was also correlated with *indianred4* (r = 0.56) and *blue2* (r = -0.58), as well as with *darkred*, *thistle3* (r ≥ 0.51), *darkorange*, *green* and *darkmagenta* (r ≥ -0.49). Chloride change had no significant associations, but post-stress chloride was correlated with *thistle3* (r = 0.56) and *lightsteelblue* (r = -0.53). Although no modules were significantly associated with cortisol (change or post-stress; p ≥ 0.01), *ivory* was close (r = -0.38; p = 0.03).

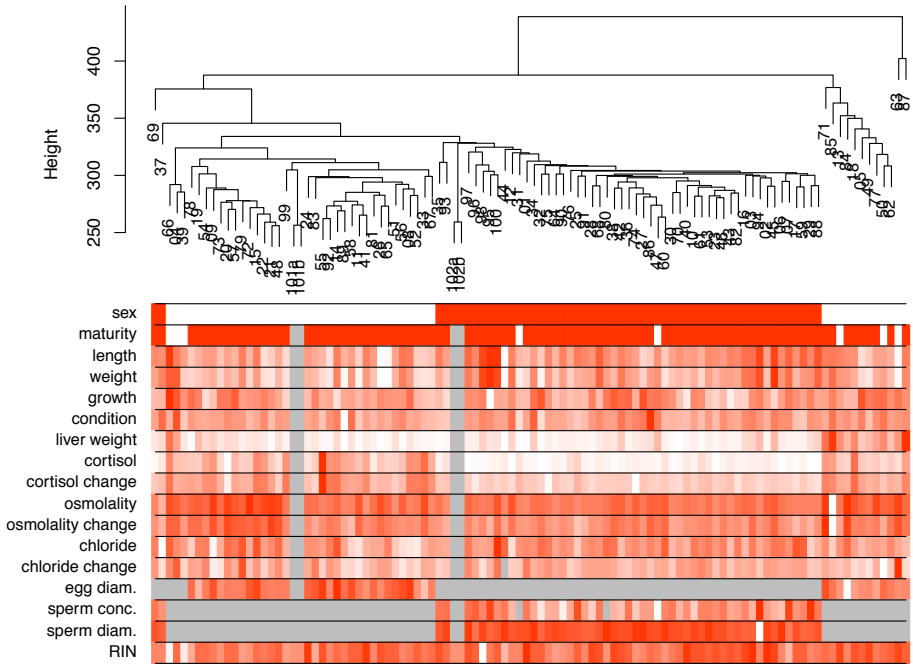
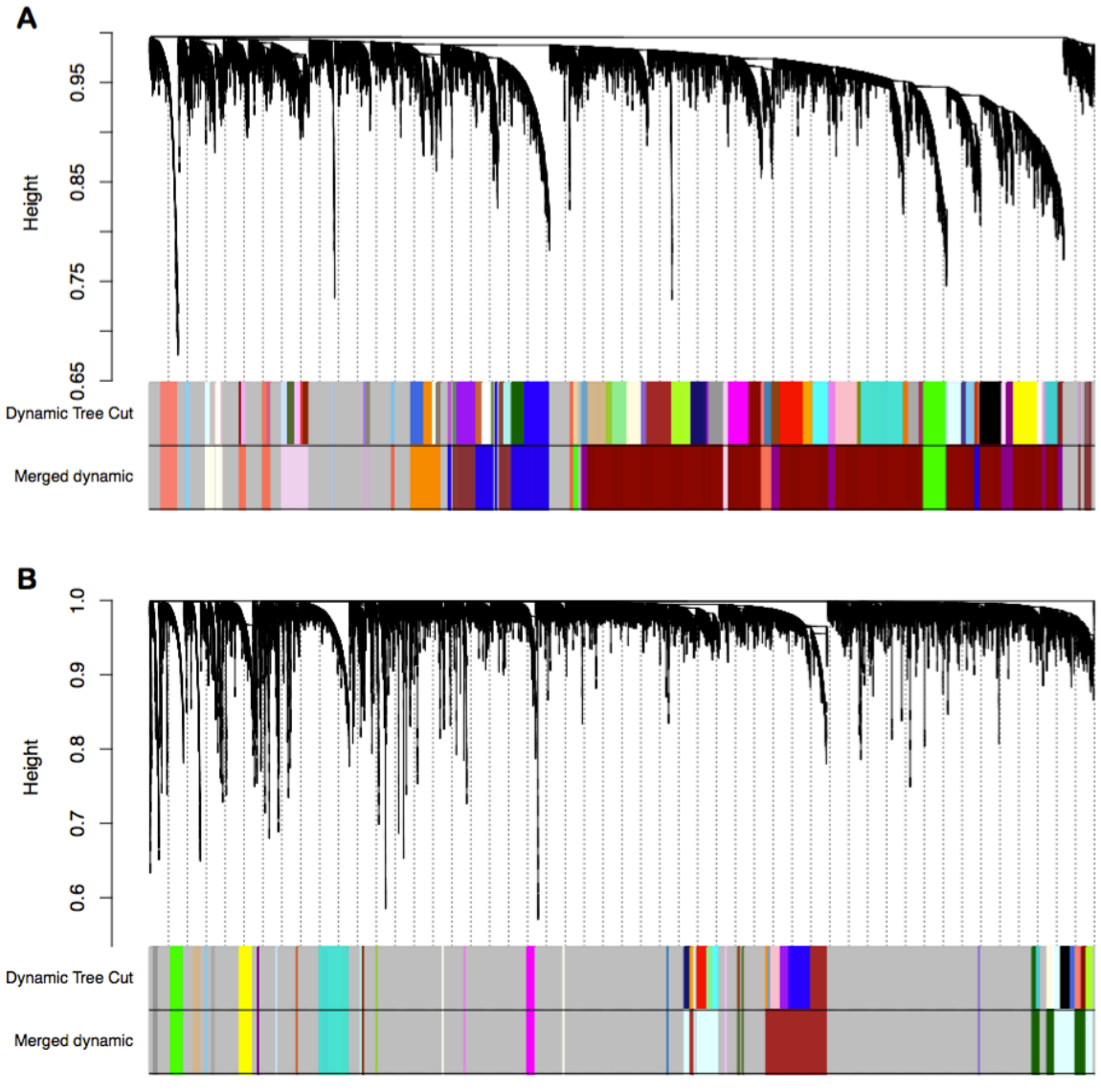


Figure 1. Brook Charr individual samples clustered by gene expression similarity using all genes (top) with corresponding trait values shown in the heatmap (bottom). Sex was the largest factor affecting the data (see heatmap sex row; white = females; red = males). Parents were sequenced in duplicate, and clustered with the offspring of their respective sex (see 101ab for mother and 102ab for father). Females with large liver weight clustered outside the other female samples (see on the right hand side of the heatmap liver weight row).

In addition to the correlations with phenotypes, functional enrichment analysis of the genes within modules was conducted (Table 1; Additional File S1). The *salmon* module (correlated with liver weight) was enriched for erythrocyte development. The *blue2* module (liver weight) was associated with ribonucleoprotein complex. *Darkorange*, *green*, and *darkmagenta* modules (all correlated with osmolality change) were enriched for small molecular metabolic process, translation and metabolism functions,

409 respectively. Some modules did not have significant enrichment of biological processes, such as
410 *lightsteelblue* (correlated with chloride change).
411



412

413 **Figure 2**. Transcripts clustered by co-expression shown as similarity to neighbouring gene shown as
414 vertical lines (top) and shown with assigned module (bottom) in (A) females and (B) males. Modules are
415 shown prior to merging (Dynamic Tree Cut) and after merging similar clusters (Merged dynamic). More
416 transcripts were assigned to modules in females (76%) than in males (28%). Note: transcript order is not
417 retained and module colors are not related in (A) and (B).
418

419 *Conservation of co-expression: female Brook Charr network*

420 The conservation of co-expression of female Brook Charr modules in Brook Charr males was primarily
421 evaluated using network adjacency comparisons, which are more sensitive and robust than cross-
422 tabulation methods (Langfelder *et al.* 2011). Most female modules were conserved in males and only

darkred had weak evidence for conservation (Table 1; Figure 4A). *Green* was the most conserved (Zsummary = 100), followed by *blue2*, *salmon*, *lightsteelblue* and *darkorange* (Zsummary ≥ 48; Figure 4A; Table 1). Modules associated with translation activities were among the highest conserved modules (e.g. *green* and *blue2*).

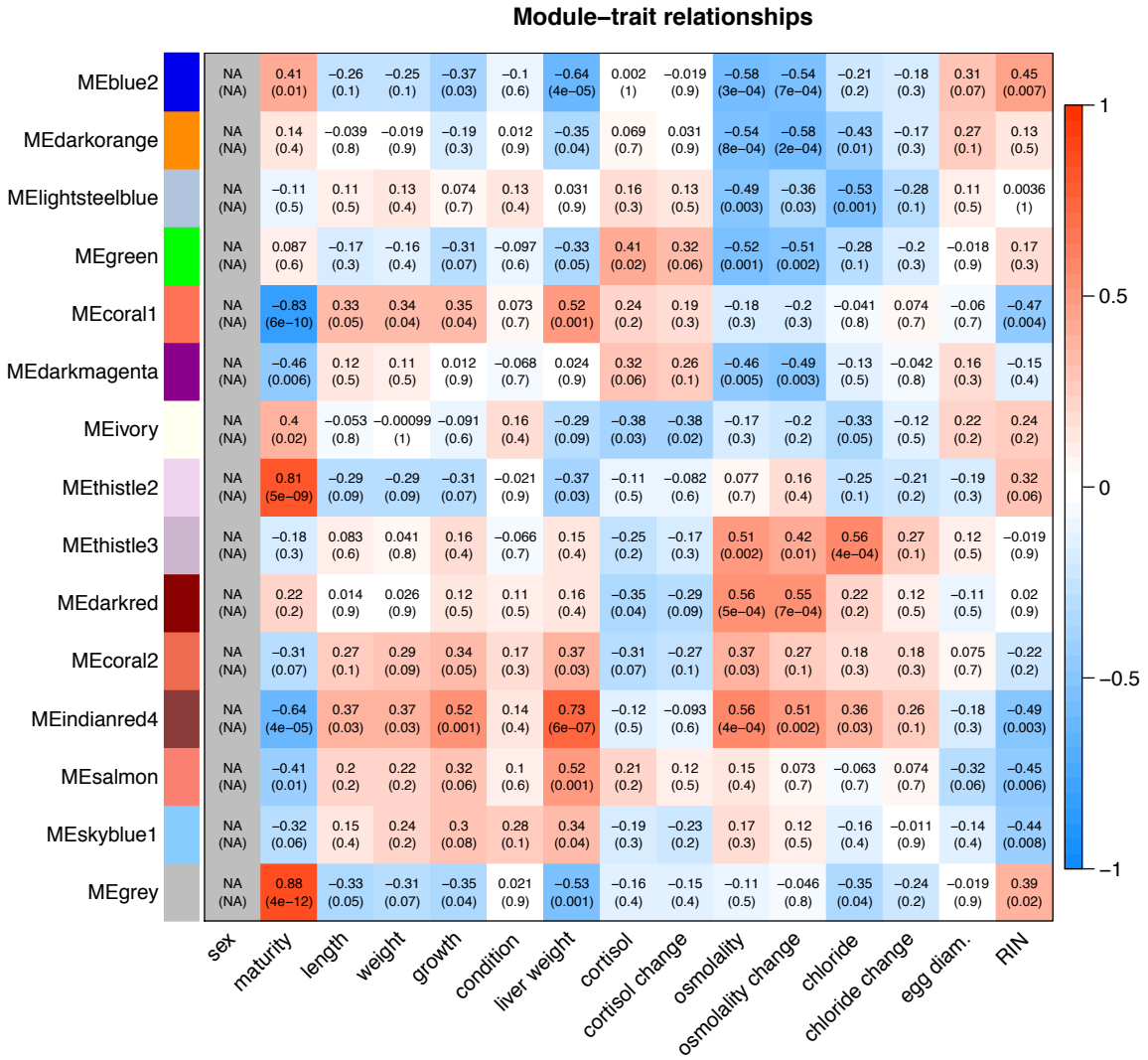


Figure 3. Module-trait relationships for Brook Charr females, including correlation r-value and p-values. The boldness of color indicates the strength of the relationship. Module-trait correlations are also shown in Table 1 with more general grouping of traits alongside other metrics such as module size and enriched Gene Ontology categories. Male module-trait relationships are shown in Figure S6.

To further inspect cross-species conservation of co-expression, samples were obtained from a recent transcriptomic analysis of male Arctic Charr liver (Prokkola *et al.* 2018). Even with caveats regarding sample size (see Methods), several female Brook Charr modules were highly preserved in

Arctic Charr males, including *blue2* (Zsummary = 34), *green* (Zsummary = 24), and *salmon* (Zsummary = 17; Figure 4B), also the most conserved in male Brook Charr (Table 1; Figure 4A). Other female Brook Charr modules with moderate evidence for conservation in male Arctic Charr included *darkorange* and *lightsteelblue* (Zsummary > 8), which were also highly conserved in male Brook Charr. It is noteworthy that the ranking of preservation of female modules in the Arctic Charr and Brook Charr males is highly similar (Table 1).

Table 1. Female modules shown with the number of transcripts within the module (n), the general category of traits correlated with the module ($p \leq 0.01$), the most significantly enriched Gene Ontology category (Biological Process), the Zsummary for preservation of the module in Brook Charr males (BC m) and Arctic Charr males (AC m), as well as the module quality (robustness). Zsummary < 2 is not conserved, 2 < Zsummary < 10 is moderately conserved, and > 10 is conserved. The *grey* module includes unassigned genes and the *gold* module is a random selection of 1000 genes from the assigned modules for testing preservation metrics. Full module-trait correlations are shown in Figure 3A, full GO enrichment in Additional File 1, and expanded summaries of this table in Additional File S2.

Module	n	Traits	GO Enrichment (BP)	Preservation BC m | AC m		Quality
green	725	blood	translation	100	24	75
blue2	1762	blood; liver; maturity; RIN	ribonucleoprotein complex biogenesis	71	32	58.5
salmon	449	liver; maturity; RIN	erythrocyte development	56	17	42
darkorange	805	blood	small molecule metabolic process	48	8.8	36
lightsteelblue	77	blood	*none*	48	8.2	25.5
ivory	451	*none*	ER-associated ubiquitin-dependent protein catabolic process	37	5.4	25.5
thistle2	834	maturity	tissue development	29	3.6	36.5
indianred4	1180	blood; growth; liver; maturity; RIN	membrane assembly	28	2.8	22
coral1	689	liver; maturity; RIN	regulation of cell growth	25	5.2	34
thistle3	235	blood; liver; size	inorganic anion transmembrane transport	24	1.2	21
skyblue1	96	RIN	intracellular signal transduction	17	1.2	21
darkmagenta	1072	blood; maturity	single-organism metabolic process	13	5	58
coral2	200	*none*	regulation of blood circulation	11	2.8	14
grey	5892	*Not a module*	*Not a module*	9	2.4	-18
gold	1k*	*blood; size*	*Not a module*	4.8	0.38	-0.86
darkred	10533	blood	protein ubiquitination	4.5	-0.55	23.5

Network construction and phenotype correlations: male Brook Charr

Highly correlated male modules (eigengene correlation r > 0.75) were merged, reducing 44 assigned male modules to 25 (Figure 2B; Figure S5B). Unlike the female network, a large proportion of the male data was unassigned. The unassigned *grey* module contained 72% of the analyzed transcripts (17,992 transcripts). Assigned modules each contained between 54-1,732 transcripts (Table 2).

Phenotypic correlations with male module eigengenes were characterized (Figure S6). Similar to females, liver weight was highly correlated with modules including *darkgreen* (r = 0.53) and *yellow* (r = -0.52). Other highly correlated module-phenotype comparisons included osmolality change with *lightcyan1* (r = 0.52), *yellow* and *steelblue* (r = -0.4). Relative to the female comparisons, there were more modules strongly correlated with growth rate including *yellow* and *steelblue* (r < -0.39), *darkgreen* and *turquoise* (r > 0.37). Post-stress chloride was correlated with *ivory* and *lightcyan1* (r > 0.4), *steelblue* and *tan* (r < -0.42). Although modules from the female network did not show significant correlations with length and weight, male modules *darkgrey* and *green* were correlated with length (r > 0.42). The male-specific phenotype sperm concentration was associated with *steelblue*, *brown* (r > 0.38), and *lightcyan1* (r = -0.38).

Several modules were enriched for immunity-related functions, including *darkmagenta* (defense response to virus) and *steelblue* (positive regulation of innate immune response) (Table 2 and Additional File S1). These two immune processes were found to belong to different modules that were not just inversely regulated, as this would put them in the same module given the unsigned network, but likely having somewhat decoupled regulation (Figure 5A), although the modules were still fairly correlated (Figure S5B). *Yellow*, *brown* and *green* modules were enriched for ribosomal or translation-related functions (Table 2). *Ivory*, the most sex- and species-specific male module (see below; Table 2; Figure 5C) was enriched for neurogenesis in GO biological process, but also transcription factor activity in GO molecular function (Additional File S1). Other non-conserved, or lowly conserved modules (see below) were enriched for membrane activity including *darkgrey* (mitochondrial inner membrane) and *lightcyan1* (membrane organization; Additional File S1).

Conservation of co-expression: male Brook Charr network

Conservation of male modules in females was also evaluated. In contrast to what was observed in the conservation of female modules in males (see above), many of the male modules were weakly to moderately conserved in the females (Figure 4C). Highly conserved modules included *yellow* and *brown* (translation-related; preservation Zsummary ≥ 34; Figure 5B), *tan* (Zsummary = 43), and *lightcyan* (Zsummary ≥ 34). Some modules were less conserved than even the randomly generated *gold* module (Table 2) and the unassigned *grey* module, including *green* (translation and size; Zsummary = 4.5),

darkgrey (mitochondrial membrane; Zsummary = 1.8) and *ivory* (transcription factor activity; Zsummary = 0.8; Figure 5C), suggesting the co-expression trends in these modules are specific to the males.

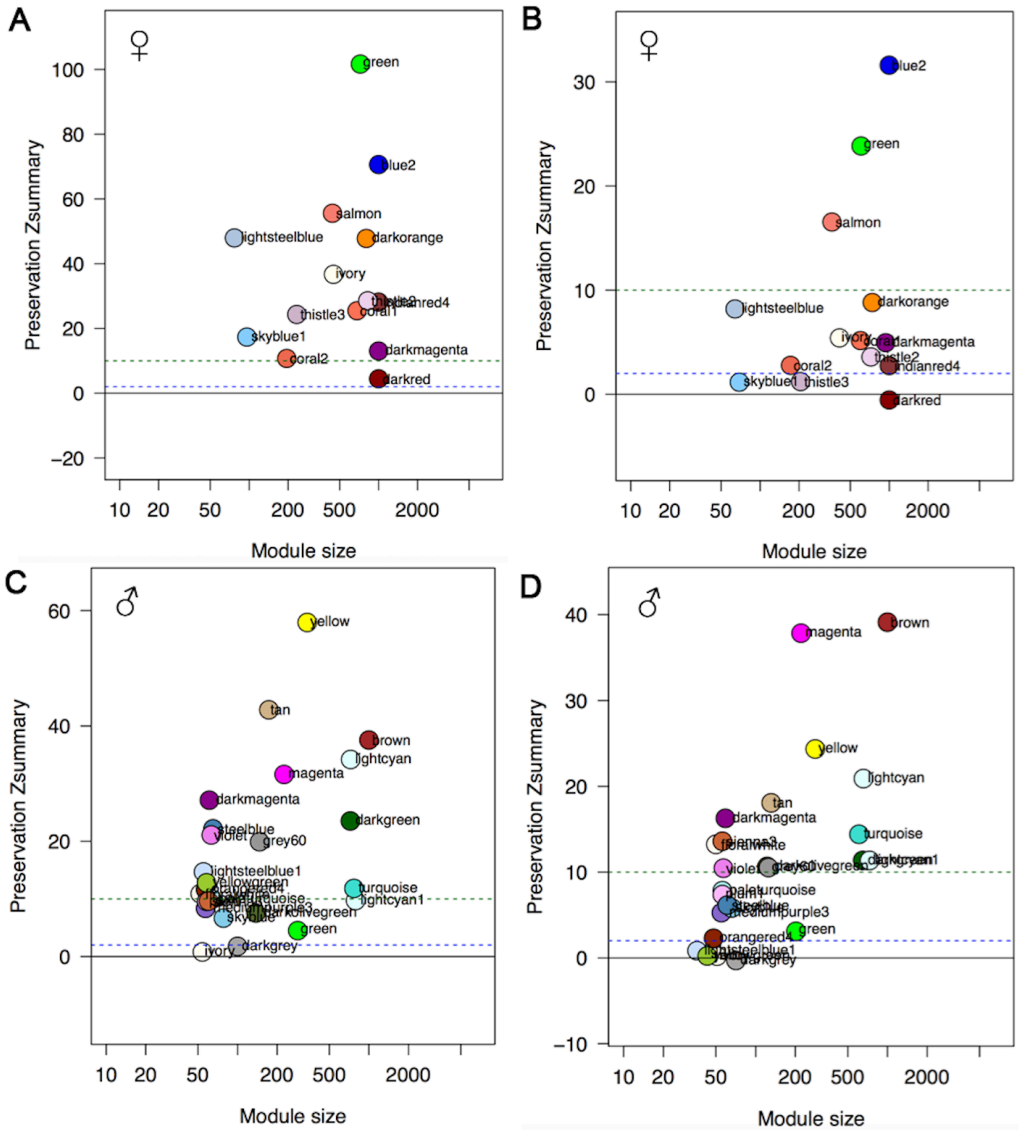


Figure 4. Module conservation Zsummary scores of Brook Charr female modules in (A) Brook Charr males and (B) Arctic Charr males, and of Brook Charr male modules in (C) Brook Charr females and (D) Arctic Charr males. Under the hatched blue line are non-conserved modules, between the hatched blue and green lines are low to moderate conservation (Zsummary 2-10), and above the hatched green line are strong evidence of conservation (Zsummary > 10). Most of the female modules had high conservation in Brook Charr males except the darkred module. Many female modules were also moderately conserved in male Arctic Charr, with some showing high conservation such as salmon, green and blue2. In the male Brook Charr modules, many modules had moderate levels of conservation with several having low conservation including ivory and darkgrey. The Arctic Charr data always showed lower conservation due to a combination of factors including lower sample sizes, the use of a different species reference transcriptome, species differences, and possibly slight environment or development differences.

Conservation of male Brook Charr modules was also explored in Arctic Charr males. Similar to that observed in the female modules, when a male Brook Charr module was conserved in female Brook Charr, it was also often conserved in male Arctic Charr (see similar ranking of conservation; Table 2; Figure 4D). Some modules, including *magenta* (response to endoplasmic reticulum stress) and *brown* (ribonucleoprotein complex biogenesis) were either equally or more conserved in the Arctic Charr males as they were in the Brook Charr females. This indicates that sufficient power is present in the Arctic Charr samples to identify conservation in modules from a separately generated network. Male Brook Charr modules that are highly or moderately conserved in female Brook Charr but not conserved in male Arctic Charr (e.g. *lightsteelblue1* and *yellowgreen*) may be conserved across sexes, but species-specific. However, this will need further data to characterize due to the caveats of the Arctic Charr data (see above).

Cross-tabulation comparisons of female and male modules

Cross-tabulation was used to gain insight into the correspondence of modules and comparison of network topology between the sexes. Most importantly, 12 of the 25 male modules had more than 25% of their genes contained within the largest female module *darkred* that contained 10,533 transcripts (Additional File S2). This could be expected given the size of the *darkred* module, but this means that the genes from these male modules were often grouped together into one very large module in females. For instance, the female *darkred* module included 87% of the genes in male *darkmagenta* (virus defense response). Even though the adjacencies of the genes within male *darkmagenta* were conserved in females (preservation Zsummary = 27; Table 2), the grouping of these genes was specific to the topology of the male network. In other cases, female modules were made up of a large number of genes from a single male module. For example, the female module *lightsteelblue* (intracellular signal transduction) was comprised of a majority of genes (82%) from male *grey60*.

By contrast, 20 of the 25 assigned male modules each did not contain more than 10% of the genes from any one female module (Additional File S2). Therefore, the grouping of many male modules into the single module female *darkred* did not occur for the female modules. Instead, the genes within the female modules often were largely present in the unassigned *grey* module of the males (Additional File S2). This further indicates that the male network was less modular than the female network, as these specific genes from female modules were unassigned in the male network.

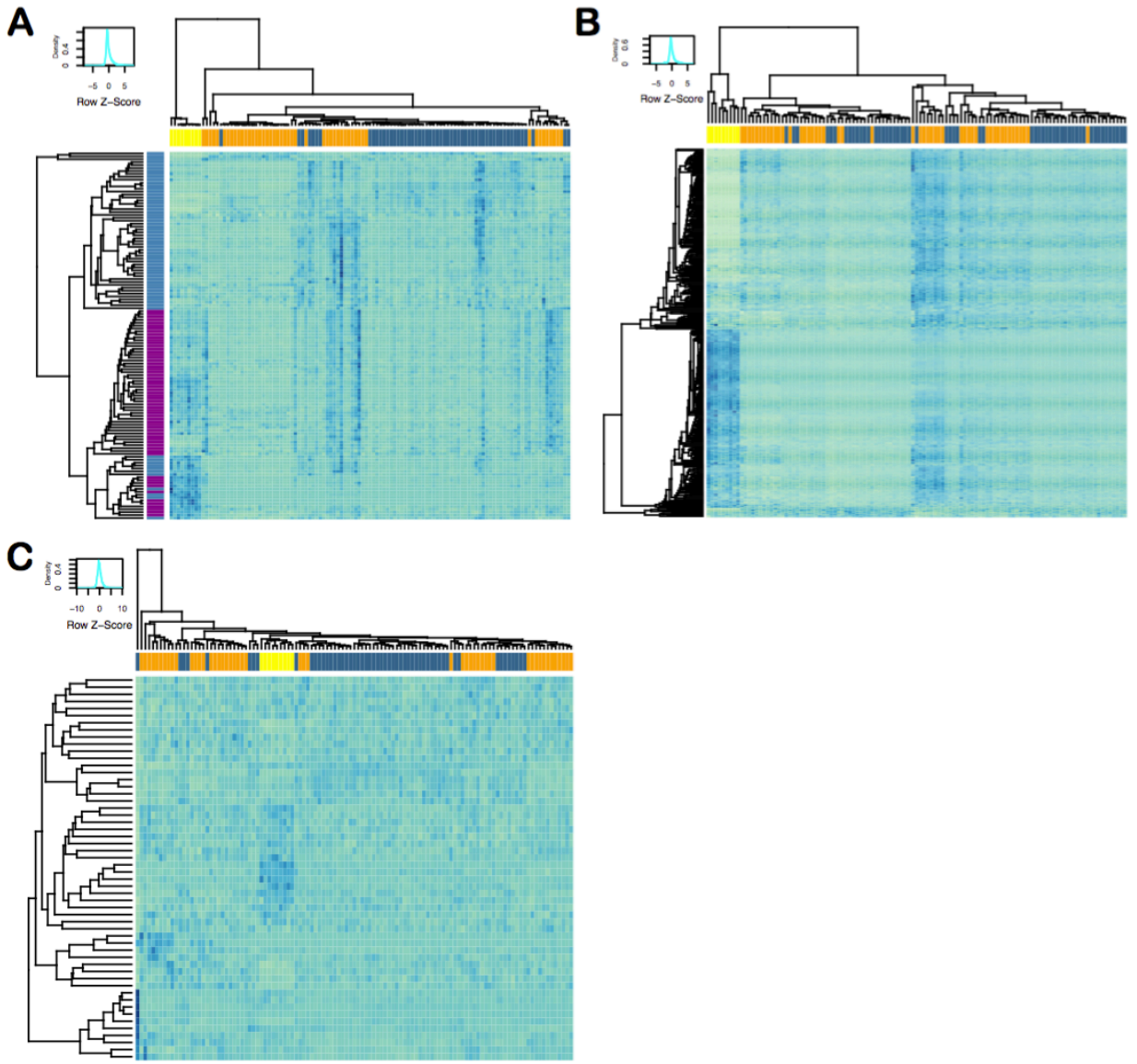


Figure 5. Heatmaps of gene expression of transcripts within (A) two immunity-related male clusters, *steelblue* and *darkmagenta*, which are related to innate immunity and innate antiviral immunity, respectively, (B) the conserved male module *yellow* (translation), and (C) the male-specific *ivory* (transcription factor activity). Samples are shown on the horizontal with Brook Charr females (orange), Brook Charr males (blue), and Arctic Charr males (in yellow). Genes from multiple modules in (A) are colored according to the module (steelblue, darkmagenta). The normalized expression values are shown within the heatmap. Genes within the modules are listed in Additional File S5.

Table 2. Male modules with the number of transcripts within the module (n), the general category of traits correlated with the module ($p \leq 0.01$), the most significantly enriched Gene Ontology category (Biol. Proc.), the Zsummary for preservation of the module in Brook Charr females (BC f) and Arctic Charr males (AC m), as well as the identified module quality (robustness). Zsummary < 2 is not conserved, 2 < Zsummary < 10 is moderately conserved, and > 10 is conserved. The *grey* module includes unassigned genes and the *gold* module is a random selection of 1000 genes from the assigned modules for testing preservation metrics. Full module-trait correlations are shown in Figure S6, full GO enrichment in Additional File 1, and expanded summaries of this table in Additional File S2.

Module	n	Traits	GO Enrichment (BP)	Preservation BC f | AC m		Quality
yellow	339	blood; liver; size	translation	58	24	51
tan	173	blood; size	*none*	43	18	41.5
brown	1732	blood; sperm	ribonucleoprotein complex biogenesis	38	39	72.5
lightcyan	734	blood; size	organic acid metabolic process	34	21	69
magenta	226	*none*	response to endoplasmic reticulum stress	32	38	45
darkmagenta	61	*none*	defense response to virus	27	16	25.5
darkgreen	724	blood; growth; liver; size	*none*	24	11	58
steelblue	65	blood; growth; sperm	positive regulation of innate immune response	22	6.2	26.5
violet	64	liver	sterol biosynthetic process	21	10	29.5
grey60	147	*none*	secretion by cell	20	11	39
lightsteelblue1	55	*none*	*none*	15	0.85	23
yellowgreen	61	blood	*none*	13	0.24	24.5
turquoise	783	*none*	immune system process	12	14	68.5
orangered4	57	sperm	regulation of apoptotic process	12	2.3	26.5
floralwhite	52	*none*	glutathione metabolic process	11	13	25
paleturquoise	64	blood	response to organonitrogen compound	10	7.8	25.5
lightcyan1	791	blood; sperm	cytokinesis	9.7	11	67.5
sienna3	61	liver; size	extracellular structure organization	9.6	14	26.5
plum1	58	*none*	*none*	9.6	7.4	23.5
mediumpurple3	57	size	ribonucleoprotein complex assembly	8.4	5.3	20.5
darkolivegreen	138	blood	*none*	7.6	11	28.5
skyblue	78	RIN	regulation of transcription, DNA-templated	6.7	5.8	28
grey	17992	*Not a module*	*Not a module*	5.5	2.3	-14
gold	1k*	*Not a module*	*Not a module*	4.9	3.2	-0.035
green	334	size	translation	4.5	3.1	39
darkgrey	100	size	small molecule metabolic process	1.8	-0.28	31.5
ivory	54	blood	neurogenesis	0.81	0.22	24

Hub genes in modules of interest

Hub genes are identified using module membership (MM) scores that estimate gene connectivity to other genes within the module (Additional Files S3 and S4). Importantly, the distribution of MM is continuous and there is no discrete change between a hub gene and a non-hub gene. Here we present the highest ranked hub genes for several modules from both the female and male networks that are of interest due to GO enrichment, module conservation or phenotypic correlation.

Female *lightsteelblue* was of interest due to conservation in the male network (77 transcripts; intracellular signal transduction), but hub genes (MM > 0.9) were often unannotated. The six of 15 that were annotated, included *selenium-binding protein,* several stonustoxin subunits (proteins with hemolytic activity), *estrogen-related receptor gamma*, and *NACHT, LRR and PYD domains-containing protein 12-like.* Several other transcripts, still with high MM (> 0.80), were annotated as *neoverrucotoxin subunit beta-like*, another toxin-related protein with hemolytic activity. Female *salmon* (449 transcripts; hemopoiesis) was also of interest as it was conserved in males. Hub genes from *salmon* had 53 transcripts with MM > 0.9, most of which were annotated, often with hemopoietic functions including *band 3* (anion transport protein in erythrocytes), *erythropoietin receptor* (promotes blood cell proliferation and prevents apoptosis), various hemoglobin subunits, and *5-aminolevulinate synthase* (involved in heme synthesis). A transcript annotated as the transcription regulator *Kruppel-like factor 4* also was in the top MM genes. The highest MM is *tubulin beta-6 chain-like*, which is a major constituent of microtubules.

For male modules, *ivory* was of interest due to enrichment for transcription factor activity, male-specificity of the module, and correlation with chloride levels. For this module, there were no genes with MM > 0.9. The highest MM transcripts included transcription regulators such as *frizzled-9* (FZD9) and *protein wnt* (WNT9) the ligand for the frizzled family of transmembrane receptors (MM = 0.79), the transcription factor *RAR-related orphan receptor gamma 2 protein* and *lysine-specific demethylase 4B* (*KDM4B*), which plays a role in the histone code, and transcription activator *nuclear factor 1 A-type*. Although these genes related to transcription factor activity were all ranked high in MM, there were also other transcripts lower in MM rank that are also putatively involved in transcription regulation, including those annotated as putative zinc finger proteins, or related to the SOX family of transcription factors. Male *darkmagenta* was also of interest as it was highly conserved and enriched for innate viral immunity-related functions. Only four transcripts were MM > 0.9, but some of the top ranking MM transcripts were associated to the module function, including the most connected *probable ATP-dependent RNA helicase* (DDX58/RIG-1; MM = 0.93), important for sensing viral infection and inducing type I interferons and pro-inflammatory cytokines, *interferon-induced protein with tetratricopeptide repeats 5* (IFIT5; MM = 0.89), an interferon induced RNA-binding immunity protein, *galectin-3-binding protein A* (L3BPA; MM = 0.89), *sacsin* (SACS; MM = 0.88) and *probable E3 ubiquitin-protein ligase* (HERC6; MM = 0.84) both

588 involved in immunity in fish, and *signal transducer and activator of transcription 1-alpha/beta* (STAT1;
589 MM = 0.83), a transcription activator that mediates responses to interferons. Male *steelblue* is of interest
590 due to conservation and enrichment of immunity-related functions. Only one transcript had MM ≥ 0.9,
591 *40S ribosomal protein S11-like*. However, other high-ranking hub genes included innate immunity-related
592 genes, namely *C-type lectin domain family 4 member E* (CLC4E; MM = 0.86), *metalloreductase* (STEA4;
593 MM = 0.86) involved in integrating inflammatory and metabolic responses. Similar to above, other lectin-
594 related transcripts were also found lower in MM, including others annotated as *CLC4E* and as *leukocyte*
595 *cell-derived chemotaxin-2* (LECT2), which has neutrophil chemotactic activity. Male *turquoise* was also
596 of interest as it was enriched for immune system processes and had 17 hub genes, 11 of which were
597 related to immune system processes. There were no correlated traits for this module. The interplay of
598 these three different immune modules can be investigated by looking at module eigengene clustering
599 (Figure S5B). The immune-related *steelblue* and *darkmagenta* modules had similarly clustering module
600 eigengenes, and these two clustered separately from the *turquoise* module (see Figure S5B), suggesting a
601 more distinct expression profile for *turquoise*. Male modules *yellow* (translation) and *brown* (translation)
602 were among the most conserved across species and had correlations to phenotypes. The four hub genes of
603 *yellow* all encode ribosomal proteins. *Brown* (related to plasma osmolality and sperm concentration) had
604 all 22 hub genes annotated (Additional File S5). Of these hub genes, 19 were related to RNA metabolism
605 and other metabolic processes. Two members of heat shock 70 kDa protein family members were
606 included as hub genes in this module, *hsp4* and *hsp14*.
607
608 *Overrepresentation of chromosomes in co-expression modules*
609 A single representative Brook Charr transcript per gene was conservatively identified based on alignment
610 position on the Atlantic Salmon genome (see Methods). Of the 69,441 total Brook Charr transcripts in the
611 reference transcriptome, 44,405 aligned to the Atlantic Salmon reference genome containing the main 29
612 chromosomes and unplaced scaffolds. Probably due to the ancestral salmonid genome duplication, gene
613 tandem duplications, and in some cases potentially due to mapping against both a chromosome and an
614 unplaced scaffold, 5,111 of these transcripts aligned to more than one location, resulting in a total of
615 49,516 alignments (MAPQ ≥ 30).
616 After keeping a single transcript per continuous alignment block on the reference genome (see
617 Methods), a total of 21,237 uniquely located gene positions were identified over all Atlantic salmon
618 scaffolds. For females, this included 20,140 unique transcripts, with 1,097 transcripts having two different
619 locations on the reference. For males, this included 20,151 unique transcripts, with 1,086 transcripts
620 having two different locations on the reference. Isolating transcripts to only the main 29 chromosomes of

621 Atlantic Salmon, not including all of the unplaced scaffolds, resulted in a total of 19,913 transcript
622 locations on the 29 chromosomes in both sexes.
623 Of the 19,913 transcripts with chromosomal locations, 7,694 were from assigned modules, 2,330
624 were in the unassigned grey module, and 9,889 were not in the top 25,000 most correlated genes from the
625 female network, and thus were not included in module construction. From the male network, 2,830
626 transcripts located to chromosomes were from assigned modules, 7,340 were in the unassigned grey
627 module, and 9,743 were not in the top 25,000 most correlated genes. The proportions of these transcripts
628 belonging to each chromosome in the 19,913 transcripts, and the proportions of the transcripts in each
629 module were compared to identify significantly enriched chromosomes from different modules (see
630 Figures 7A and 8A; Table S1 and S2).
631 Two of the 14 female modules (14%) showed overrepresentation from specific chromosomes ($p < 0.01$; Table S1; Figure 6A; all modules shown in Figure S7). The *green* module, enriched for translation
633 activity, had 208 unique, positioned genes and these genes were overrepresented on chr26 (5% vs 2% in
634 baseline; $p = 0.005$). Of the ten unique genes from this module identified on chr26, six were annotated
635 with translation-related functions (Additional File S3). This included genes annotated as *probable*
636 *ribosomal biogenesis protein RLP24*, *60S acidic ribosomal protein P2*, *60S ribosomal protein L30*, *40S*
637 *ribosomal protein S13* and *S17,* and *DNA-directed RNA polymerases I, II, and III subunit RPABC5*.
638 Importantly, in this module, one transcript was aligned twice within the same chromosome and so was
639 counted twice (see Methods), and this gene was not related to translation function. This was the only
640 observed instance of an enriched chromosome having two alignments from the same transcript. The
641 *salmon* module (hemopoiesis) had 169 positioned genes and 15 were positioned throughout chr20 ((9%
642 vs. 4% in baseline; $p = 0.006$; Figure 6B). However, only two of the 15 genes had functions related to
643 hemopoiesis, including *ankyrin-1*, involved in binding erythrocyte membrane protein band 4.2 to other
644 membrane proteins and *flavin reductase*, involved in heme catabolism.

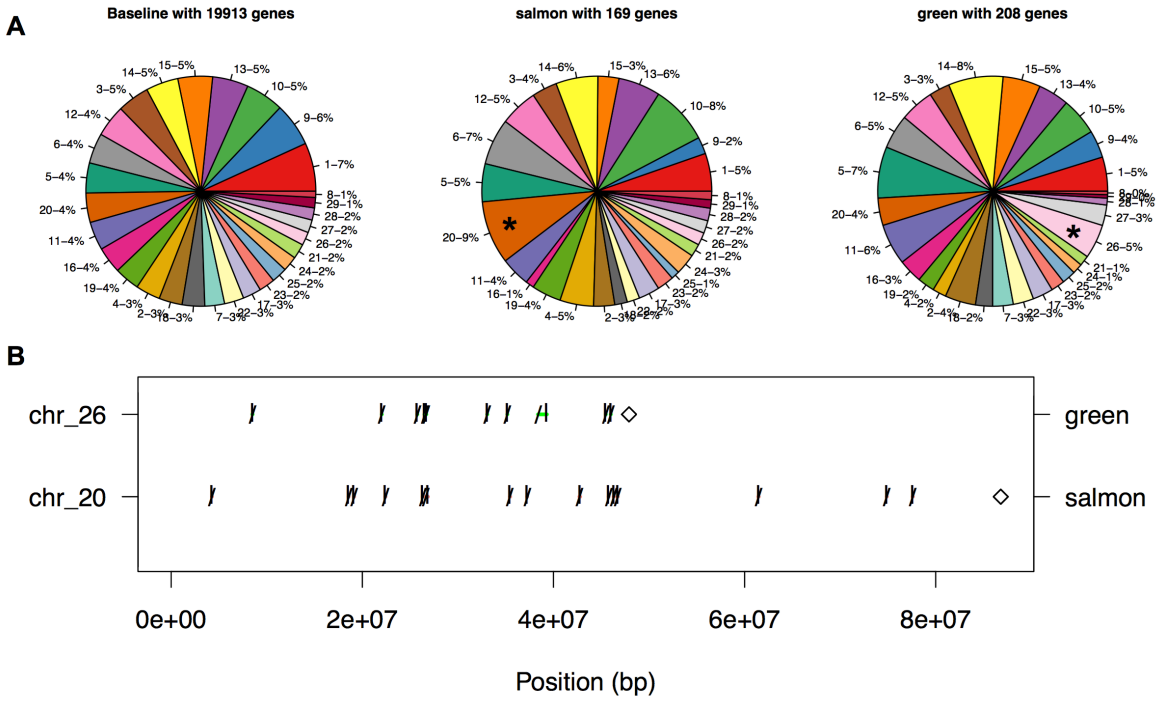


Figure 6. Chromosome overrepresentation evaluated within the female modules. The proportion of transcripts per chromosome is shown (A) for the baseline (all expressed unique genes), the female green module and the female salmon module. Significant overrepresentation of particular chromosomes is represented by an asterisk within the pie chart ($p \leq 0.01$). Positions of genes within the significantly overrepresented chromosome/module combinations are shown in (B), with the name of the chromosome on the first Y-axis and the module on the second Y-axis. Start and stop positions of genes are shown in base pairs (bp) along the x-axis with the start shown by a vertical line and stop by a slight angle line, and the end of the chromosome shown by a diamond. All female module chromosome proportions are shown in Figure S7.

In males, nine of 27 (33%) modules had at least one overrepresented chromosome, and one module (*yellowgreen*) had two overrepresented chromosomes (Figure 7A; all modules shown in Figure S8). Of the 60 positioned genes in *yellow* (translation), five were from chr26 (8% vs. 2% in baseline; $p = 0.005$). Considering that most of the genes in male *yellow* occur in female *green*, the observation that both show enrichment for chr26 is not surprising. Similarly, all five of the genes from this module-chromosome combination are related to ribosomal functions (see Additional File S3). Of the 349 positioned genes in *turquoise* (immune system process), 16 were from chr25 (5% vs. 2% in the baseline; $p < 0.002$). Five of these 16 genes were related to immunity, including *G-protein coupled receptor 183* (involved in chemotaxis of immune cells), *C-X-C chemokine receptor 4* (involved in mediating inflammatory responses to lipopolysaccharide), *interleukin-1 receptor type 2* (reduces interleukin-1B activity), *macrophage receptor MARCO* (involved in pattern recognition of bacteria), and *motile sperm domain-containing protein 2* (involved in chemotactic migration of monocytes and neutrophils). Of the 21 positioned genes in *sienna3* (extracellular structure organization), three genes were from chr21 (14%

vs. 2%; $p < 0.01$), all of which were related to the function of the module (*collagen alpha-3(VI) chain*, *collagen alpha-1(IV) chain*, and *fibronectin*). The rest of the enriched module-chromosome combinations in males were from smaller modules, including *steelblue* (innate immunity) with 19 positioned genes, four of which were in chr5 (21% vs. 4% in baseline; $p = 0.007$), all with functions related to the module function, including *ribonuclease-like 3* (host defense and antibacterial), *leukocyte cell-derived chemotaxin-2* (neutrophil chemotactic activity), *saxitoxin and tetrodotoxin-binding protein 2* (toxin accumulation/secretion) and *c-type lectin domain family 4 member E* (cell surface receptor for ligands including damaged cells, fungi and mycobacterium, involved in secreting inflammatory cytokines). *Skyblue* (14 genes) had four genes from chr13 (29% vs. 5% in baseline; $p = 0.005$), two of which were involved in transcription activity, for which the module was enriched (*zinc finger BED domain-containing protein 4* and *one cut domain family member 2*).

Interestingly, several of these enriched module-chromosome gene sets were weighted to one side of the chromosome in position. For example, the female *green* module had 9/10 genes in the second half of the chromosome (between 20M and 45Mbp; Figure 6B). Male chromosome-module combinations also were weighted to one side of the chromosome, such as *yellow*-chr26, *lightsteelblue*-chr18, *skyblue*-chr13, both *yellowgreen*-chr13 and *yellowgreen*-chr3, and *grey60*-chr3 (Figure 7B). Others, such as *steelblue*-chr5, *plum1*-chr6, *sienna3*-chr21 and *turquoise*-chr25 contained genes that were found throughout the chromosomes. This enrichment only to one side is probably due to the use of the non-focal reference genome. For those weighted to one side of the chromosome, specifically chr13 (i.e., Ssa13) and chr3 (i.e., Ssa03), the chromosomes are not in the same fusion pattern in Brook Charr (Sutherland *et al.* 2016). Importantly, the *yellowgreen* module is enriched within chr13 on the distal end (i.e., Ssa13b) and within chr3 on the distal end (i.e. Ssa03b); Ssa13b corresponds to BC08b and Ssa03b corresponds to BC08a, and therefore, if a Brook Charr reference genome were to be used in this analysis, *yellowgreen* would be highly enriched within all of BC08.

Atlantic Salmon enriched chromosomes where module genes occur throughout the chromosome are often either ancestrally fused prior to the divergence of *Salmo* and *Salvelinus*, such as chr5 (i.e., Ssa05) which corresponds to the fused BC07 (fusion 4.2-20.1 from Figure 4 of Sutherland et al. 2016), or are acrocentric chromosomes in both species, such as chr21 (i.e., Ssa21, corresponds to BC26) or chr25 (i.e., Ssa25, corresponds to BC24). Some exceptions in the examples above exist, including chromosomes with genes found throughout that are in fact two acrocentric chromosomes in Brook Charr (e.g. chr20 or Ssa20 corresponds to acrocentric BC40 and BC25; and chr6 or Ssa06 corresponds to acrocentric BC14 and BC31). Therefore, although trends observed here indicate chromosome-level co-expression and enrichment for specific modules in specific chromosomes, the trends would be stronger if a focal species reference genome was available.

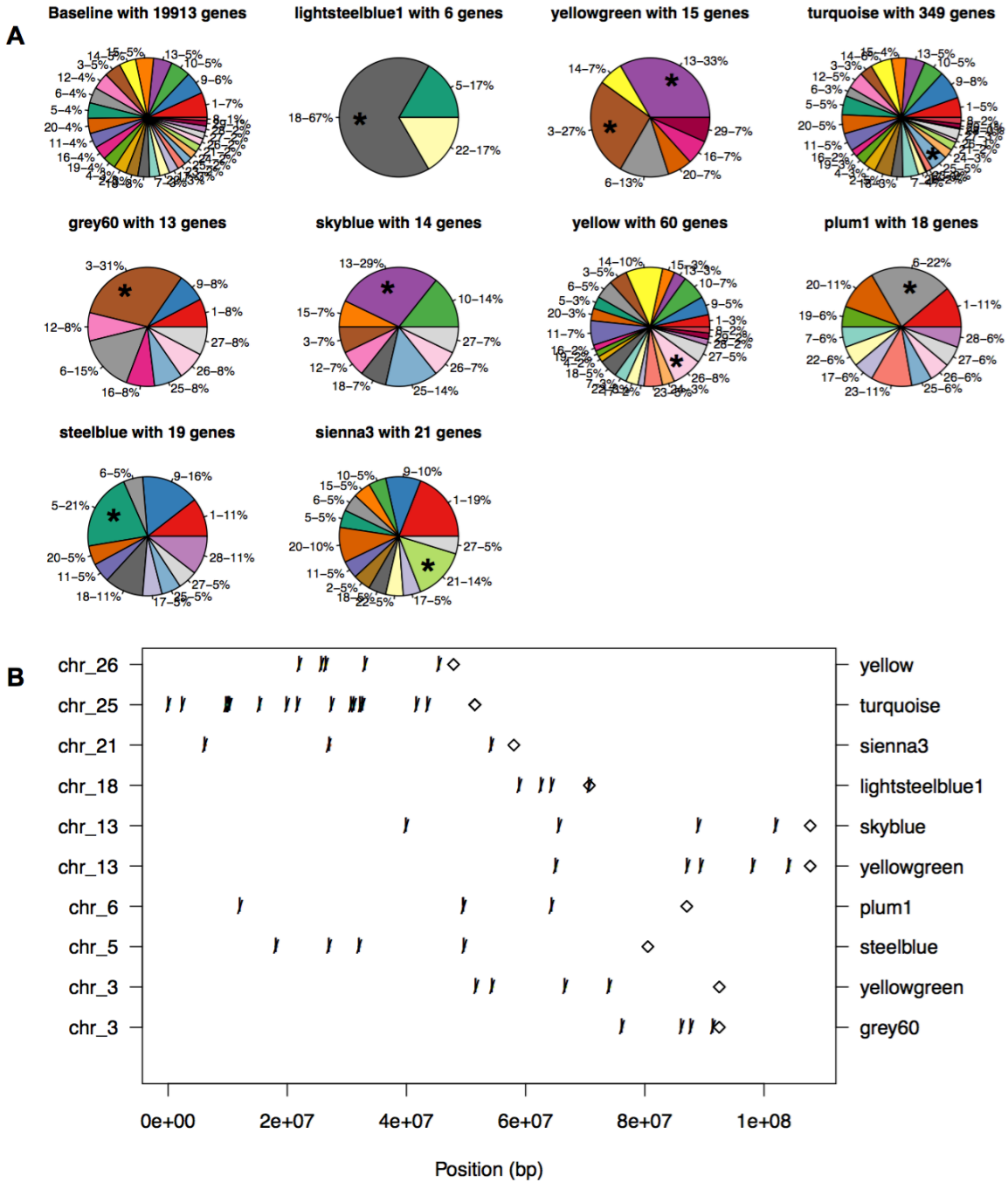


Figure 7. Chromosome overrepresentation evaluated within the male modules. The proportion of transcripts per chromosome is shown (A) for the baseline (all expressed unique genes) and for the rest of the modules with significant overrepresentation. Overrepresentation of particular chromosomes within modules are represented by an asterisk within the pie chart ($p \leq 0.01$). Positions of genes within the significantly overrepresented chromosome/module combinations are shown in (B), with the name of the chromosome on the first Y-axis and the module on the second. Start and stop positions of genes are shown in base pairs (bp) along the x-axis with the start shown by a vertical line and stop by a slight angle line, and the end of the chromosome shown by a diamond. All male module chromosome proportions are shown in Figure S7.

DISCUSSION

Gene co-expression produces complex phenotypes and may underlie key aspects of phenotypic evolution. Further, characterizing the regulatory structure in which genes respond provides valuable insight on organismal response to biotic and abiotic stimuli. Co-expressed genes may share regulatory elements, such as transcription factor binding sites (van Dam *et al.* 2017), which highlights the value of network analyses in characterizing these regulatory elements. The correlation of phenotypic variation with co-expressed genes provides further information on the function of specific modules, and can bring added information to other studies that use modules to classify samples based on gene expression.

In this study, we generated co-expression networks for both female and male Brook Charr liver transcriptomes during a reproductive period and shortly after all fish were exposed to an acute handling stress. We then evaluated module conservation in the opposite sex and in the congener Arctic Charr using adjacency comparisons. Our observations confirmed previous findings showing that co-expression patterns within most modules are conserved between sexes or closely related species (van Nas *et al.* 2009; Wong *et al.* 2014; Cheviron and Swanson 2017), but we also identified several sex-specific modules that provide insight on the evolution of gene expression and phenotypic variance, as well as being potentially involved in resolving sexual conflict through gene regulation (Thompson *et al.* 2015). For each module, we identified hub genes, evaluated functional enrichment and phenotypic correlations, as well as chromosomal positions of module genes to identify whether specific modules were overrepresented on specific chromosomes, thus providing information regarding the underlying mechanisms of regulation. Altogether, this provides a highly comprehensive analysis of the structure of the gene expression networks of a paleopolyploid non-model salmonid.

Sexual dimorphism in co-expression networks and cross-species conservation

The most striking result from an initial inspection of the transcriptome data is the large difference in modularity in the female and male Brook Charr networks. Although females had 76% of the most connected 25 k genes assigning to a module, males had only 28%, and the rest were grouped into the *grey* unassigned module. This may be related to the higher inter-individual variance in gene expression in females than in males (Figure 1). Since variance across samples in gene expression is needed to identify correlation among transcripts, the lower variance in males may have resulted in the observed lower modularity. Interestingly, if modules were built only in females and tested for conservation in males, this phenomenon may have not been noted, as the female modules often were given conserved statistics in males, but when modules were generated in males, many genes were not assigned to a module. Further, the generation of modules in the males identified many important modules that in the female network were all grouped together into the very large module *darkred*, which additionally shows the benefit of

746 generating networks in each sex independently. It will be valuable to inspect sex-specific module
747 generation in other salmonids and in tissues other than liver to understand the generality of these sex
748 differences.

749 Conserved modules between the sexes were often comprised of genes within pathways involved
750 in conserved functions. For example, the most conserved modules between the sexes and species were
751 involved in basic cellular processes involving many co-expressed subunits of a multiple subunit protein
752 complex (e.g. translation). Additionally, immunity-related modules were also conserved between the
753 sexes, and co-expression is often observed in these response pathways (Sutherland *et al.* 2014b).
754 Considering the complexity of these responses and the importance of this function, it is not surprising that
755 immunity modules are conserved. One exception to the sex conservation of translation-related modules
756 was the male *green* module, which was more male-specific, and interestingly was associated with length
757 and weight in males, whereas no female modules had significant associations to these traits. The male
758 module *darkgrey* was also correlated with length and weight and was not conserved in females. Most
759 notably, the male module *ivory*, which was correlated with blood osmolality and chloride levels, was
760 highly male-specific, enriched for transcription factor activity, and contained many putative transcription
761 factors as hub genes, including several genes from the *wnt* protein family. Wnt signaling is associated
762 with gonad differentiation and shows sex-specific expression in several studies in mammals and fish,
763 (Vainio *et al.* 1999; Nicol and Guiguen 2011; Sreenivasan *et al.* 2014; Böhne *et al.* 2014). Since
764 transcription factor activity is often different among cell types, species, or sexes, this module may contain
765 some specific features that contribute to sexual dimorphism of the liver. This may indicate that sex
766 differences in liver tissue may be generated via similar mechanisms operating in the gonads, although to
767 confirm this, more detailed studies are needed. The sex-specific modules identified here may provide
768 insight into mechanisms of gene regulation solving issues of sexual antagonism.

769 The presence of sex-specific modules in the liver may be partly explained by sex hormones
770 produced in the gonads, as they have been shown to regulate a significant proportion of the liver
771 transcriptome in mouse (van Nas *et al.* 2009). In fish, the liver has partially diverged functions between
772 the sexes, as several oocyte precursor proteins are produced in the liver in females (Qiao *et al.* 2016),
773 which may explain some of the differences between sexes observed here. Some of the strongest
774 associations of female modules were to maturity index. Maturity index is affected by sex hormones such
775 as estradiol, which controls reproduction in fish (Garcia-Reyero *et al.* 2018). Therefore, even though most
776 of the network modules were conserved in the opposite sex, phenotypic differences between the sexes are
777 reflected in the liver network differences profiled here.

778 The ranking of module conservation levels in both the opposite sex and in Arctic Charr was often
779 similar, suggesting evolutionary conservation for many gene co-expression modules. Even with the lower

780 sample size in Arctic Charr, conservation was identified for specific modules. Among the most conserved
781 male Brook Charr modules in Arctic Charr with significant phenotype correlations were *brown*, *magenta*,
782 *yellow*, and *turquoise*. *Yellow* and *turquoise* were correlated with growth rate, while *brown* was correlated
783 with sperm concentration and plasma osmolality. Considering this identified conservation, hub genes
784 from these modules may be good candidates as genes with conserved roles in the regulation of growth
785 and sperm concentration in male salmonids. *Magenta* was enriched for endoplasmic reticulum stress and
786 protein folding, also indicating the potential value of hub genes within this module to study responses to
787 stress. The conserved *turquoise* module was enriched for immune response, and marginally associated
788 with growth ($p = 0.05$), and therefore hub genes of this module may be relevant for studying
789 evolutionarily conserved associations between immunity and growth, which are known to occur in a
790 trade-off if energy supply is limited.
791
792 *Modules separated by immune response type*
793 Separate modules were identified for immune functions involving innate antiviral genes (male
794 *darkmagenta*) and innate immunity C-type lectins (male *steelblue*). This is of large interest considering
795 that these types of immune responses have been observed as responding inversely to different agents, with
796 pathogen recognition receptors (e.g. C-type lectins) potentially involved in ectoparasite defense
797 (Sutherland *et al.* 2014b) and being up-regulated during out-migration of steelhead trout *Oncorhynchus*
798 *mykiss* smolts (Sutherland *et al.* 2014a). In both of these observations, innate antiviral genes were
799 suppressed. Alternately, antiviral genes were overexpressed in non-migrating fish of the same species,
800 and were down-regulated in Pink Salmon *O. gorbuscha* infected with salmon lice relative to controls.
801 Even if the genes are not the same between these studies and ours (i.e., no 1:1 association of orthologs has
802 been done for these datasets), the observation of similar functions in separate, but related, modules in the
803 present study may indicate that these functions are hardwired into different modules given that no known
804 infection is occurring within these samples. Importantly, in the present study, unsigned networks were
805 used, and therefore if the two immune response types were completely inversely regulated, they would
806 belong to the same module, which was not observed here. Therefore they are probably more independent
807 and not completely under the same regulatory control. This is a new observation in the regulation of these
808 different immune system processes in salmon, and is a highly important avenue for further study in
809 salmonids given the relevance of these genes to immunity against pathogens, and the potential response
810 outcomes of co-infection occurring between parasitic and viral agents in nature.
811 The immune response modules observed here (i.e. male *darkmagenta*, *steelblue*, and *turquoise*)
812 were all considered as highly conserved between the sexes and moderately to highly conserved in Arctic
813 Charr. It will be valuable to see if these three modules or the genes found within them have conserved

814 expression patterns in other species as these may have important roles in defense responses. It is possible
815 that the co-expression viewed in these modules comes from the occurrence of a specific cell type that is
816 present in different levels in the sampled tissue in different individuals. Single-cell RNA-sequencing of
817 immune cells, or *in situ* gene expression hybridization techniques could address some of these questions.
818 Further, it is valuable to use a microbe profiling platform alongside transcriptome studies of wild sampled
819 individuals to best understand co-infection details (e.g. Miller *et al.* 2016). Nonetheless, these
820 observations will be strengthened when additional analyses are conducted with a broader range of species,
821 once orthologs are identified among the species.
822
823 *Module genes overrepresented on specific chromosomes*
824 Although not structured in operons like prokaryotic genomes, eukaryotic genomes are also known to
825 show clustering of functionally related genes, which can be observed at multiple levels: intra-
826 chromosomally (i.e. clustered by distance within a chromosome) (Santoni *et al.* 2013), inter-
827 chromosomally (i.e. distributed on fewer chromosomes than expected by chance), and within 3D contact
828 space with segments of other chromosomes within a nucleus (Thévenin *et al.* 2014). Co-localization in
829 such a manner may allow for expression of genes within transcription factories in which ribosomes are
830 concentrated, or where chromatin is opened (Thévenin *et al.* 2014). Eukaryotic genes from similar
831 genomic location can show correlated expression profiles even when unrelated by function (Ghanbarian
832 and Hurst 2015). In fact, this co-expression of non-related genes due to genomic location has been
833 suggested to potentially result in spurious positive co-expression at low levels of correlation, where only
834 higher-level correlation co-expression is found to contain genes belonging to the same GO classes
835 (Batada *et al.* 2007). Co-expression of unrelated genes, either at an immediate proximity (< 100 kb) or at
836 longer distances, may be due to chromatin dynamics resulting in a neutral shift in expression of
837 neighbouring genes "piggybacking" selective changes to a focal gene (Ghanbarian and Hurst 2015).
838 Enrichment within a specific chromosome is only one component of profiling these dynamics, but
839 provides a starting point for the analysis.
840 Although in general the co-expression modules were not concentrated into specific chromosomes,
841 some overrepresentation was observed. For example, enrichment of genes related to the module function
842 were observed for female *green* (translation) and *male* turquoise (immunity), *sienna3* (extracellular
843 matrix), *steelblue* (innate immunity), and *skyblue* (transcription activity). Although these few examples
844 are interesting, the generally low proportion of overrepresentation suggests that chromosomal enrichment
845 is probably not a major driver of co-expression patterns in modules identified here. However, Hi-C
846 chromosome maps are not available for salmonids, and so this additional level of proximity cannot be yet
847 evaluated in this system. Nonetheless, this work identifies coexpressed genes related to enriched module

functions that are overrepresented on specific chromosomes, giving evidence for clustering of functionally and transcriptionally related genes, which is an important finding and will be valuable to consider in future studies identifying co-expressed or differentially regulated panels of response genes. Interestingly, the sex-specific modules in the male (i.e., *green*, *darkgrey*, *ivory*) were not found to be overrepresented on the corresponding Atlantic Salmon chromosome (Ssa09a) to the Brook Charr sex chromosome (BC35) here, suggesting that the maintenance of this sexual dimorphic gene regulation is not due to sequestration of these genes on the sex chromosome but rather due to different regulatory architecture.

The discovery of enrichment of the male module *yellowgreen* across the second arm of two different chromosomes in Atlantic Salmon, which in Brook Charr are fused into a single chromosome is an important observation. If co-expression of this module for genes from throughout the Brook Charr chromosome is equally conserved in Atlantic Salmon, it would indicate that chromosomal proximity may matter, but belonging to different chromosomes also allows for co-expression. If co-expression for this module is stronger in Brook Charr, then it may suggest that the fusion of the two chromosome arms matters for co-expression, thus indicating relevance for chromosome fusions and fissions to gene regulation. Additional observations of a single module enriched across an entire chromosome in Atlantic Salmon that was in fact two acrocentric chromosomes in Brook Charr (e.g. *salmon* module enriched on chromosome Ssa20) reduces the likelihood of fused chromosomes enabling higher co-expression. Nonetheless, these will be important questions to investigate as other transcriptome datasets are generated within salmonids, and orthologs between species are identified.

Future comparative approaches

The lack of large effect factors on the data within each sex, for example a treatment regime specific to groups of samples, or contrasting conditions permitted the characterization of the reproductive, post-acute stress response state underlying co-expression modules. Further, the use of all full-sib offspring is expected to remove a large amount of variation that could occur in a similar analysis of outbred individuals. This enabled the generation of networks expected to be largely driven by individual variance in allelic states, slight variances in development, or stochastic inter-individual differences that occur through other biological phenomena such as epigenetic imprinting or feed acquisition. The current dataset fulfills all of the suggested requirements for a successful network analysis including more than 20 samples per condition and > 10M reads per sample (Ballouz *et al.* 2015). For these reasons, this dataset is optimal for understanding the architecture of gene expression and for applying an expression QTL analysis, as these genetic factors are among the only influences expected to be influencing transcription difference among individuals here. Now that hub genes are identified, eQTL investigations can use

882 module gene connectivity information to link network metrics to eQTL effect sizes (Mähler *et al.* 2017).
883 The dataset presented here will also be useful for investigating relationships between cis and trans eQTL
884 with allele-specific expression while considering the co-expression context of the specific genes explored,
885 their positions within the modules, and the general functions of the modules to which they are assigned.
886 Collectively this will improve our understanding of the selective constraints on gene expression
887 regulation as well as the evolution of species-specific or conserved phenotypes.

888 As comparable datasets are produced in other salmonids, it will be valuable to identify whether
889 the conserved modules are conserved outside of the genus *Salvelinus*. However, importantly this will
890 require identification of 1:1 orthologs among the species, which would enable cross-species analyses.
891 Salmonid ortholog identification across reference transcriptomes has recently been conducted for Atlantic
892 Salmon, Brown Trout *Salmo trutta*, Arctic Charr, and European Whitefish *Coregonus lavaretus*, which
893 also included identification of paralogs (Carruthers *et al.* 2018). This type of approach, combined with
894 non-redundant reference transcriptomes will be invaluable in future studies to enable cross species
895 comparisons. If modules are largely conserved between species, as our study suggests within *Salvelinus*
896 liver, this would indicate that large-scale rewiring of baseline transcription networks has not occurred
897 since the base of the lineage. The largest amount of rediploidization is thought to have occurred in the
898 salmonids at the base of the lineage (Kodama *et al.* 2014; Lien *et al.* 2016), although a substantial
899 proportion of ohnologs experienced lineage-specific rediploidization post-speciation events later in
900 evolutionary time (Robertson *et al.* 2017). Given the impact of less-permanent regulatory mechanisms
901 such as epigenetic control in regulating gene expression, one could expect large lineage specific changes
902 in co-expression networks. The impact of the genome duplication and rediploidization on transcriptome
903 networks, including lineage specific changes are important avenues for future study.
904

905 # CONCLUSIONS

906 The co-expression networks of the liver tissue of female and male Brook Charr in a reproductive season
907 shortly after an acute handling stressor were extensively characterized in the present study for correlated
908 phenotypes, enriched gene functions, hub genes, conservation in the opposite sex or in the closely related
909 Arctic Charr, and in terms of chromosomal location of module genes. This study supports previous
910 observations of moderate to high conservation of modules between sexes and closely related species.
911 Highly conserved modules were involved in basic cellular functions and immune functions. Sex-specific
912 modules were enriched for transcription factor activities and associated with sex-specific or sexually
913 antagonistic phenotypes, such as body size. Lower modularity was identified in males, likely explained by
914 less inter-individual variance in the transcriptome. Our results on different immune system response types
915 suggest not only the potential for inverse regulation but also decoupled regulation, which has implications

for responses to co-infections and requires further study. This work found evidence for functionally related genes co-expressed and co-located on specific chromosomes, identifying relevance of genomic location on co-expression regulatory dynamics. The relevance of chromosome arm fusions in co-expression dynamics merits further investigation. In addition to providing new insights on the underlying structure of transcription regulation in the salmonid genus *Salvelinus*, this is also a low-noise, high sample size dataset where expression patterns are putatively driven by slight variations in allelic combinations and stochastic minor differences in development, and will subsequently be a valuable resource for identifying eQTL while considering the co-expression network positions of genes.

DATA ACCESSIBILITY

Brook Charr Phylofish transcriptome assembly (Pasquier *et al.* 2016)
http://phylofish.sigenae.org/ngspipelines/#!/NGSpipelines/Salvelinus%20fontinalis

Atlantic Salmon genome assembly (Lien *et al.* 2016)
https://www.ncbi.nlm.nih.gov/assembly/GCF_000233375.1

Complete documented bioinformatics pipeline: https://github.com/bensutherland/sfon_wgcna

SUPPLEMENTAL INFORMATION

Supplemental Results. Additional figures and tables that support the main text.

Additional File S1. All enrichment analyses for Gene Ontology biological process and molecular function categories for all results with five or more genes and with an enrichment $p \leq 0.01$.

Additional File S2. Overview table for female and male modules, including module size, trait association, GO enrichment, preservation Zsummary and medianRank, and cross-tabulation results.

Additional File S3. Transcripts and annotation for each overrepresented module-chromosome combination for female and male networks ($p \leq 0.01$). Genes with functions related to the function of the module are in bold text.

Additional File S4. Female genes in modules with gene significance and module membership values.

Additional File S5. Male genes in modules with gene significance and module membership values.

ACKNOWLEDGEMENTS

Thanks to Guillaume Côté for laboratory assistance, Jérémy Le Luyer for discussions about WGCNA, Eric Normandeau for discussion on gene annotation, Yann Guiguen for discussion on details of transcriptome assembly, and Aimee-Lee Houde for comments on the manuscript. This work was funded by a Fonds de Recherche du Québec (FRQ) Nature et Technologies research grant awarded to Céline Audet, Louis Bernatchez, and Nadia Aubin-Horth; a grant from the Société de Recherche et de Développement en Aquaculture Continentale (SORDAC) awarded to L.B. and C.A. J.M.P. is supported by the Finnish Cultural Foundation. During this work, B.J.G.S. was supported first by a Natural Sciences and Engineering Research Council of Canada postdoctoral fellowship and subsequently by an FRQ Santé postdoctoral fellowship.

References

Allendorf F. W., Thorgaard G. H., 1984 Tetraploidy and the evolution of salmonid fishes. In:Turner BJ (Ed.), *Evolutionary genetics of fishes*, Plenum Publishing Corporation, New York, pp. 1–53.

Anttila K., Lewis M., Prokkola J. M., Kanerva M., Seppänen E., Kolari I., Nikinmaa M., 2015 Warm acclimation and oxygen depletion induce species-specific responses in salmonids. J. Exp. Biol. **218**: 1471–1477.

Ballouz S., Verleyen W., Gillis J., 2015 Guidance for RNA-seq co-expression network construction and analysis: safety in numbers. Bioinformatics **31**: 2123–2130.

Barson N. J., Aykanat T., Hindar K., Baranski M., 2015 Sex-dependent dominance at a single locus maintains variation in age at maturity in salmon. Nature **528**: 405–408.

Batada N. N., Urrutia A. O., Hurst L. D., 2007 Chromatin remodelling is a major source of coexpression of linked genes in yeast. Trends Genet. **23**: 480–484.

Blackmon H., Brandvain Y., 2017 Long-term fragility of Y chromosomes is dominated by short-term resolution of sexual antagonism. Genetics: genetics.300382.2017–9.

Bolger A. M., Lohse M., Usadel B., 2014 Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics **30**: 2114–2120.

Böhne A., Sengstag T., Salzburger W., 2014 Comparative transcriptomics in East African Cichlids reveals sex- and species-specific expression and new candidates for sex differentiation in fishes. Genome Biology and Evolution **6**: 2567–2585.

Bush W. S., Moore J. H., 2012 Chapter 11: Genome-Wide Association Studies. PLoS Comput Biol **8**: e1002822.

Carruthers M., Yurchenko A. A., Augley J. J., Adams C. E., Herzyk P., Elmer K. R., 2018 *De novo* transcriptome assembly, annotation and comparison of four ecological and evolutionary model salmonid fish species. BMC Genomics **19**: 32.

978 Cheviron Z. A., Swanson D. L., 2017 Comparative transcriptomics of seasonal phenotypic flexibility in
979 two North American songbirds. Integr. Comp. Biol. **57**: 1040–1054.

980 Crête-Lafrenière A., Weir L. K., Bernatchez L., 2012 Framing the Salmonidae family phylogenetic
981 portrait: a more complete picture from increased taxon sampling. PLoS ONE **7**: e46662.

982 Ellegren H., Parsch J., 2007 The evolution of sex-biased genes and sex-biased gene expression. Nat Rev
983 Genet **8**: 689–698.

984 Filteau M., Pavey S. A., St-Cyr J., Bernatchez L., 2013 Gene coexpression networks reveal key drivers of
985 phenotypic divergence in Lake Whitefish. Molecular Biology and Evolution **30**: 1384–1396.

986 Gaiteri C., Ding Y., French B., Tseng G. C., Sibille E., 2014 Beyond modules and hubs: the potential of
987 gene coexpression networks for investigating molecular mechanisms of complex brain disorders.
988 Genes Brain Behav. **13**: 13–24.

989 Garcia-Reyero N., Jayasinghe B. S., Kroll K. J., Sabo-Attwood T., Denslow N. D., 2018 Estrogen
990 signaling through both membrane and nuclear receptors in the liver of Fathead Minnow. General and
991 Comparative Endocrinology **257**: 50–66.

992 Ghanbarian A. T., Hurst L. D., 2015 Neighboring genes show correlated evolution in gene expression.
993 Molecular Biology and Evolution **32**: 1748–1766.

994 Gillis J., Pavlidis P., 2012 "Guilt by association" is the exception rather than the rule in gene networks. (A
995 Rzhetsky, Ed.). PLoS Comput Biol **8**: e1002444.

996 Haas B. J., Papanicolaou A., Yassour M., Grabherr M., Blood P. D., Bowden J., Couger M. B., Eccles D.,
997 Li B., Lieber M., MacManes M. D., Ott M., Orvis J., Pochet N., Strozzi F., Weeks N., Westerman R.,
998 William T., Dewey C. N., Henschel R., LeDuc R. D., Friedman N., Regev A., 2013 *De novo*
999 transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation
1000 and analysis. Nat Protoc **8**: 1494–1512.

1001 Horreo J. L., 2017 Revisiting the mitogenomic phylogeny of Salmoninae: new insights thanks to recent
1002 sequencing advances. PeerJ **5**: e3828–10.

1003 Huang D. W., Sherman B. T., Lempicki R. A., 2009 Systematic and integrative analysis of large gene
1004 lists using DAVID bioinformatics resources. Nat Protoc **4**: 44–57.

1005 Kodama M., Brieuc M. S. O., Devlin R. H., Hard J. J., Naish K. A., 2014 Comparative mapping between
1006 Coho Salmon (*Oncorhynchus kisutch*) and three other salmonids suggests a role for chromosomal
1007 rearrangements in the retention of duplicated regions following a whole genome duplication event.
1008 G3 - Genes|Genomes|Genetics **4**: 1717–1730.

1009 Langfelder P., Horvath S., 2008 WGCNA: an R package for weighted correlation network analysis. BMC
1010 Bioinformatics **9**: 559.

1011 Langfelder P., Horvath S., 2012 Fast R functions for robust correlations and hierarchical clustering. J.
1012 Stat. Soft. **46**: 1–17.

1013 Langfelder P., Luo R., Oldham M. C., Horvath S., 2011 Is my network module preserved and
1014 reproducible? (PE Bourne, Ed.). PLoS Comput Biol **7**: e1001057.

1015 Langfelder P., Mischel P. S., Horvath S., 2013 When is hub gene selection better than standard meta-
1016 analysis? (T Ravasi, Ed.). PLoS ONE **8**: e61505.

1017 Langmead B., Salzberg S. L., 2012 Fast gapped-read alignment with Bowtie 2. Nat Meth **9**: 357–359.

1018 Li H., Handsaker B., Wysoker A., Fennell T., Ruan J., Homer N., Marth G., Abecasis G., Durbin R., 1000
1019 Genome Project Data Processing Subgroup, 2009 The Sequence Alignment/Map format and
1020 SAMtools. Bioinformatics **25**: 2078–2079.

1021 Li W., Godzik A., 2006 CD-HIT: a fast program for clustering and comparing large sets of protein or
1022 nucleotide sequences. Bioinformatics **22**: 1658–1659.

1023 Lien S., Koop B. F., Sandve S. R., Miller J. R., Kent M. P., Nome T., Hvidsten T. R., Leong J. S.,
1024 Minkley D. R., Zimin A., Grammes F., Grove H., Gjuvsland A., Walenz B., Hermansen R. A.,
1025 Schalburg von K., Rondeau E. B., Di Genova A., Samy J. K. A., Olav Vik J., Vigeland M. D., Caler
1026 L., Grimholt U., Jentoft S., Våge D. I., de Jong P., Moen T., Baranski M., Palti Y., Smith D. R.,
1027 Yorke J. A., Nederbragt A. J., Tooming-Klunderud A., Jakobsen K. S., Jiang X., Fan D., Hu Y.,
1028 Liberles D. A., Vidal R., Iturra P., Jones S. J. M., Jonassen I., Maass A., Omholt S. W., Davidson W.
1029 S., 2016 The Atlantic Salmon genome provides insights into rediploidization. Nature **533**: 200–205.

1030 Mackay T. F., 2001 The genetic architecture of quantitative traits. Annu. Rev. Genet. **35**: 303–339.

1031 Mackay T. F. C., Stone E. A., Ayroles J. F., 2009 The genetics of quantitative traits: challenges and
1032 prospects. Nat Rev Genet **10**: 565–577.

1033 MacManes M. D., 2014 On the optimal trimming of high-throughput mRNA sequence data. Front. Genet.
1034 **5**: 13.

1035 Macqueen D. J., Johnston I. A., 2014 A well-constrained estimate for the timing of the salmonid whole
1036 genome duplication reveals major decoupling from species diversification. Proceedings of the Royal
1037 Society B: Biological Sciences **281**: 20132881–20132881.

1038 Mähler N., Wang J., Terebieniec B. K., Ingvarsson P. K., Street N. R., Hvidsten T. R., 2017 Gene co-
1039 expression network connectivity is an important determinant of selective constraint. (NM Springer,
1040 Ed.). PLoS Genet **13**: e1006402.

1041 Miller K. M., Gardner I. A., Vanderstichel R., Burnley T., Li S., Tabata A., Kaukinen K. H., Ming T. J.,
1042 Ginther N. G., 2016 *Report on the performance evaluation of the Fluidigm BioMark platform for*
1043 *high-throughput microbe monitoring in salmon*.

1044 Miller K. M., Günther O. P., Li S., Kaukinen K. H., Ming T. J., 2017 Molecular indices of viral disease
1045 development in wild migrating salmon. Conserv Physiol **5**.

1046 Mueller A. J., Canty-Laird E. G., Clegg P. D., Tew S. R., 2017 Cross-species gene modules emerge from
1047 a systems biology approach to osteoarthritis. NPJ Syst Biol Appl **3**: 13.

1048 Nicol B., Guiguen Y., 2011 Expression profiling of wnt signaling genes during gonadal differentiation
1049 and gametogenesis in Rainbow Trout. Sex Dev **5**: 318–329.

1050 Oldham M. C., Horvath S., Geschwind D. H., 2006 Conservation and evolution of gene coexpression
1051 networks in human and chimpanzee brains. Proceedings of the National Academy of Sciences **103**:
1052 17973–17978.

1053 Pasquier J., Cabau C., Nguyen T., Jouanno E., Severac D., Braasch I., Journot L., Pontarotti P., Klopp C.,
1054 Postlethwait J. H., Guiguen Y., Bobe J., 2016 Gene evolution and gene expression after whole
1055 genome duplication in fish: the PhyloFish database. BMC Genomics **17**: 368.

1056 Pavey S. A., Bernatchez L., Aubin-Horth N., Landry C. R., 2012 What is needed for next-generation
1057 ecological and evolutionary genomics? Trends in Ecology & Evolution **27**: 673–678.

1058 Prokkola J. M., Nikinmaa M., Lewis M., Anttila K., Kanerva M., Ikkala K., Seppänen E., Kolari I., Leder
1059 E. H., 2018 Cold temperature represses daily rhythms in the liver transcriptome of a stenothermal
1060 teleost under decreasing day length. J. Exp. Biol. **221**: jeb170670.

1061 Qiao Q., Le Manach S., Sotton B., Huet H., Duvernois-Berthet E., Paris A., Duval C., Ponger L., Marie
1062 A., Blond A., Mathéron L., Vinh J., Bolbach G., Djediat C., Bernard C., Edery M., Marie B., 2016
1063 Deep sexual dimorphism in adult medaka fish liver highlighted by multi-omic approach. Sci. Rep. **6**:
1064 32459.

1065 Quinlan A. R., Hall I. M., 2010 BEDTools: a flexible suite of utilities for comparing genomic features.
1066 Bioinformatics **26**: 841–842.

1067 R Core Team, 2018 R: A language and environment for statistical computing. R Foundation for Statistical
1068 Computing.

1069 Roberts A., Pachter L., 2013 Streaming fragment assignment for real-time analysis of sequencing
1070 experiments. Nat Meth **10**: 71–73.

1071 Robertson F. M., Gundappa M. K., Grammes F., Hvidsten T. R., Redmond A. K., Lien S., Martin S. A.
1072 M., Holland P. W. H., Sandve S. R., Macqueen D. J., 2017 Lineage-specific rediploidization is a
1073 mechanism to explain time-lags between genome duplication and evolutionary diversification.
1074 Genome Biol. **18**: 111.

1075 Robinson M. D., Oshlack A., 2010 A scaling normalization method for differential expression analysis of
1076 RNA-seq data. Genome Biol. **11**: R25.

1077 Robinson M. D., McCarthy D. J., Smyth G. K., 2010 edgeR: a Bioconductor package for differential
1078 expression analysis of digital gene expression data. Bioinformatics **26**: 139–140.

1079 Rose N. H., Seneca F. O., Palumbi S. R., 2015 Gene networks in the wild: identifying transcriptional
1080 modules that mediate coral resistance to experimental heat stress. Genome Biology and Evolution **8**:
1081 243–252.

1082 Santoni D., Castiglione F., Paci P., 2013 Identifying correlations between chromosomal proximity of
1083 genes and distance of their products in protein-protein interaction networks of yeast. (C Schönbach,
1084 Ed.). PLoS ONE **8**: e57707.

1085 Sauvage C., Vagner M., Derôme N., Audet C., Bernatchez L., 2012a Coding gene single nucleotide
1086 polymorphism mapping and quantitative trait loci detection for physiological reproductive traits in
1087 Brook Charr, *Salvelinus fontinalis*. G3 - Genes|Genomes|Genetics **2**: 379–392.

1088 Sauvage C., Vagner M., Derôme N., Audet C., Bernatchez L., 2012b Coding gene SNP mapping reveals
1089 QTL linked to growth and stress response in Brook Charr (*Salvelinus fontinalis*). G3 -
1090 Genes|Genomes|Genetics **2**: 707–720.

1091 Sreenivasan R., Jiang J., Wang X., Bártfai R., Kwan H. Y., Christoffels A., Orban L., 2014 Gonad
1092 differentiation in Zebrafish is regulated by the canonical wnt signaling pathway. Biology of
1093 Reproduction **90**: e34397–10.

1094 Sutherland B. J. G., Gosselin T., Normandeau E., Lamothe M., Isabel N., Audet C., Bernatchez L., 2016
1095 Salmonid chromosome evolution as revealed by a novel method for comparing RADseq linkage

maps. Genome Biology and Evolution **8**: 3600–3617.

Sutherland B. J. G., Hanson K. C., Jantzen J. R., Koop B. F., Smith C. T., 2014a Divergent immunity and energetic programs in the gills of migratory and resident Oncorhynchus mykiss. Mol. Ecol. **23**: 1952–1964.

Sutherland B. J. G., Koczka K. W., Yasuike M., Jantzen S. G., Yazawa R., Koop B. F., Jones S. R. M., 2014b Comparative transcriptomics of Atlantic *Salmo salar*, Chum *Oncorhynchus keta* and Pink Salmon *O. gorbuscha* during infections with salmon lice *Lepeophtheirus salmonis*. BMC Genomics **15**: 200.

Sutherland B. J. G., Rico C., Audet C., Bernatchez L., 2017 Sex chromosome evolution, heterochiasmy, and physiological QTL in the salmonid Brook Charr *Salvelinus fontinalis*. G3 - Genes|Genomes|Genetics **7**: 2749–2762.

Thévenin A., Ein-Dor L., Ozery-Flato M., Shamir R., 2014 Functional gene groups are concentrated within chromosomes, among chromosomes and in the nuclear space of the human genome. Nucleic Acids Research **42**: 9854–9861.

Thompson D., Regev A., Roy S., 2015 Comparative analysis of gene regulatory networks: from network reconstruction to evolution. Annu. Rev. Cell Dev. Biol. **31**: 399–428.

Vainio S., Heikkilä M., Kispert A., Chin N., McMahon A. P., 1999 Female development in mammals is regulated by Wnt-4 signalling. Nature **397**: 405–409.

van Dam S., Võsa U., van der Graaf A., Franke L., de Magalhães J. P., 2017 Gene co-expression analysis for functional classification and gene-disease predictions. Brief Bioinform.

van Nas A., Guhathakurta D., Wang S. S., Yehya N., Horvath S., Zhang B., Ingram-Drake L., Chaudhuri G., Schadt E. E., Drake T. A., Arnold A. P., Lusis A. J., 2009 Elucidating the role of gonadal hormones in sexually dimorphic gene coexpression networks. Endocrinology **150**: 1235–1249.

Warnes G. R., Bolker B., Bonebakker L., Gentleman R., Hubert W., Liaw A., Lumley T., Maechler M., Magnusson A., Moeller S., Schwartz M., Venables B., gplots: Various R Programming Tools for Plotting Data.

Wijchers P. J., Festenstein R. J., 2011 Epigenetic regulation of autosomal gene expression by sex chromosomes. Trends Genet. **27**: 132–140.

Wong R. Y., McLeod M. M., Godwin J., 2014 Limited sex-biased neural gene expression patterns across strains in Zebrafish (Danio rerio). BMC Genomics **15**: 905–10.

Wu T. D., Watanabe C. K., 2005 GMAP: a genomic mapping and alignment program for mRNA and EST sequences. Bioinformatics **21**: 1859–1875.